

MANUAL DE INTRODUCCIÓN A SPSS 10.0

Profesores:

Vicente Candela Moreno
Francisco Ferrández Agulló
Javier Montoyo Bojo
María del Mar Pujol López
Antonio Requena Ruiz
José Requena Ruiz
Ramón Rizo Aldeguer

Departamento de Ciencia de la Computación e Inteligencia Artificial
Report Interno DCCIA-2001-1

1. INTRODUCCIÓN

1.1. Descripción del producto

SPSS es un potente sistema de análisis estadístico y gestión de datos.

Ofrece un rápido entorno de modelización visual que abarca desde lo más simple hasta lo más complejo para crear modelos de manera interactiva y realizar cambios utilizando técnicas analíticas probadas y acreditadas. Permite sacar partido a los datos utilizando una completa gama de productos respaldados por más de 30 años de experiencia en el campo del software analítico. Este manual está basado en la versión 10.0.6.

SPSS opera en las más conocidas plataformas: MS Windows, Macintosh, Digital y UNIX. En el caso de MS Windows, los requisitos de hardware y de software mínimos para ejecutar el sistema son:

- ♦ Windows 95, Windows 98, Windows NT 4.0 o Windows 2000.
- ♦ Procesador Pentium o similar a una velocidad de 90MHz o superior.
- ♦ 16MB o más de memoria de acceso aleatorio (RAM). Se recomienda 64MB o más.
- ♦ Disco duro con un mínimo de 80MB de espacio disponible (para el sistema Base).
- ♦ Unidad de CD-ROM.
- ♦ Adaptador gráfico con una resolución mínima de 800 x 600 (SVGA).
- ♦ Para la conexión con un servidor SPSS, es necesario un adaptador de red con el protocolo TCP/IP.

La interfaz del entorno de ventanas proporciona una relación con el sistema rápida y cómoda, pero internamente funciona mediante comandos.

El sistema está compuesto por módulos que pueden instalarse por separado pero funcionan conjuntamente. Los módulos más importantes, y que han sido utilizados en la elaboración de este manual, son:

- ♦ **Base (Base)**
Módulo con funciones fundamentales necesario para utilizar los otros. Estadísticas, gráficos y tablas de pivote multidimensionales.
- ♦ **Modelos de regresión (Regression Models)**
Modelos predictivos avanzados respecto a los de regresión lineal simple. Medición de distancias entre datos.
- ♦ **Modelos avanzados (Advanced Models)**
Análisis de relaciones complejas con un conjunto de procedimientos avanzado, y análisis de datos de supervivencia.
- ♦ **Tablas (Tables)**
Resultados dispuestos en potentes tablas con los estadísticos que se requieran.
- ♦ **Tendencias (Trends)**
Análisis de series temporales como información histórica. Creación de modelos y acontecimientos futuros.

- ♦ **Categorías (Categories)**

Completo grupo de herramientas para la investigación de productos con el análisis de conjunto y la familia de procedimientos de escalamiento.

- ♦ **Análisis conjunto (Conjoint)**

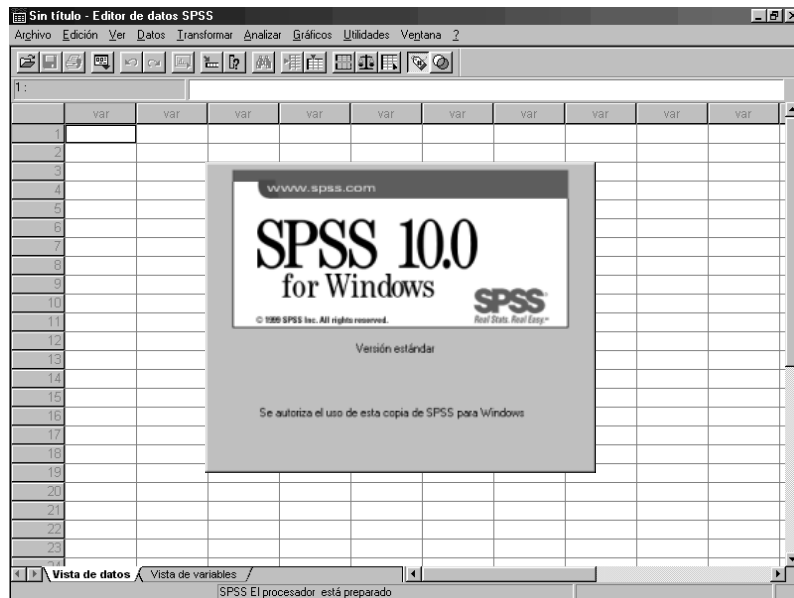
Creado con el objetivo de proporcionar una forma de evaluación del impacto de los atributos de productos individuales sobre determinadas preferencias.

- ♦ **Pruebas exactas**

Proporciona dos métodos adicionales para calcular los niveles de significación de los estadísticos disponibles mediante tablas de contingencia y pruebas no paramétricas.

1.2. Ejecución

Para ejecutar SPSS se accederá al botón Inicio y desde allí a Programas / SPSS for Windows / SPSS 10.0 para Windows. Aparece la presentación del programa con el *Editor de datos* al fondo:

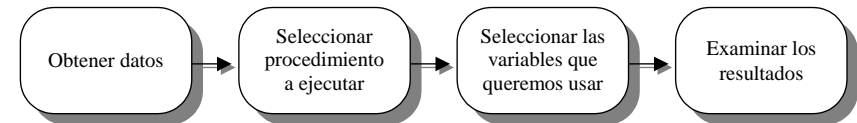


A continuación nos preguntará cómo queremos comenzar: ejecutando el tutorial, introduciendo datos nuevos, abriendo datos de disco, etc. Una vez hecho esto, nos dejará con el *Editor de datos*.

Se pueden abrir varias sesiones de SPSS independientes para poder trabajar con diferentes ficheros de datos.

1.3. Forma de trabajar: pasos básicos

Para trabajar con SPSS debemos seguir los 4 pasos básicos que se muestran a continuación:



1.3.1. Obtención de datos

Existen dos formas de obtener datos en SPSS. La primera es la introducción directa de los datos en las casillas que nos muestra el *Editor de datos*. Las columnas nos indican las variables que tenemos, y las filas los casos de las variables. Al introducir un dato en una casilla, se activa la variable de la columna donde nos encontremos y se define automáticamente según el tipo de dato que hayamos introducido. Si queremos definir las variables nosotros mismos, debemos seleccionar la *Vista de variables* donde podremos especificar su nombre, tipo, tamaño y otros atributos.

La segunda forma de obtener datos es extrayéndolos de un archivo de disco. Éste puede ser de formato SPSS (*.sav); de formato Excel, Lotus, texto, etc.; o de formato de base de datos como Access, dBase, FoxPro, etc.

Si los datos son nuevos o los hemos modificado es conveniente almacenarlos en disco a través del menú *Archivo*.

1.3.2. Selección de procedimiento

Una vez tenemos los datos en el *Editor de datos*, debemos elegir un procedimiento en el menú *Analizar* para poder obtener resultados. El sistema nos permite realizar informes, estadísticos descriptivos, tablas, correlaciones, análisis de regresión, etc. Asimismo si deseamos un gráfico descriptivo de los datos, elegiremos uno de los existentes en el menú *Gráficos*. Éste puede ser normal o interactivo, en cuyo caso podremos modificar posteriormente las variables usadas y los parámetros del mismo.

1.3.3. Selección de variables y opciones

Ahora se nos solicita la selección de las variables que queremos usar en el procedimiento de análisis o en el gráfico. El cuadro de selección de variables es similar en todos los procedimientos: nos muestra una lista de todas las variables y una o varias casillas de destino donde debemos introducirlas para que el sistema pueda calcular los estadísticos o valores que precise para la presentación de resultados.

1.3.4. Examen de resultados

El último paso es el examen de resultados en una nueva ventana llamada *Visor*. Esta ventana nos muestra dos paneles. El de la izquierda es el conjunto de los titulares de los elementos gráficos situados en el panel de la derecha. Su apariencia y funcionamiento es similar al del explorador de Windows. El de la derecha muestra los elementos gráficos, que pueden ser textos, tablas o gráficos.

Los resultados o elementos gráficos que vayamos generando no se sustituyen, sino que se van acumulando en el visor donde les podremos dar formato para su posterior impresión, así como para su almacenamiento en disco. En el caso de los gráficos interactivos podremos modificarlos directamente, sin necesidad de volver a generarlos de nuevo.

1.4. Ventanas

Ya hemos comentado la existencia de dos ventanas: el *Editor de datos* y el *Visor*, pero SPSS utiliza además otras. He aquí un resumen de las diferentes ventanas existentes:

Editor de datos	Muestra el contenido del fichero de datos. Permite crear nuevos archivos de datos o modificar los existentes. Esta ventana se abre automáticamente cuando se inicia una sesión de SPSS. No se puede tener más de un archivo de datos abierto al mismo tiempo. Para ello debemos abrir varias sesiones. Está compuesto de dos vistas: <i>Vista de datos</i> y <i>Vista de variables</i> .
Visor	Muestra los resultados estadísticos, tablas y gráficos que se generen. Permite editar los resultados, almacenarlos e imprimirlos. Esta ventana se abre automáticamente la primera vez que se ejecuta un procedimiento que genera resultados.
Visor de borrador	Esta ventana muestra los resultados como texto normal (en lugar de como tablas pivote interactivas).
Editor de tablas pivote	Permite modificar los resultados mostrados en este tipo de tablas. Podemos editar el texto, intercambiar los datos de las filas y las columnas, añadir colores, crear tablas multidimensionales y ocultar y mostrar los resultados de manera selectiva.
Editor de gráficos	Permite modificar los gráficos y diagramas. Es posible cambiar los colores, seleccionar diferentes tipos de fuentes y tamaños, intercambiar los ejes horizontal y vertical, rotar diagramas de dispersión 3-D e incluso cambiar el tipo de gráfico.
Editor de resultados de texto	Los resultados de texto que no aparecen en las tablas pivote pueden modificarse con este editor. Es posible variar el texto y cambiar las características de las fuentes (tipo, estilo, color y tamaño).
Editor de sintaxis	Permite escribir conjuntos de comandos para que el sistema los ejecute (SPSS funciona internamente a base de comandos). Podemos pegar las selecciones de un cuadro de diálogo en una ventana de sintaxis, donde aparecerán en forma de comandos. Este editor permite utilizar las funciones especiales de SPSS que no se encuentran disponibles en los cuadros de diálogo. También podemos guardar los comandos en un archivo para utilizarlos en sesiones posteriores.
Editor de procesos	El procesamiento y la automatización OLE permiten personalizar y automatizar muchas tareas en SPSS. Este editor permite crear y modificar procesos básicos.

1.5. Barra de menús

Desde esta barra podemos acceder a todas las funciones que implementa el sistema. Cada ventana posee su propia barra de menús pero varios de ellos son comunes a todas las ventanas.

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

Los menús más importantes se describen a continuación:

Archivo	⇒	Permite abrir, cerrar, guardar e imprimir archivos.
Edición	⇒	Para cortar, copiar y pegar elementos, además de acceder a las opciones generales.
Ver	⇒	Podemos hacer que se muestren u oculten diversos elementos del sistema.
Datos	⇒	Permite realizar cambios en los datos como combinar ficheros, transponer variables y casos, o crear grupos de casos para analizar.
Transformar	⇒	Permite modificar los datos de las variables a partir de los ya existentes, en sí mismas o en otras nuevas.
Analizar	⇒	Encontramos en este menú los procedimientos de análisis estadístico.
Gráficos	⇒	Crea gráficos de barras, histogramas, etc., normales o interactivos.
Utilidades	⇒	Varias utilidades: definir conjuntos, ejecutar procesos, editar menús, etc.
Ventana	⇒	Gestión de ventanas abiertas y cambio de unas a otras.
?	⇒	Acceso a la ayuda de SPSS, basada en temas.

1.6. Barras de herramientas

Además de la barra de menús, la interfaz de SPSS incorpora una o varias barras de herramientas según la ventana en la que nos encontremos. Estas barras contienen las funciones más usadas en el trabajo usual para facilitar el acceso a ellas.

Si detenemos el puntero del ratón unos segundos sobre una de las herramientas (o botones) nos presentará un pequeño cuadro con una breve descripción de la utilidad del mismo:



Las barras, además, son personalizables, pudiendo añadir o eliminar los botones que queramos así como moverlos; incluso permite cambiar de sitio la barra o borrarla completamente.

1.7. Ayuda

El acceso a la ayuda a través de su menú nos permite obtener todos los temas expuestos en el formato de ayuda de Windows. Estos temas tratan los aspectos básicos del sistema. Uno de ellos se expone a modo de breve tutorial de SPSS.

Otras opciones de la ayuda son la *Guía de sintaxis*, para aprender la sintaxis de los comandos internos, y el *Asesor estadístico*, que nos va guiando a través de una serie de pasos hasta obtener los resultados que deseamos.

La otra forma de obtener ayuda es la forma contextual. Por ejemplo, en los cuadros de diálogo, pulsando el botón derecho del ratón nos muestra una breve descripción del elemento sobre el que nos hallamos; o bien nos muestra un menú contextual con la opción “¿Qué es esto?”. Esta posibilidad la tenemos incluso en los elementos gráficos en la ventana *Visor*. Por último, en muchos cuadros aparece un botón de ayuda para acceder directamente al tema referente al cuadro en el que nos encontremos.

2. OBTENCIÓN DE DATOS

2.1. Matriz de datos

La matriz que nos presenta el *Editor de datos* contiene todas las observaciones de las variables que necesitamos para poder realizar el análisis estadístico.

Cada columna representa una variable única y cada fila un caso individual de cada variable. Para introducir datos basta con pinchar sobre una casilla y teclear el dato. Si donde tecleamos el dato la variable no está definida, SPSS asigna automáticamente un nombre de variable a la columna y la define con sus parámetros por defecto.

VARIABLES						
	id	sexo	fechnac	edu	catlab	salario
CASOS	1	Hombre	03.02.52	15	Directivo	\$57,000
	2	Hombre	23.05.58	16	Administ	\$40,200
	3	Mujer	26.07.29	12	Administ	\$21,450
	4	Mujer	15.04.47	8	Administ	\$21,900
	5	Hombre	09.02.55	15	Administ	\$45,000
	6	Hombre	22.08.58	15	Administ	\$32,100
	7	Hombre	26.04.56	15	Administ	\$36,000

2.2. Definición de variables

Si pulsamos en la pestaña *Vista de variables* accederemos a una matriz similar a la de datos, pero en este caso contiene las variables definidas. Cada fila representa una variable y cada columna un parámetro o especificación de esa variable. Éstos se asignan por defecto, pero podemos modificarlos a nuestra conveniencia. La lista de parámetros es:


Nombre	➤	Debemos introducir el nombre de la variable, de 8 caracteres como máximo, que pueden ser letras, números o el símbolo de subrayado (_). El primer carácter únicamente puede ser una letra, y no se hacen distinciones entre mayúsculas y minúsculas.
Tipo	➤	A través del botón que aparece al lado del tipo se accede a un panel donde seleccionar el tipo de variable: numérica, coma, fecha, moneda, cadena, etc., su anchura y los decimales (que también se pueden seleccionar mediante las columnas siguientes). Además, dentro del tipo cadena debemos distinguir entre la corta (≤8 caracteres) y la larga (>8).
Anchura	➤	Número de caracteres o dígitos de la variable.
Decimales	➤	Número de decimales si la variable es de tipo contable (numérico, moneda, etc.).
Etiqueta	➤	Dada la imposibilidad de dar un nombre de variable mayor de 8 caracteres, se usan etiquetas para describir la variable de una forma más clara. Aunque internamente se trabaje con los nombres de las variables, en la presentación de resultados podremos utilizar las etiquetas.

Valores	➤	Permite asignar etiquetas a determinados valores de las variables. Por ejemplo, una variable numérica puede contener valores de los datos que sean: 0 si es mujer y 1 si es hombre, entonces asignamos a los valores 0 y 1 las etiquetas de valor <i>mujer</i> y <i>hombre</i> respectivamente. De esta forma cuando trabajemos con ellos, nos aparecerá <i>mujer</i> y <i>hombre</i> en lugar de 0 y 1.
Perdidos	➤	Nos permite definir los valores que se tratarán como perdidos. Hemos de distinguir entre los valores declarados por el sistema (ausencia de dato) y los valores declarados como perdidos por nosotros mismos.
Columnas	➤	Es la anchura de la columna en la <i>Vista de datos</i> .
Alineación	➤	Posición del dato en la casilla en la <i>Vista de datos</i> .
Medida	➤	Permite especificar el nivel de medida como escala (datos numéricos de una escala de intervalo o de razón), ordinal (representan categorías ordenadas) o nominal (representan categorías sin orden alguno). Esta especificación sólo se usa en procedimientos gráficos.

En general, es importante saber distinguir entre variable categórica y variable de escala. Una variable categórica es aquella que tiene un número limitado de valores o categorías distintas. Todas las variables de cadena y las variables numéricas con etiquetas de valor definidas, o las variables numéricas definidas como nominales u ordinales, se tratan como categóricas. El resto se tratarán como de escala.

2.3. Introducción de datos

Con las variables definidas podemos pasar a la introducción de los datos. Para desplazarnos por la matriz de datos podemos usar las flechas del cursor así como la tecla [Intro] para pasar al siguiente caso de la misma variable, y la tecla [Tab] para pasar a la siguiente variable.

Los valores de datos definidos con etiquetas nos permiten introducirlos como tales o con el nombre de la etiqueta. Con la opción del menú *Ver / Etiquetas de valor* o con el botón  podemos mostrar las etiquetas en lugar del valor original. Al seleccionar una casilla nos aparece una lista desplegable con todas las opciones que tenemos para la variable.




Para modificar un dato es suficiente con pinchar su casilla y editarlo, bien en la misma casilla, bien en el *Editor de casillas* (situado entre la barra de herramientas y el *Editor de datos*). Debemos tener en cuenta que en el *Editor de casillas* no podemos introducir ninguna fórmula u operación.

2.4. Uso del Editor de datos y variables

El *Editor de datos* nos permite realizar múltiples operaciones con la matriz de datos. Básicamente con las operaciones de cortar, copiar y pegar del menú *Edición* podemos realizar cualquier desplazamiento o copia de los datos que se nos antoje. Podemos copiar un dato y pegarlo en otra o varias celdas, o copiar uno o varios casos completos en otra u otras filas, o copiar grupos de casos de una o varias variables, o variables completas, sustituyendo a otras o

creando otras nuevas. En resumen, casi cualquier operación es posible. Los datos que no le proporcionemos directamente los completará el sistema con los valores definidos por defecto.

En el caso de la ventana de variables el funcionamiento es el mismo, con la restricción de que no podemos modificar las especificaciones definidas de las variables.

Asimismo podemos insertar casos y variables mediante el menú *Datos* o los botones  o desplazarnos a un caso determinado con el comando *Ir a caso* () o a una variable ().

2.5. Menú Archivo

Dentro del menú *Archivo* podemos acceder a multitud de operaciones sobre los archivos. La primera opción es la de crear un nuevo archivo, ya sea de datos (*.sav), de sintaxis (*.sps), de resultados (*.spo), de resultados de borrador (*.rtf), o de procesos (*.sbs). En una misma sesión SPSS no podremos tener más de un archivo de datos abierto simultáneamente. La segunda opción es la de abrir un archivo existente en disco de cualquiera de los tipos mencionados, pero en este caso no sólo podremos leer datos propios de SPSS, sino que podremos acceder a otros tipos de archivo como los pertenecientes a Excel, Lotus, dBase, etc.

También podemos obtener un fichero de datos a partir de una base de datos externa. Para ello SPSS dispone de un asistente para realizar consultas a la base de datos y obtener la matriz de datos que precisemos.

La última opción para obtener datos es la de ficheros de texto. Para importar con éxito un fichero de texto deberemos seguir una serie de pasos guiados por un asistente.

Las siguientes opciones son las de *Guardar* y *Guardar como* que nos permiten almacenar nuestros datos en disco según el formato propio de SPSS o de Excel, dBase, Lotus, texto, etc.

Mostrar información de datos nos presenta a través del *Visor* la información de cualquier archivo de disco.

Aplicar diccionario de datos aplica al archivo de trabajo la información (etiquetas, valores perdidos y formatos) del archivo de datos de disco que nos solicita. La información se aplica basándose en los nombres de las variables coincidentes, de modo que el resto de variables permanecen intactas.

Hacer caché de datos nos crea una copia temporal de la tabla de datos en disco. Esto puede ser interesante para acelerar las lecturas de los datos siempre que tengamos espacio suficiente en disco.

SPSS también posee una opción de *Presentación preliminar* que nos muestra cómo quedarán los datos una vez impresos.

El análisis en modo distribuido permite utilizar un equipo remoto, normalmente de mayor potencia que el nuestro, para realizar trabajos de gran esfuerzo computacional. Para ello precisamos de una versión local (o cliente) en nuestro equipo y de una versión servidor en el equipo remoto. Pulsaremos en la opción *Cambiar servidor*, seleccionaremos el servidor remoto al que deseamos conectarnos, e introduciremos el IDentificador y la contraseña, y el dominio si es necesario. Una vez conectados abriremos un archivo remoto a través del servidor. Si deseamos un archivo de nuestro propio equipo, deberemos llegar hasta él a través del servidor (ruta de acceso UNC, Convención de Denominación Universal). El resto del análisis distribuido lo hará el sistema por sí solo, de forma transparente al usuario.

2.6. Menú Edición

En el menú *Edición* tenemos dos opciones de inestimable ayuda: *Deshacer* (Ctrl+Z) y *Rehacer* (Ctrl+Y), que lógicamente permanecerá desactivada hasta que deshagamos algún entuerto). No debemos confiarnos siempre a la opción de *Deshacer* ya que no siempre se nos permite ejecutarla, sobre todo cuando actuamos directamente sobre el fichero de disco.

El resto de opciones son las usuales de cualquier programa basado en Windows: *Cortar*, *Copiar*, *Pegar*, *Eliminar*, etc. Además de la opción *Buscar* que nos encuentra cualquier conjunto de caracteres numéricos o alfanuméricos en la matriz de datos.

Por último, el comando *Opciones* nos presenta una panel con las especificaciones generales de SPSS. Aquí se nos permite cambiar los directorios usados, los sistemas de medida, las diferentes configuraciones y formatos de los visores, de los gráficos, de las tablas, de los datos, de los procesos, etc.

2.7. Menú Ver

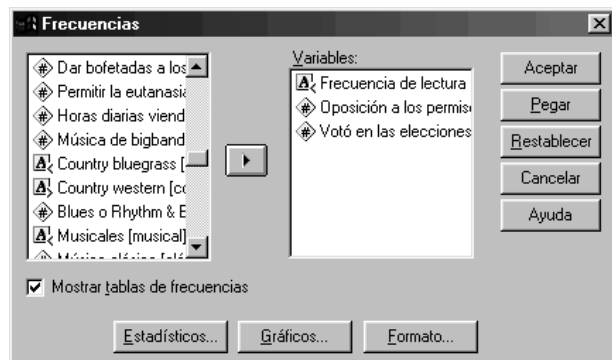
A través del menú *Ver* se accede al formato de visualización de los datos como la fuente empleada o la visibilidad de la cuadrícula o de la barra de estado. Esta última (situada en la parte inferior de la ventana) nos muestra información relativa al sistema y a los datos. Además podemos configurar las barras de herramientas, crear nuevas y personalizarlas a nuestro antojo.


Por último, nos permite conmutar entre la *Vista de datos* y la *Vista de variables*.

2.8. Uso de los cuadros de diálogo: selección de variables y opciones




Antes de continuar debemos conocer el uso de los cuadros de diálogo utilizados para seleccionar variables y opciones. Casi todos los procedimientos nos muestran este tipo de cuadros como método visual y compacto de personalización de la tarea a ejecutar.

Un cuadro de diálogo tiene el siguiente aspecto:



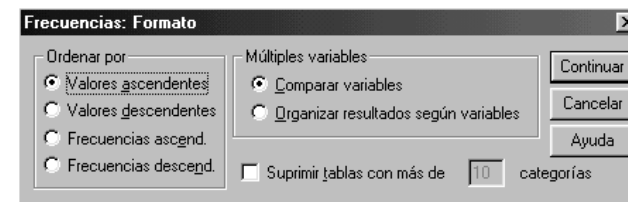
En la lista de la izquierda aparecen las variables susceptibles de ser empleadas en el procedimiento. Para seleccionarlas hemos de pinchar primero sobre ellas y después sobre el botón . Podemos seleccionar varias, si el procedimiento lo permite, usando la tecla [Ctrl]. En la imagen mostrada hemos seleccionado las variables *Frecuencia de...*, *Oposición a...* y *Votó en...* de forma que ahora aparecen en la lista de la derecha. Además pueden aparecer varias listas de destino según las necesidades del procedimiento.

A la izquierda de las variables aparece un símbolo indicativo de su naturaleza:

-  Variable numérica de cualquier tipo.
-  Variable de cadena menor o igual a 8 caracteres.
-  Variable de cadena mayor de 8 caracteres.

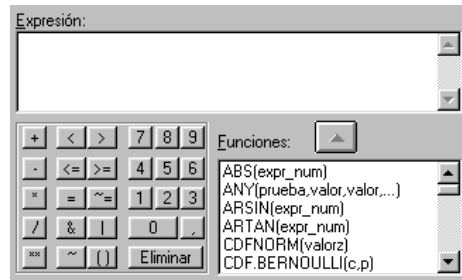
Un cuadro puede contener diferentes controles (casillas de selección, listas desplegadas, etc.) para configurar el análisis a nuestro antojo. En la figura mostrada sólo aparece uno de ellos, una casilla de selección, para indicar que queremos ver las tablas de frecuencias.

Cuando las opciones de configuración son demasiado extensas se proporcionan otros cuadros de diálogo auxiliares accesibles desde determinados botones mostrados normalmente en la parte inferior. Por ejemplo, pulsando el botón *Formato* se nos presenta un nuevo cuadro:

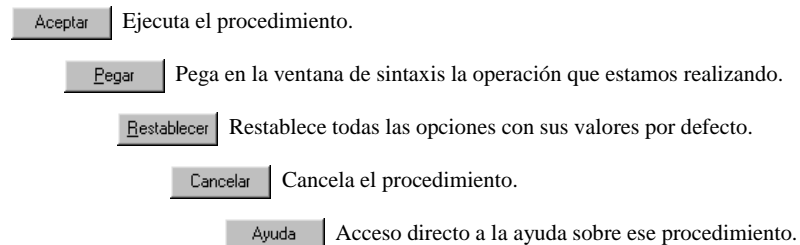


en el que podremos decidir la ordenación y otros parámetros del análisis. Cuando hayamos terminado sólo es preciso pulsar el botón *Continuar* para volver al cuadro de diálogo inicial.

A veces es preciso utilizar un cuadro de introducción de expresiones que puede contener variables, operaciones, constantes, funciones, etc. Debemos tener en cuenta que las constantes de cadena deben ir entre comillas o apóstrofes, y que las constantes numéricas se deben escribir en formato americano, con el punto como separador de la parte decimal.

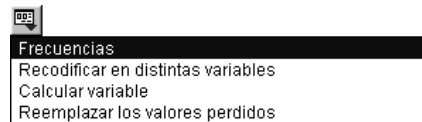


Por último, la parte extrema derecha del cuadro (a veces la parte inferior) nos muestra un conjunto de botones comunes a casi todos los cuadros:



2.9. Rellamada de los cuadros de diálogo

En la barra de herramientas encontramos un botón para acceder rápidamente a los cuadros de diálogo usados con mayor asiduidad. Al pulsar este botón, aparece una lista ordenada de forma que los procedimientos usados más recientemente se situarán en posiciones superiores:



Al terminar la sesión, esta lista se almacenará en disco para poder usarla en sesiones posteriores. De esta manera siempre tendremos rápidamente disponibles los procedimientos que usemos con elevada frecuencia.

2.10. Menú Datos

Definir fechas genera variables de fecha que se utilizan para establecer la periodicidad de una serie temporal (cada caso corresponde a un momento de tiempo distinto y deben

espaciarse de forma uniforme) y para etiquetar los resultados de los análisis de series temporales. Genera una anotación en el *Visor*.

Además de las ya comentadas opciones de inserción y desplazamiento podemos *Ordenar los casos* por una o más variables de forma ascendente o descendente.

Transponer crea un archivo de datos nuevo en el que se transponen las filas y las columnas del archivo de datos original de manera que los casos (las filas) se convierten en variables, y las variables (las columnas) en casos. Los nombres de las nuevas variables se crean automáticamente si no les asignamos nosotros una variable determinada en la que todos sus casos sean diferentes. Una vez realizada la transposición, aparecerá en primer lugar una variable llamada *case_lbl* cuyos casos serán los nombre de las antiguas variables.

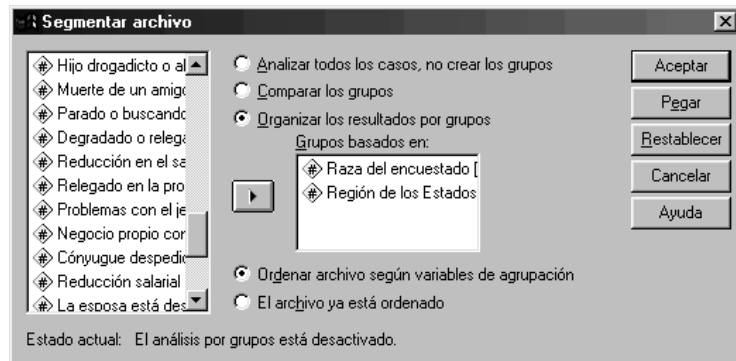
Fundir archivos nos permite unir los casos o las variables de dos archivos distintos. El sistema nos pedirá un archivo de disco para fundir con el que tenemos abierto. A través de un cuadro de diálogo indicaremos las variables con las que nos quedaremos de ambos archivos. Las variables coincidentes se asignarán automáticamente al nuevo archivo (aparecerán en el panel de la derecha) y las que no, vendrán indicadas con (+) o con (*) en el panel de la izquierda según el archivo al que pertenezcan. Desde este panel podremos incluir variables en el nuevo archivo, bien de forma independiente, bien casadas por parejas.

Agregar datos combina grupos de casos en casos de resumen únicos y crea un nuevo archivo de datos agregado (llamado por defecto AGR.SAV). Debemos indicar las variables por las que se va a segmentar el archivo y las variables que aparecerán agrupadas. El nuevo archivo de datos contiene un caso para cada grupo.

La generación de *Diseño ortogonal* crea un archivo de datos que contiene un diseño de efectos principales ortogonales que permite contrastar de forma estadística varios factores sin probar cada combinación de los niveles de factores. Este diseño puede mostrarse mediante *Mostrar diseño*. Se debe definir al menos un factor, cuyo nombre puede ser cualquier nombre de variable, y definir los valores de cada uno de ellos.

Si no reemplazamos el actual archivo de trabajo, se crea automáticamente un archivo en disco llamado por defecto ORTO.SAV, que contiene el diseño ortogonal.

Segmentar archivo divide el archivo de datos en distintos grupos para el análisis basándose en los valores de una o más variables de agrupación. Si selecciona varias variables de agrupación, los casos se agruparán por variable dentro de las categorías de la variable anterior de la lista *Grupos basados en*. Cuando pasemos al análisis obtendremos resultados divididos según los valores de las variables que hayamos seleccionado con la opción *Organizar los resultados por grupos*.



Si elegimos comparar los grupos, los resultados se mostrarán juntos según los grupos para poder compararlos. Es necesario que el archivo esté ordenado según las variables de agrupación, de lo contrario tendremos que indicárselo al sistema.

Seleccionar casos proporciona varios métodos para seleccionar un subgrupo de casos basándose en criterios que incluyen variables y expresiones complejas. Asimismo se puede seleccionar una muestra aleatoria de casos. Los criterios usados para definir un subgrupo pueden incluir valores y rangos de las variables (del tipo que sean), números de caso (filas), expresiones aritméticas y lógicas, y funciones. Los casos no seleccionados porque no cumplan los criterios pueden ser filtrados o eliminados. Con este procedimiento conseguimos que cualquier análisis posterior se realice únicamente con los casos seleccionados.

Ponderar casos permite asignar un peso o ponderación a cada caso según la variable de frecuencia que indiquemos. Es decir, esta variable debe indicar el número de observaciones únicas representadas por un único caso. Los casos con valores perdidos, negativos o cero para la variable de ponderación se excluyen del análisis, pero sí se computan los valores fraccionarios (en las tablas de contingencia se redondearía al entero más próximo). La variable de ponderación aplicada se mantiene activa hasta que se seleccione otra o se desactive la ponderación. Si se guarda el archivo, la información de ponderación se guardará con él.

2.11. Menú Transformar

La opción *Calcular* nos permite asignar a una variable nueva o ya existente una expresión numérica o alfanumérica que puede contener operaciones y funciones matemáticas preincorporadas (más de 70) con números y variables existentes. Es una forma inmediata de crear nuevas variables con valores constantes, pero también nos permite seleccionar los casos a los que aplicar la operación según las condiciones que establezcamos.

La semilla de aleatorización se establece a 2.000.000 cada vez que se inicia una sesión de SPSS. Esta semilla es utilizada por el generador de números pseudo-aleatorios para los procedimientos que lo precisen y se regenera cada vez que se necesita. La opción *Semilla de*

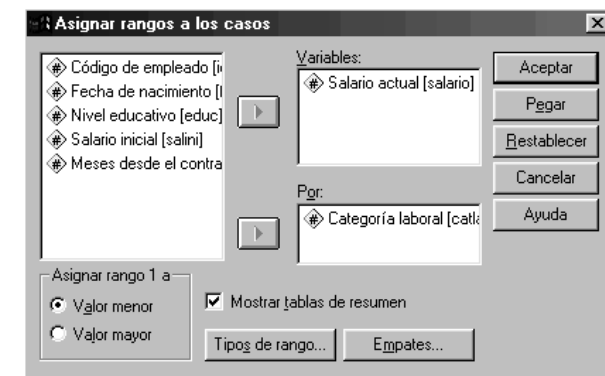
aleatorización nos permite establecer la semilla de forma que podamos repetir una secuencia previamente generada de números.

Contar apariciones, como su propio nombre indica, cuenta las veces que se repite un determinado valor o conjunto de valores en más de una variable. Dicho de otra forma, cuenta las veces que se repite un valor en un mismo caso. El cuadro de diálogo nos solicita la variable de destino (nueva o existente) donde se almacenará el número de veces que aparece el valor. Debemos indicar las variables donde contar las apariciones, definir el valor o rango de valores que se pretende contar y, si queremos, realizar una selección previa de casos según determinadas condiciones (*Si...*).

Recodificar, en las mismas o distintas variables, reasigna los valores de las variables existentes. Para una o más variables (del mismo tipo) debemos definir el valor o rango de valores antiguos y el cambio directo (ya que no hay operación intermedia) por otros valores nuevos. El cuadro de diálogo permite especificar los casos a tratar.

Categorizar variables convierte datos numéricos continuos en un número discreto de categorías que se almacenan en nuevas variables. Los datos se categorizan según grupos percentiles; por lo que cada grupo contiene aproximadamente el mismo número de valores diferentes. Por ejemplo, la especificación de cuatro grupos asignará el valor 1 a los casos situados bajo el percentil 25, 2 a los casos entre el percentil 25 y el 50, 3 a los casos situados entre el percentil 50 y el 75 y 4 a los casos por encima del percentil 75.

Asignar rangos a casos crea nuevas variables que contienen rangos, puntuaciones de Savage y normales, y los valores de los percentiles para las variables numéricas. Por defecto se crean tantos rangos como casos existen; el menor elemento sería el 1 y el mayor el número de casos; los empates se resuelven con la media de los valores del rango.



Los nombres de las nuevas variables así como sus etiquetas descriptivas se generan automáticamente en función del nombre de la variable original y de las medidas seleccionadas. Además se crea una tabla resumen con la lista de las variables originales, las

nuevas variables y las etiquetas de variable. Los rangos se pueden establecer en orden ascendente o descendente. Se puede indicar más de una variable, con lo que el rango se calcularía para cada combinación de las mismas. La lista *Por* define variables de agrupación con lo que el procedimiento se ejecutaría independientemente para cada grupo definido por la o las variables indicadas. Se puede establecer uno o varios *Tipos de rango* que crearán sus respectivas variables. Por último los empates pueden dirimirse con la media (por defecto), el mayor, el menor, o con valores secuenciales.

Recodificación automática convierte los valores numéricos y de cadena en valores enteros consecutivos. La recodificación se realiza sobre una nueva variable que conserva las etiquetas de valor de la variable antigua. Para los valores que no tienen una etiqueta de valor ya definida se utiliza el valor original como etiqueta del valor recodificado. Además se genera una tabla en el *Visor* que muestra los valores antiguos, los nuevos y las etiquetas de valor.

La recodificación puede realizarse en orden ascendente o descendente. Los valores de cadena se recodifican por orden alfabético, con las mayúsculas antes que las minúsculas. Los valores perdidos se recodifican como valores perdidos mayores que cualquier valor no perdido y conservando el orden. Por ejemplo, si la variable original posee 10 valores no perdidos, el valor perdido mínimo se recodificará como 11, y el valor 11 será un valor perdido para la nueva variable.

Crear serie temporal crea nuevas variables basadas en funciones de variables de series temporales numéricas existentes. Una serie temporal es una secuencia de valores de una variable donde cada caso representa una observación en un momento de tiempo distinto. Los valores obtenidos al crear una serie de este tipo son útiles en muchos procedimientos de análisis de series temporales. Existen varias funciones disponibles como la diferencia, medias móviles, medianas móviles, retardo y adelanto. Por ejemplo, la diferencia asignaría a cada caso de la nueva variable la diferencia entre el valor del mismo caso y del caso anterior de la variable original.

El sistema genera una tabla en el *Visor* con información sobre el procedimiento realizado. Los nombres de las nuevas variables se crean automáticamente en base a la variable utilizada para crearlas y conservan cualquier etiqueta de valor ya definida.

Algunas medidas de series temporales no se pueden calcular si hay valores perdidos en la serie y pueden causar problemas en determinados análisis que necesitan la existencia de todas las observaciones. Para solventar este problema tenemos el procedimiento *Reemplazar valores perdidos* que crea nuevas variables de series temporales a partir de las existentes pero reemplazando los valores perdidos por estimaciones calculadas. El método de estimación puede ser: la media de los valores existentes (por defecto) que asignaría la media obtenida a todos los valores perdidos; la media o la mediana de puntos adyacentes; la interpolación lineal; y la tendencia lineal en un punto.

De la misma forma que en el procedimiento anterior, el sistema genera una tabla de información, los nombres de las nuevas variables se crean automáticamente y se conserva cualquier etiqueta de valor definida.

Las transformaciones que hagamos se ejecutarán automáticamente salvo que hayamos indicado lo contrario mediante las opciones generales del programa. Una vez en este panel seleccionaríamos la pestaña *Datos* y marcaríamos la opción *Calcular los valores antes de*

usarlos. Ahora, las transformaciones generadas con los procedimientos *Calcular* o *Recodificar*, entre otros (en general los que no requieren una lectura diferente de los datos), se van acumulando sin ser ejecutadas hasta que se lo indiquemos mediante el comando *Ejecutar transformaciones pendientes*. Esto puede ser útil trabajando con archivos grandes para retrasar la ejecución y ganar tiempo de procesamiento.

2.12. Menú Utilidades

Cuando necesitemos información detallada sobre las variables disponemos de dos opciones:

- ♦ *Variables*: muestra un cuadro dividido en dos paneles, uno contiene la lista de variables y el otro presenta información sobre la que esté seleccionada.
- ♦ *Información del archivo*: muestra a través del *Visor* de resultados un informe más o menos completo sobre las variables que contiene el archivo de datos que estamos usando.

Si utilizamos una cantidad grande de variables puede ocurrir que los cuadros de diálogo se ralenticen en su apertura y que sea tedioso localizar y seleccionar las variables que queremos usar en cada procedimiento. Para evitar esto podemos definir conjuntos pequeños de variables que restrinjan las listas de origen de los cuadros. Esto incluso puede mejorar el rendimiento.

El cuadro *Definir conjuntos de variables* nos permite definir los nombres de los conjuntos y las variables que asignamos a cada uno. Los nombres tienen un máximo de 12 caracteres, incluidos espacios, y no distingue entre mayúsculas y minúsculas. Posteriormente, cuando queramos utilizarlos, iremos a la opción *Usar conjuntos* y los asignaremos.

Por defecto se utilizan dos conjuntos predefinidos por el sistema:

- ♦ *ALLVARIABLES*: contiene todas las variables del archivo de datos, incluidas las nuevas creadas durante la sesión.
- ♦ *NEWVARIABLES*: sólo contiene las creadas durante la sesión.

Estos conjuntos pueden quitarse de la lista y situar otros. En concreto debemos eliminar el de *ALLVARIABLES*, de lo contrario los conjuntos que adjuntemos no tendrán efecto.

A partir de ese momento, en las listas de variables sólo se visualizarán las que estén incluidas en algún conjunto en uso. El orden de las mismas en los conjuntos no se mantendrá en las listas.

3. ANÁLISIS

Dentro del menú *Analizar* encontramos todos los procedimientos de análisis estadístico (exceptuando los gráficos) de SPSS.

3.1. Informes

Cubos OLAP → (Procedimiento analítico interactivo). Crea una tabla con varias capas que contienen totales, medias y otros estadísticos univariados para variables de resumen continuas según una o más variables categóricas de agrupación. En la tabla se creará una nueva capa para cada categoría de cada variable de agrupación.

Resúmenes de casos → Calcula estadísticos de subgrupo para las variables dentro de las categorías de una o más variables de agrupación. Se cruzan todos los niveles de las variables de agrupación. Permite elegir el orden en el que se mostrarán los estadísticos. También se muestran estadísticos de resumen para cada variable a través de todas las categorías. Los valores de los datos en cada categoría pueden mostrarse en una lista o suprimirse. Con grandes conjuntos de datos se pueden listar sólo los primeros n casos.

Informe de estadísticos en filas → Genera informes en los cuales se presentan distintos estadísticos de resumen en filas. También se encuentran disponibles listados de los casos, con o sin estadísticos de resumen.

Columnas de datos son las variables del informe para obtener el listado de los casos o los estadísticos de resumen y controla el formato de presentación de las columnas de datos.

Romper columnas por muestra una lista de las variables de ruptura opcionales que dividen el informe en grupos y controla los estadísticos de resumen y los formatos de presentación de las columnas de ruptura.

Informe controla las características globales, incluyendo los estadísticos de resumen globales, la presentación de los valores perdidos, la numeración de las páginas y los títulos.

Mostrar casos muestra los valores reales (o etiquetas) de las variables de la columna de datos para cada caso.

Presentación preliminar muestra sólo la primera página del informe. Esto es útil para examinar previamente el formato sin procesar el informe completo.

Los datos ya están ordenados. Para los informes con variables de ruptura, los datos se deben ordenar por estas variables antes de generar el informe. Si los datos ya están ordenados, se puede ahorrar tiempo de proceso con esta opción.

Informe de estadísticos en columnas → Genera informes en los que diversos estadísticos de resumen aparecen en columnas distintas.

Las opciones son similares a las del procedimiento anterior.

3.2. Estadísticos descriptivos

Frecuencias → Proporciona estadísticos y representaciones gráficas útiles para describir o inspeccionar muchos tipos de variables. Los valores pueden organizarse en orden ascendente o descendente, y se pueden ordenar las categorías por sus frecuencias. Se puede suprimir el informe de frecuencias cuando una variable posee muchos valores diferentes, y permite etiquetar los gráficos con las frecuencias o con los porcentajes.

Descriptivos → Muestra estadísticos de resumen para varias variables en una única tabla y calcula valores tipificados (puntuaciones z). Las variables se pueden ordenar por el tamaño de sus medias (en orden ascendente o descendente), alfabéticamente o por el orden en el que se seleccionen (por defecto). Se pueden guardar las puntuaciones z , añadiéndose al *Editor de datos*, para su posterior utilización. Cuando las variables se registran en unidades de medida diferentes (que no tengan nada que ver), una transformación de puntuación z pondrá las variables en una escala común para una comparación visual más fácil.

Explorar → Genera estadísticos de resumen y representaciones gráficas, bien para todos los casos o bien de forma separada para grupos de casos. Este procedimiento se puede usar para inspeccionar los datos, identificar valores atípicos o extremos, descubrir discontinuidades, obtener descripciones, comprobar supuestos y caracterizar diferencias entre subpoblaciones (grupos de casos), etc. La exploración de los datos puede ayudar a determinar si son adecuadas las técnicas de análisis que se están empleando.

Tablas de contingencia → Crea tablas de clasificación doble y múltiple, y además proporciona una serie de pruebas y medidas de asociación para las tablas de doble clasificación. La estructura de la tabla y el hecho de que las categorías estén ordenadas o no determinan las pruebas o medidas que se utilizan.

Los estadísticos de tablas de contingencia y las medidas de asociación sólo se calculan para las tablas de doble clasificación. Si especifica una fila, una columna y un factor de capa (variable de control), se creará un panel de medidas y estadísticos asociados para cada valor del factor de capa (o una combinación de valores para dos o más variables de control).

3.3. Tablas personalizadas

Tablas básicas → Genera tablas que muestran estadísticos de clasificación cruzada y de subgrupo. Las variables de resumen seleccionadas se resumen en las casillas definidas por las variables de subgrupo. Para todas las variables resumidas se emplean los mismos estadísticos (la media por defecto). Si no existen variables de resumen, se muestran los recuentos. Los nombres de las variables de resumen se muestran a lo largo de la parte izquierda de la tabla. Se pueden definir casillas para mostrar subgrupos de casos empleando las combinaciones:

- ♦ *Hacia abajo* de la página (como filas distintas).
- ♦ *A través* de la página (como columnas distintas). La combinación de *Hacia abajo* y *A través* genera una presentación en forma de tabla de contingencia.

- ♦ Repartidas en *Tablas distintas*. Las variables de esta lista subdividen la tabla en capas o en grupos de capas, de manera que sólo puede verse una capa de la tabla cada vez. Una vez en el *Visor*, para ver las otras capas se deben pivotar.

Anidar significa que se muestran las categorías de una variable bajo cada una de las categorías de la variable anterior. La apilación muestra las categorías de cada variable como un bloque; se puede entender como la extracción de tablas diferentes y su unión en la misma presentación.

Tablas generales → Genera tablas con estadísticos de clasificación cruzada y de subgrupo. Se pueden generar tablas que muestren diferentes estadísticos para distintas variables, variables de respuestas múltiples, anidación y apilación mixta o totales complejos.

En la mayoría de las circunstancias es necesario seleccionar una variable de fila. También se puede seleccionar una o más variables para definir columnas o capas. Por ejemplo, para generar una tabla de contingencia simple debemos seleccionar una variable para definir las filas y otra para definir las columnas. Si tiene una variable de capa, también es necesario tener una variable de fila y una de columna.

Cada variable seleccionada se debe especificar si es de agrupación para definir casillas o si es una variable de resumen. También se pueden especificar diferentes estadísticos para cada variable. Si seleccionamos más de una variable para una dimensión dada, deberemos especificar si deben estar apiladas o anidadas en la dimensión.

La mayoría de las tablas que muestran los mismos estadísticos de resumen para cada variable se pueden obtener más fácilmente utilizando el procedimiento *Tablas básicas o de respuestas múltiples*.

Se pueden mostrar los estadísticos (*Las etiquetas de los estadísticos aparecen*) en la parte superior (columnas), en el lateral (filas) o en capas separadas.

Para cada variable seleccionada, se puede especificar si es de agrupación o de resumen. Los estadísticos asociados a una variable que define casillas son los recuentos y porcentajes. La primera variable de resumen que se especifique definirá la dimensión de resumen. Las palabras *Dimensión resumen* aparecerán sobre la lista de variables para esa dimensión, por lo que no se podrán resumir variables en otra dimensión.

Los estadísticos seleccionados en la *edición de estadísticos* se aplican sólo a la variable seleccionada. La primera vez que se editen los estadísticos de una variable, se definirá la dimensión de los estadísticos.

Insertar total inserta una variable de total después de la variable seleccionada. No se puede obtener un total de otro total o de una variable que tiene un total anidado debajo de ella.

Respuestas múltiples muestra una lista de los conjuntos de respuestas múltiples definidos y permite definir otros.

Tablas de respuestas múltiples → Crea tablas de frecuencia y de contingencia básicas en las que una o más variables es un conjunto de respuestas múltiples (si no, sería conveniente utilizar *Tablas básicas*).

Si no seleccionamos la opción *Anidar* para una dimensión, se producirán tablas separadas para cada variable pero conectadas físicamente (concatenadas).

Las restantes opciones son similares a las descritas anteriormente.

Tablas de frecuencias → Genera tablas especiales que contienen varias variables con los mismos valores. Por defecto, las variables forman columnas y las categorías filas (todas las variables de *Frecuencias para* deben tener las mismas categorías). Cada casilla muestra el número de casos en esa categoría. Permite variables de subgrupo.

Mediante *Subgrupos* se puede dividir la tabla en columnas dentro de cada tabla o en tablas diferentes (capas). Al dividirla se mostrarán todas las variables para cada subgrupo.

Si hay diversas variables en una de las listas de la sección *Subgrupos*, se puede elegir entre anidarlas o apilarlas (se crean las tablas por separado pero se unen en la presentación).

3.4. Comparar medias

Medias → Calcula medias de subgrupo y estadísticos univariados relacionados para variables dependientes dentro de las categorías de una o más variables independientes. Pueden obtenerse el análisis de varianza de un factor, la *t* y pruebas de linealidad.

Prueba T para una muestra → Contrasta si la media de una sola variable difiere de una constante especificada.

Prueba T para muestras independientes → Compara las medias de dos grupos de casos. Los sujetos deben asignarse aleatoriamente a dos grupos, de forma que cualquier diferencia en la respuesta sea debida al tratamiento (o falta de tratamiento) y no a otros factores. Cuando esto no es posible (por ejemplo comparando una variable entre hombres y mujeres ya que el sexo de una persona no se asigna aleatoriamente), debemos asegurarnos de que las diferencias en otros factores no enmascaren o resalten una diferencia significativa entre las medias (por ejemplo los estudios respecto al sexo).

Prueba T para muestras relacionadas → Compara las medias de dos variables de un solo grupo. Calcula las diferencias entre los valores de las dos variables de cada caso y contrasta si la media difiere de 0.

ANOVA de un factor → Genera un análisis de varianza de un factor para una variable dependiente cuantitativa respecto a una única variable de factor (la variable independiente). El análisis de varianza se utiliza para contrastar la hipótesis de que varias medias son iguales. Esta técnica es una extensión de la prueba *t* para dos muestras. Además de determinar que existen diferencias entre las medias, es posible que se desee saber qué medias difieren. Existen dos tipos de contrastes para comparar medias: los contrastes *a priori* y las pruebas *post hoc*. Los contrastes *a priori* se plantean antes de ejecutar el experimento y las pruebas *post hoc* se realizan después de haber llevado a cabo el experimento. También se pueden contrastar las tendencias existentes a través de las categorías.

3.5. Modelo lineal general

Univariante → Proporciona un análisis de regresión y un análisis de varianza para una variable dependiente mediante uno o más factores o variables. Las variables de factor

dividen la población en grupos. Con este procedimiento se pueden contrastar hipótesis nulas sobre los efectos de otras variables en las medias de varias agrupaciones de una única variable dependiente. Se pueden investigar las interacciones entre los factores así como los efectos de los factores individuales, algunos de los cuales pueden ser aleatorios. Además, se pueden incluir los efectos de las covariables y las interacciones de covariables con los factores. Para el análisis de regresión, las variables independientes (predictoras) se especifican como covariables.

Se pueden contrastar tanto los modelos equilibrados como los no equilibrados. Se considera que un diseño está equilibrado si cada casilla del modelo contiene el mismo número de casos. Además de contrastar hipótesis se generan estimaciones de los parámetros.

También se encuentran disponibles los contrastes de hipótesis de uso más habitual. Además, si una prueba F global ha mostrado cierta significación, pueden emplearse las pruebas post hoc para evaluar las diferencias entre las medias específicas. Las medias marginales estimadas ofrecen estimaciones de valores de las medias pronosticados para las casillas del modelo; los gráficos de perfil (gráficos de interacciones) de estas medias permiten observar fácilmente algunas de estas relaciones.

En el archivo de datos podemos guardar residuos, valores pronosticados, distancia de Cook y valores de influencia como variables nuevas para comprobar los supuestos.

Ponderación MCP permite especificar una variable usada para aplicar a las observaciones una ponderación diferente en un análisis de mínimos cuadrados ponderados (MCP), por ejemplo para compensar la distinta precisión de las medidas.

Métodos. Las sumas de cuadrados de Tipo I, Tipo II, Tipo III y Tipo IV pueden emplearse para evaluar las diferentes hipótesis. El Tipo III es el valor por defecto.

Gráficos. Diagramas de dispersión por nivel, gráficos de residuos y gráficos de perfil (interacción).

Multivariante → Proporciona un análisis de regresión y un análisis de varianza para variables dependientes múltiples por una o más covariables o variables de factor. Las variables de factor dividen la población en grupos. Con este procedimiento es posible contrastar hipótesis nulas sobre los efectos de las variables de factor sobre las medias de varias agrupaciones de una distribución conjunta de variables dependientes.

En un modelo multivariado, las sumas de cuadrados debidas a los efectos del modelo y las sumas de cuadrados error se encuentran en forma de matriz en lugar de en la forma escalar del análisis univariado. Estas matrices se denominan matrices SCPC (sumas de cuadrados y productos cruzados). Si se especifica más de una variable dependiente, se proporciona el análisis multivariado de varianzas usando la traza de Pillai, la lambda de Wilks, la traza de Hotelling y el criterio de mayor raíz de Roy con el estadístico F aproximado, así como el análisis univariado de varianza para cada variable dependiente.

Las pruebas de comparaciones múltiples post hoc se realizan por separado para cada variable dependiente.

También se hallan disponibles: una matriz SCPC residual, que es una matriz cuadrada de las sumas de cuadrados y los productos cruzados de los residuos; una matriz de covarianza residual, que es la matriz SCPC residual dividida por los grados de libertad de los residuos; y la matriz de correlaciones residual, que es la forma tipificada de la matriz de covarianza residual.

El resto es similar al procedimiento anterior.

Medidas repetidas → Analiza grupos de variables dependientes relacionadas que representan diferentes medidas del mismo atributo. Permite definir uno o varios factores intra-sujetos (no confundir con las variables existentes), donde cada factor constituye un nivel dentro del factor precedente.

Si los sujetos se comparan en más de una medida cada vez, hemos de seleccionar *Medida* para definirlos. El nombre de las medidas no existe como un nombre de variable en el propio archivo de datos sino que se define aquí.

Componentes de la varianza → Estima la contribución de cada efecto aleatorio, para modelos de efectos mixtos, a la varianza de la variable dependiente. Este procedimiento resulta de interés para el análisis de modelos mixtos, como los diseños split-plot, los diseños de medidas repetidas univariados y los diseños de bloques aleatorios. Al calcular las componentes de la varianza, se puede determinar dónde centrar la atención para reducir la varianza.

Se dispone de cuatro métodos diferentes para estimar las componentes de la varianza:

- ♦ Estimador mínimo no cuadrático insesgado (EMNCI, MINQUE).
- ♦ Análisis de varianza (ANOVA).
- ♦ Máxima verosimilitud (MV, ML).
- ♦ Máxima verosimilitud restringida (MVR, RML).

Además, existen diversas especificaciones para los diferentes métodos.

Los resultados por defecto para todos los métodos incluyen las estimaciones de componentes de la varianza. Si se usa el método MV o el método MVR, se mostrará también una tabla con la matriz de covarianza asíntótica. Otros resultados disponibles incluyen una tabla de ANOVA y las medias cuadráticas esperadas para el método ANOVA, y la historia de iteraciones para los métodos MV y MVR. Este procedimiento es totalmente compatible con el procedimiento MLG (Modelo Lineal General) Factorial general.

3.6. Correlaciones

Bivariadas → Calcula el coeficiente de correlación de Pearson, la rho de Spearman y la tau-b de Kendall con sus niveles de significación. Las correlaciones miden cómo están relacionadas las variables o los órdenes de los rangos. Antes de calcular un coeficiente de correlación, se deben inspeccionar los datos para detectar valores atípicos (que pueden producir resultados equívocos) y evidencias de una relación lineal. Las correlaciones son medidas de asociación lineal. Dos variables pueden estar perfectamente relacionadas, pero si la relación no es lineal, el coeficiente de correlación no es un estadístico adecuado para medir su asociación.

Parciales → Calcula los coeficientes de correlación parcial, los cuales describen la relación lineal existente entre dos variables mientras se controlan los efectos de una o más variables adicionales.

Distancias → Calcula una gran variedad de estadísticos midiendo las similitudes o las disimilitudes (distancias), bien entre pares de variables o entre pares de casos. Estas medidas de similitud o de distancia se pueden emplear a continuación con otros procedimientos como el análisis factorial, el análisis de conglomerados o el escalamiento multidimensional para ayudar a analizar conjuntos de datos complejos.

3.7. Regresión

Lineal → Estima los coeficientes de la ecuación lineal, con una o más variables independientes, que mejor prediga el valor de la variable dependiente.

Estimación curvilínea → Genera estadísticos de estimación curvilínea por regresión y gráficos relacionados para 11 modelos diferentes. Se produce un modelo diferente para cada variable dependiente. También se pueden guardar valores pronosticados, residuos e intervalos pronosticados como nuevas variables.

Logística binaria → Es útil cuando se desea predecir la presencia o ausencia de una característica o resultado según los valores de un conjunto de variables predictoras. Es similar a un modelo de regresión lineal pero está adaptado para modelos en los que la variable dependiente es dicotómica. Los coeficientes pueden utilizarse para estimar la razón de las ventajas (odds ratio) de cada variable independiente del modelo. La regresión logística se puede aplicar a un rango más amplio de situaciones de investigación que el análisis discriminante.

Se pueden estimar modelos utilizando la entrada en bloque de las variables o cualquiera de los siguientes métodos por pasos:

- ◆ Condicional hacia adelante.
- ◆ LR hacia adelante.
- ◆ Wald hacia adelante.
- ◆ Condicional hacia atrás.
- ◆ LR hacia atrás.
- ◆ Wald hacia atrás.

Logística multinomial → Resulta útil si se desea clasificar a los sujetos según los valores de un conjunto de variables predictoras. Este tipo de regresión es similar a la regresión logística, pero más general, ya que la variable dependiente no está restringida a dos categorías.

Se ajusta un modelo logit multinomial para el modelo factorial completo o para un modelo especificado por el usuario. La estimación de los parámetros se realiza a través de un algoritmo iterativo de máxima verosimilitud.

Ordinal → Permite dar forma a la dependencia de una respuesta ordinal politémica sobre un conjunto de predictores, que pueden ser factores o covariables. El diseño de la regresión ordinal se basa en la metodología de McCullagh y en la sintaxis se hace referencia al procedimiento como PLUM.

Probit → Mide la relación entre la intensidad de un estímulo y la proporción de casos que presentan una cierta respuesta a dicho estímulo. Es útil para las situaciones en las que se dispone de una respuesta dicotómica que se piensa puede estar influenciada o causada por los niveles de alguna o algunas variables independientes, y es particularmente adecuada para datos experimentales. Este procedimiento permite estimar la intensidad necesaria para que un estímulo llegue a inducir una determinada proporción de respuestas, como la dosis efectiva para la mediana.

No lineal → Sirve para encontrar un modelo no lineal para la relación entre la variable dependiente y un conjunto de variables independientes. A diferencia de la regresión lineal tradicional, que está restringida a la estimación de modelos lineales, la regresión no lineal puede estimar modelos con relaciones arbitrarias entre las variables independientes y la dependiente. Esto se lleva a cabo usando algoritmos de estimación iterativos. Este procedimiento no es necesario para modelos polinómicos simples que se pueden estimar usando métodos tradicionales como la *Regresión lineal*.

Estimación ponderada → Los modelos de regresión lineal típicos asumen que la varianza es constante en la población objeto de estudio. Cuando éste no es el caso (por ejemplo cuando los casos con puntuaciones mayores en un atributo muestran más variabilidad que los casos con puntuaciones menores en ese atributo), la regresión lineal mediante mínimos cuadrados ordinarios (MCO, OLS) deja de proporcionar estimaciones óptimas para el modelo. Si las diferencias de variabilidad se pueden pronosticar a partir de otra variable, el procedimiento *Estimación ponderada* permite calcular los coeficientes de un modelo de regresión lineal mediante mínimos cuadrados ponderados (MCP, WLS), de forma que se les dé mayor ponderación a las observaciones más precisas (es decir, aquellas con menos variabilidad) al determinar los coeficientes de regresión. El procedimiento *Estimación ponderada* contrasta un rango de transformaciones de ponderación e indica cuál se ajustará mejor a los datos.

Mínimos cuadrados en dos fases → Los modelos de regresión lineal típica asumen que los errores de la variable dependiente no están correlacionados con la variable o variables independientes. Cuando éste no es el caso (por ejemplo, cuando las relaciones entre las variables son bidireccionales), la regresión lineal mediante mínimos cuadrados ordinarios (OLS) deja de proporcionar estimaciones óptimas del modelo. La regresión por mínimos cuadrados en dos fases utiliza variables instrumentales que no estén correlacionadas con los términos de error para calcular los valores estimados de los predictores problemáticos (en la primera fase) y después utiliza dichos valores calculados para estimar un modelo de regresión lineal para la variable dependiente (la segunda fase). Dado que los valores calculados se basan en variables que no están correlacionadas con los errores, los resultados del modelo en dos fases son óptimos.

Escalamiento óptimo → Cuantifica los datos categóricos mediante la asignación de valores numéricos a las categorías, obteniéndose una ecuación de regresión lineal óptima para las variables transformadas.

El análisis de regresión lineal ordinario implica minimizar las diferencias de la suma de los cuadrados entre una variable de respuesta (la dependiente) y una combinación

ponderada de las variables predictoras (las independientes). Las variables son normalmente cuantitativas, con los datos categóricos (nominales) recodificados como variables binarias o de contraste. Como resultado, las variables categóricas sirven para separar grupos de casos y la técnica estima conjuntos separados de parámetros para cada grupo. Los coeficientes estimados reflejan cómo los cambios en los predictores afectan a la respuesta. El pronóstico de la respuesta es posible para cualquier combinación de los valores predictores.

Una aproximación alternativa incluye la regresión de la respuesta respecto a los propios valores predictores categóricos. Como consecuencia, se estima un coeficiente para cada variable. Sin embargo, para las variables categóricas, los valores categóricos son arbitrarios. La codificación de las categorías de diferentes maneras proporciona diferentes coeficientes, dificultando las comparaciones entre los análisis de las mismas variables.

CATREG (regresión categórica mediante escalamiento óptimo) amplía la aproximación típica mediante un escalamiento de las variables nominales, ordinales y numéricas simultáneamente. El procedimiento cuantifica las variables categóricas de manera que las cuantificaciones reflejen las características de las categorías originales. Además, trata a las variables categóricas cuantificadas como si fueran variables numéricas. La utilización de transformaciones no lineales permite a las variables ser analizadas en varios niveles para encontrar el modelo que más se ajusta.

3.8. Loglineal

General → Analiza las frecuencias de las observaciones incluidas en cada categoría de la clasificación cruzada de una tabla de contingencia. Cada una de las clasificaciones cruzadas de la tabla constituye una casilla y cada variable categórica se denomina factor. La variable dependiente es el número de casos (la frecuencia) en una casilla de la tabla de contingencia, y las variables explicativas son los factores y las covariables. Este procedimiento estima los parámetros de máxima verosimilitud de modelos loglineales jerárquicos y no jerárquicos utilizando el método de Newton-Raphson. Es posible analizar una distribución multinomial o de Poisson.

Se pueden seleccionar hasta 10 factores para definir las casillas de una tabla. Una variable de estructura de casilla permite definir ceros estructurales para tablas incompletas, incluir en el modelo un término de desplazamiento, ajustar un modelo log-tasa o implementar el método de corrección de las tablas marginales. Las variables de contraste permiten el cálculo del logaritmo de la razón de ventajas generalizadas (GLOR).

SPSS muestra automáticamente información sobre el modelo y estadísticos de bondad de ajuste. Además es posible mostrar una variedad de estadísticos y gráficos, o guardar los valores pronosticados y los residuos en el archivo de datos de trabajo.

Logit → Analiza la relación entre variables dependientes (o de respuesta) y variables independientes (o explicativas). Las variables dependientes siempre son categóricas, mientras que las variables independientes pueden ser categóricas (factores). Otras variables independientes, las covariables de casilla, pueden ser continuas pero no se aplican en forma de caso por caso. A una casilla dada se le aplica la media ponderada de la covariable para los casos de esa casilla. El logaritmo de las ventajas de las variables dependientes se expresa como una combinación lineal de parámetros. Se supone automáticamente una distribución multinomial; estos modelos se denominan a veces modelos logit multinomiales. Este procedimiento estima los parámetros de los modelos loglineales logit utilizando el algoritmo de Newton-Raphson.

Selección de modelo → Analiza tablas de contingencia de varios factores. Ajusta modelos loglineales jerárquicos a las tablas de contingencia multidimensionales utilizando un algoritmo de ajuste proporcional. Ayuda a encontrar cuáles de las variables categóricas están asociadas. Para construir los modelos se encuentran disponibles métodos de entrada forzada y de eliminación hacia atrás. Para los modelos saturados, es posible solicitar estimaciones de los parámetros y pruebas de asociación parcial. Un modelo saturado añade 0,5 a todas las casillas.

3.9. Clasificar

Conglomerados de K medias → Intenta identificar grupos de casos relativamente homogéneos basándose en las características seleccionadas y utilizando un algoritmo que puede gestionar un gran número de casos. Sin embargo, el algoritmo requiere que el usuario especifique el número de conglomerados. Se pueden especificar los centros iniciales de los conglomerados si se conoce de antemano dicha información. Se pueden elegir uno de los dos métodos disponibles para clasificar los casos: la actualización de los centros de los conglomerados de forma iterativa o sólo la clasificación. Asimismo, se puede guardar la pertenencia a los conglomerados, información de la distancia y los centros de los conglomerados finales. Permite especificar una variable cuyos valores sean utilizados para etiquetar los resultados por casos. También se pueden solicitar los estadísticos F de los análisis de varianza. Aunque estos estadísticos son oportunistas (ya que el procedimiento trata de formar grupos que de hecho difieran), el tamaño relativo de los estadísticos proporciona información acerca de la contribución de cada variable a la separación de los grupos.

Conglomerados jerárquicos → Intenta identificar grupos relativamente homogéneos de casos (o de variables) basándose en las características seleccionadas, mediante un algoritmo que comienza con cada caso (o cada variable) en un conglomerado diferente y combina los conglomerados hasta que sólo queda uno. Es posible analizar las variables brutas o elegir de entre una variedad de transformaciones de estandarización. Las medidas de distancia o similitud se generan mediante el procedimiento *Proximidades*. Los estadísticos se muestran en cada etapa para ayudar a seleccionar la mejor solución.

Discriminante → Resulta útil para las situaciones en las que se desea construir un modelo predictivo para pronosticar el grupo de pertenencia de un caso a partir de las características observadas de cada caso. El procedimiento genera una función discriminante (o, para más de dos grupos, un conjunto de funciones discriminantes) basada en combinaciones lineales de las variables predictoras que proporcionan la mejor discriminación posible entre los grupos. Las funciones se generan a partir de una muestra de casos para los que se conoce el grupo de pertenencia; posteriormente, las funciones pueden ser aplicadas a nuevos casos que dispongan de medidas para las variables predictoras pero de los que se desconozca el grupo de pertenencia.

La variable de agrupación puede tener más de dos valores. Los códigos de la variable de agrupación han de ser números enteros y es necesario especificar sus valores máximo y mínimo. Los casos con valores fuera de estos límites se excluyen del análisis.

3.10. Reducción de datos

Análisis factorial → Intenta identificar variables subyacentes, o factores, que expliquen la configuración de las correlaciones dentro de un conjunto de variables observadas. Se suele utilizar en la reducción de los datos para identificar un pequeño número de factores que explique la mayoría de la varianza observada en un número mayor de variables manifestadas. También puede utilizarse para generar hipótesis relacionadas con los mecanismos causales o para inspeccionar las variables para análisis subsiguientes (por ejemplo, para identificar la colinealidad antes de realizar un análisis de regresión lineal).

Este procedimiento dispone de 7 métodos de extracción factorial; 5 métodos de rotación, entre ellos el oblimin directo y el promax para rotaciones no ortogonales; y 3 métodos para calcular las puntuaciones factoriales, que pueden guardarse como variables para análisis adicionales.

Análisis de correspondencias → Describe las relaciones existentes entre dos variables nominales, recogidas en una tabla de correspondencias, sobre un espacio de pocas dimensiones, mientras que al mismo tiempo se describen las relaciones entre las categorías de cada variable. Para cada variable, las distancias sobre un gráfico entre los puntos de categorías reflejan las relaciones entre las categorías, con las categorías similares representadas próximas unas a otras. La proyección de los puntos de una variable sobre el vector desde el origen hasta un punto de categoría de la otra variable describe la relación entre ambas variables.

El análisis de las tablas de contingencia a menudo incluye examinar los perfiles de fila y de columna, así como contrastar la independencia a través del estadístico de chi-cuadrado. Sin embargo, el número de perfiles puede ser bastante grande y la prueba de chi-cuadrado no revelará la estructura de la dependencia. El procedimiento *Tablas de contingencia* ofrece varias medidas y pruebas de asociación pero no puede representar gráficamente ninguna relación entre las variables.

El análisis factorial es una técnica típica para describir las relaciones existentes entre variables en un espacio de pocas dimensiones. Sin embargo, requiere datos de intervalo y el número de observaciones debe ser cinco veces el número de variables. Por su parte, el análisis de correspondencias asume que las variables son nominales y permite describir las relaciones entre las categorías de cada variable, así como la relación entre las variables. Además, el análisis de correspondencias se puede utilizar para analizar cualquier tabla de medidas de correspondencia que sean positivas.

Escalamiento óptimo → La combinación del nivel de medida y el número de conjunto seleccionado determina este procedimiento mediante mínimos cuadrados alternantes que realiza.

Nivel de medida. Permite especificar el nivel correspondiente a las variables utilizadas en el análisis:

- ♦ *Todas las variables son nominales múltiples.* Todas las variables tienen cuantificaciones de categorías que pueden diferir para cada dimensión.
- ♦ *Alguna variable no es nominal múltiple.* Una o más variables se escalan a un nivel diferente del nominal múltiple. Otros niveles de escala posibles son: nominal simple, ordinal y numérica discreta.

Número de conjuntos de variables. Permite especificar cuántos grupos de variables se van a comparar con otros grupos de variables:

- ♦ *Un conjunto.* Los datos contienen un grupo de variables.
- ♦ *Múltiples conjuntos.* Los datos contienen más de un grupo de variables. Si se selecciona esta opción, se elegirá Correlación canónica no-lineal (OVERALS).

Análisis seleccionado. Las opciones para *Nivel de medida* y *Número de conjuntos de variables* deciden el análisis según las combinaciones.

- ♦ *Análisis de homogeneidad (HOMALS).* Seleccionando *Todas las variables son nominales múltiples* y *Un conjunto*.
- ♦ *Análisis de componentes principales categóricos (CATPCA).* Seleccionando *Alguna variable no es nominal múltiple* y *Un conjunto*.
- ♦ *Análisis de correlación canónica no lineal (OVERALS).* Seleccionando *Múltiples conjuntos*.

3.11. Escalas

Análisis de fiabilidad → Permite estudiar las propiedades de las escalas de medición y de los elementos que las constituyen. Calcula un número de medidas de fiabilidad de escala que se utilizan normalmente y también proporciona información sobre las relaciones entre elementos individuales de la escala. Se pueden utilizar los coeficientes de correlación intraclass para calcular estimaciones de la fiabilidad inter-evaluadores.

Modelos de fiabilidad disponibles:

- ♦ *Alfa (Cronbach).* Modelo de consistencia interna basado en la correlación inter-elementos promedio.
- ♦ *Dos mitades.* Divide la escala en dos partes y examina la correlación entre dichas partes.
- ♦ *Guttman.* Calcula los límites inferiores de Guttman para la fiabilidad verdadera.
- ♦ *Paralelo.* Asume que todos los elementos tienen varianzas iguales y varianzas error iguales a través de las réplicas.
- ♦ *Paralelo estricto.* Asume los supuestos del modelo paralelo y también asume que las medias son iguales a través de los elementos.

Escalamiento multidimensional → Trata de encontrar la estructura de un conjunto de medidas de distancia entre objetos o casos. Esto se logra asignando las observaciones a posiciones específicas en un espacio conceptual (normalmente de dos o tres dimensiones) de modo que las distancias entre los puntos en el espacio concuerden al máximo con las disimilaridades dadas. En muchos casos, las dimensiones de este espacio conceptual son interpretables y se pueden utilizar para comprender mejor los datos. Si las variables se han medido objetivamente, se puede utilizar como técnica de reducción de datos (permitirá calcular las distancias a partir de los datos multivariados, si es necesario). El escalamiento multidimensional puede también aplicarse a valoraciones subjetivas de disimilaridad entre objetos o conceptos. Además, puede tratar datos de disimilaridad procedentes de múltiples fuentes, como podrían ser múltiples evaluadores o múltiples sujetos evaluados por un cuestionario.

Escalamiento multidimensional (PROXSCAL) → Trata de encontrar la estructura existente en un conjunto de medidas de proximidades entre objetos. Esto se logra asignando las observaciones a posiciones específicas en un espacio conceptual de pocas dimensiones, de modo que las distancias entre los puntos en el espacio concuerden al máximo con las similitudes (o disimilitudes) dadas. El resultado es una representación de mínimos cuadrados de los objetos en dicho espacio de pocas dimensiones que, en muchos casos, ayuda a entender mejor los datos.

3.12. Pruebas no paramétricas

Chi-cuadrado → Tabula una variable en categorías y calcula un estadístico de chi-cuadrado. Esta prueba de bondad de ajuste compara las frecuencias observadas y esperadas en cada categoría para contrastar si todas las categorías contienen la misma proporción de valores o si cada categoría contiene una proporción de valores especificada por el usuario.

Binomial → Compara las frecuencias observadas de las dos categorías de una variable dicotómica con las frecuencias esperadas en una distribución binomial con un parámetro de probabilidad especificado. Por defecto, el parámetro de probabilidad para ambos grupos es 0,5. Para cambiar las probabilidades, puede introducirse una proporción de prueba para el primer grupo. La probabilidad del segundo grupo será 1 menos la probabilidad especificada para el primer grupo.

Rachas → Contrasta si es aleatorio el orden de aparición de dos valores de una variable. Una racha es una secuencia de observaciones similares. Una muestra con un número excesivamente grande o pequeño de rachas sugiere que la muestra no es aleatoria.

K-S de 1 muestra → La prueba de Kolmogorov-Smirnov para una muestra compara la función de distribución acumulada observada de una variable con una distribución teórica determinada, que puede ser la normal, la uniforme, la de Poisson o la exponencial. La Z de Kolmogorov-Smirnov se calcula a partir de la diferencia mayor (en valor absoluto) entre las funciones de distribución acumuladas teórica y observada. Esta prueba de bondad de ajuste contrasta si las observaciones podrían razonablemente proceder de la distribución especificada.

2 muestras independientes → Compara dos grupos de casos existentes en una variable.

K muestras independientes → Compara dos o más grupos de casos respecto a una variable.

2 muestras relacionadas → Compara las distribuciones de dos variables.

K muestras relacionadas → Compara las distribuciones de dos o más variables.

3.13. Series temporales

Suavizado exponencial → Suaviza componentes irregulares de datos de series temporales, para ello hace uso de una variedad de modelos que incorporan diferentes supuestos acerca de la tendencia y la estacionalidad.

Seleccionando un modelo estacional y definiendo la periodicidad mediante *Definir fechas*, se puede seleccionar una variable que represente los factores estacionales.

Los modelos difieren en los componentes estacionales y de tendencia. Se dispone de los modelos Simple, Holt, Winters y Personalizado.

Autorregresión → Estima un modelo de regresión lineal con errores autorregresivos de primer orden. Los métodos disponibles son: Máxima verosimilitud exacta, Cochrane-Orcutt y Prais-Winsten. Se debe desactivar *Incluir constante en el modelo* si no se desea estimar un término constante en la ecuación de regresión.

ARIMA → Estima modelos Arima (Modelo Autorregresivo Integrado de Media Móvil) univariados estacionales y no estacionales (también conocidos como modelos "Box-Jenkins").

Se pueden transformar las series antes de la estimación mediante una de las opciones de la lista *Transformar*:

- ♦ *Ninguna*. No se lleva a cabo ninguna transformación.
- ♦ *Log natural*. Transforma las series antes de la estimación con el logaritmo en base e.
- ♦ *Log base 10*. Transforma las series antes de la estimación con el logaritmo en base 10.

El grupo *Modelo* permite especificar los tres parámetros del modelo Arima: *Autorregresivo*, *Diferenciación* y *Media móvil*. Estos parámetros se conocen de forma más común como p, d y q, respectivamente. También se pueden definir los parámetros estacionales correspondientes, si se tiene un modelo estacional y se ha definido la estacionalidad de los datos en *Definir fechas*.

Además, si se ha definido la periodicidad (que se muestra en la parte inferior del cuadro), se pueden especificar los parámetros estacionales correspondientes introduciendo los valores en los cuadros sp, sd y sq estacionales.

Descomposición estacional → Estima factores estacionales multiplicativos o aditivos para las series temporales. Precisa que esté definida la periodicidad mediante *Definir fechas*. Permite seleccionar el tipo de modelo utilizado para la descomposición estacional: Multiplicativo o Aditivo.

Las opciones de *Ponderación de la media móvil* permiten especificar la manera de tratar las series en el cálculo de medias móviles. Estas opciones sólo están disponibles si la periodicidad de la serie es par (si es impar, todos los puntos son ponderados por igual):

- ♦ *Todos los puntos por igual* calcula las medias móviles con una amplitud igual a la periodicidad y todos los puntos ponderados por igual.

- ♦ *Puntos finales ponderados por .5* calcula las medias móviles con una amplitud igual a la periodicidad más uno y con los puntos finales ponderados por 0,5.

Mediante *Mostrar el listado por casos* se imprime un listado por casos que contiene un resumen en una línea en cada iteración, así como los estadísticos finales.

3.14. Supervivencia

Tablas de mortalidad → Existen muchas situaciones en las se desea examinar la distribución de un periodo entre dos eventos, como la duración del empleo (tiempo transcurrido entre el contrato y el abandono de la empresa). Sin embargo, este tipo de datos suele incluir algunos casos para los que no se registra el segundo evento (censurados); por ejemplo, la gente que todavía trabaja en la empresa al final del estudio. Las razones para que no se verifique el segundo evento pueden ser muy variadas: en algunos casos, el evento simplemente no tiene lugar antes de que finalice el estudio; en otros, el investigador puede haber perdido el seguimiento de su estado en algún momento anterior a que finalice el estudio; y existen además casos que no pueden continuar por razones ajenas al estudio (como el caso en que un empleado caiga enfermo y se acoja a una baja laboral). Estos casos se conocen globalmente como casos censurados y hacen que el uso de técnicas tradicionales como las pruebas t o la regresión lineal sea inapropiado para este tipo de estudio.

Existe una técnica estadística útil para este tipo de datos llamada tabla de mortalidad de “seguimiento”. La idea básica de la tabla de mortalidad es subdividir el periodo de observación en intervalos de tiempo más pequeños. En cada intervalo, se utiliza toda la gente que se ha observado como mínimo durante ese periodo de tiempo para calcular la probabilidad de que un evento terminal tenga lugar dentro de ese intervalo. Las probabilidades estimadas para cada intervalo se utilizan para estimar la probabilidad global de que el evento tenga lugar en diferentes puntos temporales.

Kaplan-Meier → Es un método de estimación de modelos hasta el evento en presencia de casos censurados. Se basa en la estimación de las probabilidades condicionales en cada punto temporal cuando tiene lugar un evento y en tomar el límite del producto de esas probabilidades para estimar la tasa de supervivencia en cada punto temporal.

Regresión de Cox → Del mismo modo que los anteriores, la *Regresión de Cox* es un método para crear modelos para datos de tiempos de espera hasta un evento con casos censurados presentes. Sin embargo, la regresión de Cox permite incluir en los modelos variables predictoras (covariables). Por ejemplo, permite construir un modelo de la duración en el empleo como función del nivel educativo y de la categoría laboral. Gestiona los casos censurados correctamente y proporciona las estimaciones de los coeficientes para cada una de las covariables, permitiendo evaluar el impacto de múltiples covariables en el mismo modelo. Además, es posible utilizar este método para examinar el efecto de covariables continuas.

Cox con covariable dep. del tiempo → Existen ciertas situaciones en las que interesa calcular un modelo de regresión de Cox, pero no se cumple el supuesto de tasas de impacto proporcionales. Es decir, que las tasas de impacto cambian con el tiempo: los valores de una (o de varias) de las covariables son diferentes en los distintos puntos del tiempo. En esos

casos, es necesario utilizar un modelo de regresión de Cox extendido, que permita especificar covariables dependientes del tiempo ($T_COV_$).

Con el fin de analizar dicho modelo, se debe definir primero una covariable dependiente del tiempo. Para facilitar esta tarea contamos con una variable del sistema, llamada $T_$, que representa el tiempo. Se utiliza para definir covariables dependientes del tiempo empleando dos métodos generales:

- ♦ Para contrastar el supuesto de tasas de impacto proporcionales con respecto a una covariable particular, o para estimar un modelo de regresión de Cox extendido que permita impactos no proporcionales, se define la covariable dependiente del tiempo como una función de la variable de tiempo $T_$ y la covariable en cuestión.
- ♦ Algunas variables pueden tener valores distintos en periodos diferentes del tiempo, pero no están sistemáticamente relacionadas con el tiempo. En tales casos es necesario definir una covariable dependiente del tiempo segmentada, lo cual puede llevarse a cabo a partir de un conjunto de medidas usando las expresiones lógicas (1:verdadero, 0:falso).

Para las covariables dependientes del tiempo segmentadas, los casos en los que falte cualquiera de los valores quedarán eliminados del análisis. Por tanto, todos los casos deben tener valores para todos los puntos del tiempo medidos en la covariable, incluso para los puntos del tiempo posteriores a la eliminación del caso del conjunto bajo riesgo (ya sea por el evento o por la censura).

3.15. Respuestas múltiples

Definir conjuntos → Agrupa variables elementales en conjuntos de categorías múltiples y de dicotomías múltiples, para los que se pueden obtener tablas de frecuencias y tablas de contingencia. Se pueden definir hasta 20 conjuntos de respuestas múltiples. Cada conjunto debe tener un nombre exclusivo.

Las variables elementales se pueden codificar como dicotomías o categorías:

- ♦ *Dicotomías*. En *Valor contado* debe introducirse un entero. Cada variable que tenga al menos una aparición del valor contado se convierte en una categoría del conjunto de dicotomías múltiples.
- ♦ *Categorías*. Crear un conjunto de categorías múltiples con el mismo rango de valores que las variables que lo componen. Deben introducirse dos enteros como valores máximo y mínimo del rango para las categorías del conjunto de categorías múltiples. El procedimiento suma cada valor entero distinto en el rango inclusivo para todas las variables que lo componen. Las categorías vacías no se tabulan.

A cada conjunto de respuestas múltiples se le debe asignar un nombre exclusivo de hasta 7 caracteres. El procedimiento coloca delante del nombre asignado un signo dólar (\$). El nombre del conjunto de respuestas múltiples sólo se encuentra disponible para su uso en los procedimientos de respuestas múltiples.

Frecuencias → Produce tablas de frecuencias para conjuntos de respuestas múltiples. Es necesario definir previamente uno o más conjuntos de respuestas múltiples.

Para los conjuntos de dicotomías múltiples, los nombres de categorías que se muestran en los resultados proceden de etiquetas de variable definidas para variables elementales del grupo. Para los conjuntos de categorías múltiples, las etiquetas de categoría proceden de las etiquetas de valor de la primera variable del grupo. Si las categorías perdidas para la primera variable están presentes para otras variables del grupo, debe definirse una etiqueta de valor para las categorías perdidas.

Los casos con *Valores perdidos* se excluyen en base a tabla por tabla. Un caso se considera perdido para un conjunto de dicotomías (o categorías múltiples) sólo si ninguno de sus componentes contiene el valor (o tiene valores válidos en el rango). Se pueden excluir los casos con valores perdidos en cualquier variable de la tabulación del conjunto de dicotomías, o categorías, o ambas.

Tablas de contingencia → Presenta, en forma de tabla de contingencia, conjuntos de respuestas múltiples, variables elementales o una combinación. También permite obtener porcentajes de casilla basados en casos o respuestas, modificar la gestión de los valores perdidos u obtener tablas de contingencia emparejadas. Es necesario definir previamente uno o más conjuntos de respuestas múltiples.

Los nombres en los resultados se muestran de la misma forma que en el procedimiento anterior. Las etiquetas de categoría por columnas se muestran en tres líneas, con un máximo de 8 caracteres por línea. Para evitar la división de palabras, se pueden invertir los elementos de las filas y las columnas o volver a definir las etiquetas.

4. GRÁFICOS

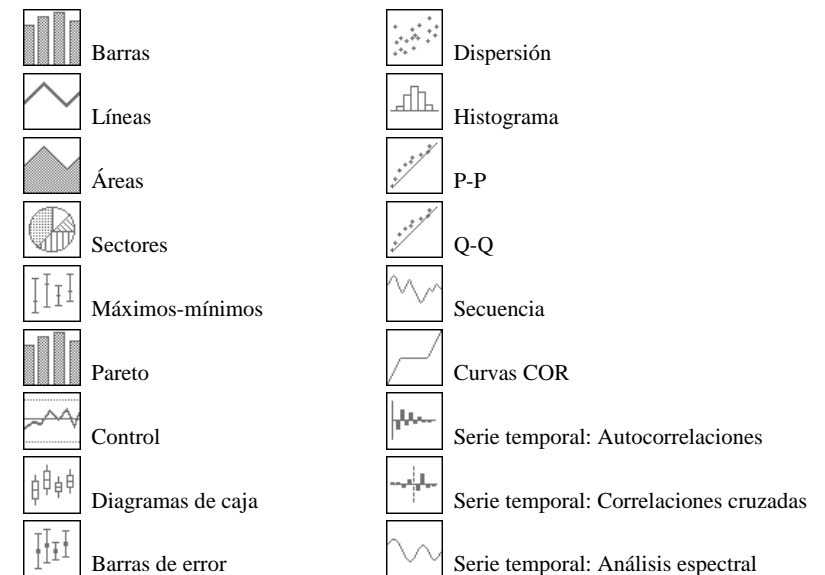
El menú *Gráficos* permite realizar una gran cantidad de gráficos orientados al análisis de los datos. La primera opción, *Galería*, no realiza un gráfico en sí, sino que nos traslada directamente a un tema de la ayuda que nos presenta una galería de gráficos.

La segunda opción es la de gráficos *Interactivos*. Con esta opción se generan gráficos que podrán ser modificados posteriormente, en cuanto a variables y elementos gráficos, de forma mucho más efectiva que los gráficos normales. Sin embargo sólo son convenientes cuando se trabaje con archivos de datos pequeños o subconjuntos de otros archivos mayores ya que los requerimientos de recursos son enormes.

Las restantes opciones de este menú son una lista de todos los gráficos posibles (no interactivos). Posteriormente se podrán modificar, pero sólo a nivel de formato y de una forma más precaria.

4.1. Galería de gráficos

La galería nos presenta todos los gráficos posibles:



Una vez seleccionado uno de los tipos, pasamos a otra pantalla de ayuda que nos muestra los subtipos existentes. De cada subtipo nos presenta un ejemplo de lo que se puede hacer y de cómo hacerlo.

4.2. Creación de un gráfico

Para la creación de un gráfico se nos muestra, generalmente, un cuadro de diálogo inicial de selección del gráfico deseado. Esto es debido a que la mayoría de tipos de gráfico están disponibles en varios subtipos y configuraciones. Ejemplo:



Una vez elegido el gráfico en cuestión, pasamos al cuadro de diálogo usual que aparece cuando realizamos un análisis.

Veamos con más detalle los tipos de gráficos disponibles.

4.2.1. Barras

Las posibilidades en los gráficos de barras son:



Simples: Las barras representan una función de la variable o variables.



Agrupadas: Las barras representan una función de la variable o variables dentro de categorías de otra variable.



Apiladas: Las barras representan una función de la variable o variables dentro de categorías de otra variable y se apilan por segmentos.

E, independientemente de cada posibilidad, tenemos:

- ♦ **Resúmenes para grupos de casos:** Resume categorías de una sola variable.
- ♦ **Resúmenes para distintas variables:** Resume categorías de 2 o más variables.
- ♦ **Valores individuales de los casos:** Resume una sola variable por cada caso.

Con lo cual tenemos 9 posibilidades:

Resúmenes para grupos de casos en gráficos de barras simples: Crea un gráfico que resume las categorías de una sola variable. La altura de la barra representa una función de la propia variable de categoría u otra variable de resumen.

Resúmenes para distintas variables en gráficos de barras simples: Crea un gráfico que resume dos o más variables. Hay una barra por cada variable.

Valores individuales de los casos en gráficos de barras simples: Crea un gráfico que resume una sola variable por cada caso.

Resúmenes para grupos de casos en gráficos de barras agrupadas: Crea un gráfico que resume categorías de una variable dentro de categorías de otra variable. La altura de la barra puede representar una función de la variable para la definición de las agrupaciones, o bien otra variable de resumen.

Resúmenes para distintas variables en gráficos de barras agrupadas: Crea un gráfico que resume dos o más variables dentro de categorías de otra variable.

Valores individuales de los casos en gráficos de barras agrupadas: Crea un gráfico que resume valores de dos o más variables por cada caso.

Resúmenes para grupos de casos en gráficos de barras apiladas: Crea un gráfico en el cual los segmentos de las barras se apilan uno encima de otro. Hay una barra por cada valor de la variable del eje de categorías. Los segmentos de cada barra representan las categorías de la variable de pila.

Resúmenes para distintas variables en gráficos de barras apiladas: Crea un gráfico que resume dos o más variables dentro de categorías.

Valores individuales de los casos en gráficos de barras apiladas: Crea un gráfico que resume dos o más variables para cada caso.

4.2.2. Líneas

Las posibilidades en los gráficos de líneas son:



Simples: Igual que los de barras.



Múltiples: Igual que los de barras agrupadas.



Líneas verticales: Igual que el anterior.

E independientemente de cada posibilidad, tenemos:

- ♦ **Resúmenes para grupos de casos**
- ♦ **Resúmenes para distintas variables**
- ♦ **Valores individuales de los casos**

Que son las mismas opciones que con los gráficos de barras ya comentados.

4.2.3. Áreas

Las posibilidades en los gráficos de áreas son:



Simples: Igual que los de barras.



Apiladas: Igual que los de barras agrupadas.

E, independientemente de cada posibilidad, tenemos:

- ♦ **Resúmenes para grupos de casos**
- ♦ **Resúmenes para distintas variables**
- ♦ **Valores individuales de los casos**

Que son las mismas opciones de antes.

4.2.4. Sectores



La buena percepción visual de los gráficos de sectores se basa en su sencillez, que por otra parte imposibilita el mostrar variables según categorías de otras variables. Con lo cual, no hay distintos gráficos posibles, sólo podemos determinar la estructura de los datos como ya hemos visto:

- ♦ **Resúmenes para grupos de casos**
- ♦ **Resúmenes para distintas variables**
- ♦ **Valores individuales de los casos**

4.2.5. Máximos y mínimos

Las posibilidades en los gráficos de máximos y mínimos son:



Máximo-mínimo-cierre simple: Crea un gráfico que resume una variable dicotómica (de 2 valores distintos) o tricotómica (de 3 valores distintos) en categorías de otra variable.



Máximo-mínimo-cierre agrupado: Crea un gráfico que resume dos o tres variables en agrupaciones definidas por una variable categórica. Cada barra de una agrupación se define mediante una segunda variable categórica.



Línea de diferencias: Crea un gráfico que resume una o dos variables dicotómicas en categorías de otra variable.



Barras de rango simples: Crea un gráfico que resume una o dos variables numéricas en categorías de otra variable.



Barras de rango agrupadas: Crea un gráfico que resume dos o más parejas de variables en agrupaciones definidas por una variable categórica. Cada barra de una agrupación representa una de las parejas de variables.

Para los máximos y mínimos tenemos las opciones ya vistas:

- ♦ **Resúmenes para grupos de casos**
- ♦ **Resúmenes para distintas variables**
- ♦ **Valores individuales de los casos**

4.2.6. Pareto

Tenemos 2 posibilidades gráficas:



Simple: Crea un gráfico de barras que resume categorías de una variable o varias (según la opción seleccionada), ordenadas en sentido descendente. Una línea muestra la suma acumulada.



Apilado: Crea un gráfico de barras que resume categorías de una variable o varias en categorías de otra, con las barras ordenadas en sentido descendente. Una línea representa la suma acumulada.

De la misma forma que en los gráficos anteriores, tenemos:

- ♦ **Recuentos o sumas para grupos de casos**
- ♦ **Sumas para distintas variables**
- ♦ **Valores individuales de los casos**

4.2.7. Control

Los gráficos de control de procesos nos ofrecen cuatro subtipos:



X-barra, R, s: Crea gráficos X-barra, R y s que representan la media, el rango y la desviación típica, respectivamente, en medidas de algunos procesos en ejecución (a lo largo del tiempo).



Individuos, rango móvil: Representa una única variable de medida del proceso por caso individual. En los gráficos de rango móvil, cada punto representa la diferencia entre el valor actual y el anterior.



p, np: Resume la proporción o el número de unidades disconformes en los subgrupos.



c, u: Crea un gráfico de control de proceso que resume el número de disconformidades (o disconformidades por unidad) dentro de los subgrupos.

Las opciones posibles son:

- ♦ **Los casos son unidades:** Se resume una variable de proceso.
- ♦ **Los casos son subgrupos:** Se resumen las variables en subgrupos de casos definidos por una variable categórica. Cada punto resume las variables seleccionadas en un único caso. Esta opción no está disponible para los gráficos de control de *Individuos*, *rango móvil*.

4.2.8. Diagramas de caja

Dos tipos de diagramas:



Simple: Crea un diagrama de caja que resume una variable numérica en categorías de otra variable, o bien resume varias variables. Cada caja muestra la mediana, los cuartiles y los valores extremos según un grupo de casos o según una variable.



Agrupado: En este caso se representa dentro de las agrupaciones definidas por una variable categórica. Cada caja de una agrupación se define mediante una segunda variable categórica.

Y dos opciones ya vistas:

- ♦ **Resúmenes para grupos de casos**

- ♦ **Resúmenes para distintas variables**

4.2.9. Barras de error

Dos posibilidades como en el caso anterior:



Simples: Crea un gráfico que resume la distribución de una única variable numérica en categorías de otra variable, o resume varias variables numéricas.



Agrupadas: Crea un gráfico que resume la distribución de una o más variables numéricas en agrupaciones definidas por una variable categórica.

Igual que el caso anterior:

- ♦ **Resúmenes para grupos de casos**
- ♦ **Resúmenes para distintas variables**

4.2.10. Dispersión

Este tipo de gráficos son más específicos en contraste con los anteriores. Sólo tenemos cuatro opciones sin más posibilidades:



Simple: Crea un gráfico de dispersión en dos dimensiones que relaciona entre sí dos variables numéricas. Cada variable se representa en un eje.



Superpuesto: Crea un gráfico con dos o más pares de variables. Representa simultáneamente los valores de la primera variable de cada par frente a los de la segunda.



Matricial: Crea un diagrama con todas las combinaciones posibles de dos o más variables numéricas entre sí. Es una representación, en dos dimensiones, de los valores de cada una de las variables de matrices frente a los de cada una de las restantes.



3-D: Crea un gráfico de tres variables numéricas en tres dimensiones. Representa los valores de una variable frente a los de las otras dos.

4.2.11. Histograma



Crea un histograma que muestra la distribución de una única variable numérica. Un histograma representa la distribución de una variable cuantitativa que muestra la

concentración relativa de los datos a lo largo de diferentes intervalos o secciones de la escala en la que están medidos dichos datos.

4.2.12. P-P



Crea un gráfico de las proporciones acumuladas de una variable respecto a las de una distribución cualquiera de prueba. Es decir, por cada variable especificada proporciona dos gráficos: en el primero se representan los valores de la función de distribución acumulativa esperada bajo el supuesto de normalidad (por defecto) frente a los observados. En el segundo se representan los residuos.

Los gráficos de probabilidad suelen emplearse para determinar si la distribución de una variable coincide con una distribución dada. Si la variable seleccionada coincide con la distribución de prueba, los puntos se concentran en torno a una línea recta.

Entre las distribuciones de prueba disponibles se encuentran la beta, chi-cuadrado, exponencial, gamma, semi-normal, Laplace, logístico, Lognormal, normal, Pareto, t de Student, Weibull y uniforme. Según la distribución elegida, pueden especificarse distintos grados de libertad y otros parámetros.

4.2.13. Q-Q



Crea un gráfico con los cuantiles de distribución de una variable respecto a los cuantiles de una distribución cualquiera de prueba. Es decir, por cada variable especificada proporciona dos gráficos: en el primero se representan los valores de los cuantiles esperados bajo el supuesto de distribución normal (por defecto) de media cero y varianza uno frente a los observados. En el segundo, en lugar de los cuantiles esperados, se representan los residuos.

Las distribuciones de prueba disponibles son las mismas que en P-P.

4.2.14. Secuencia



Crea un gráfico de casos en secuencia. Este procedimiento requiere datos de serie temporal, u otros en los que se ordenen los casos con un criterio significativo.

Representa, en dos dimensiones, los valores de cada una de las variables especificadas frente al número de secuencia en el archivo de datos. Como hemos dicho, si la variable es una serie temporal a cuyos valores se han asignado fechas (opción *Definir fechas*), la representación se realiza frente al instante de tiempo correspondiente.

4.2.15. Curvas COR



Este procedimiento es un método útil para evaluar la realización de esquemas de clasificación en los que exista una variable con dos categorías por las que se clasifiquen los sujetos. Permite calcular la estimación del área situado bajo la curva COR de forma paramétrica o no paramétrica mediante un modelo exponencial binegativo.

Además, permite especificar si se debe incluir o excluir el valor del punto de corte al realizar una clasificación positiva, y es posible especificar la dirección de la escala según la categoría positiva.

4.2.16. Serie temporal: autocorrelaciones



Crea un gráfico con la función de autocorrelación simple (FAS) y la función de autocorrelación parcial de una o varias series. Este procedimiento sólo es adecuado en datos de serie temporal.

- ♦ *Autocorrelación simple*: correlaciona los valores de una serie con los valores retardados en uno o más casos. Las autocorrelaciones se calculan para los retardos de orden 1, 2, ..., hasta un número especificado.
- ♦ *Autocorrelación parcial*: correlaciona los valores de una serie con los valores retardados en uno o más casos, después de haber eliminado los efectos de las correlaciones para los retardos intercalados.

4.2.17. Serie temporal: correlaciones cruzadas



Crea un gráfico con la función de correlación cruzada de dos o más series para retardos positivos, negativos y de valor cero. Es decir, representa para cada par de variables especificadas, los valores de la función de correlación cruzada. Este procedimiento sólo es apropiado en datos de serie temporal.

Correlación cruzada: correlación que se da entre dos series temporales. Las observaciones de una serie se correlacionan con las observaciones de otra serie para diversos retardos y adelantos. Se representan habitualmente en un gráfico que ayuda a identificar las variables que son indicadores precedentes de otras variables. El procedimiento CCF, del módulo *Tendencias*, genera las correlaciones cruzadas.

4.2.18. Serie temporal: análisis espectral



El análisis espectral analiza la variación de las series como un todo en componentes periódicos de frecuencia diferente en lugar de analizar la variación desde un punto temporal al siguiente, que es lo usual en las técnicas de serie temporal. Las series suaves tienen más variación a bajas frecuencias; la variación aleatoria ("ruido blanco") dispersa la variación a través de todas las frecuencias.

Este análisis representa para cada variable especificada los valores del periodograma (u otra función de análisis espectral) con objeto de obtener una estimación de densidad espectral. Para suavizar el periodograma se dispone de Tukey-Hamming, Tukey, Parzen, Bartlett, Unidad de Daniell, o de ninguno.

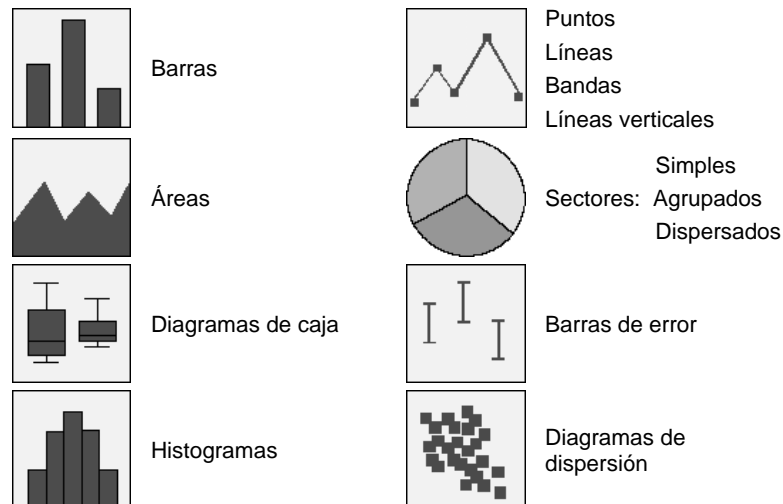
Las series con datos perdidos no pueden analizarse con este procedimiento.

4.3. Galería de gráficos interactivos

Los gráficos interactivos son mucho más potentes que los gráficos normales. Permiten especificar su formato con multitud de opciones y su creación es mucho más gráfica e intuitiva.

Sin embargo, como ya hemos comentado, los requerimientos de recursos (memoria, espacio en disco y procesador) son muy grandes, por lo que es conveniente usarlos con una cantidad reducida de datos, de lo contrario el tiempo necesario para crearlos puede ser excesivo. Además, hay ciertos tipos de gráficos que no es posible generar en su forma interactiva, como los de pareto o lo de control.

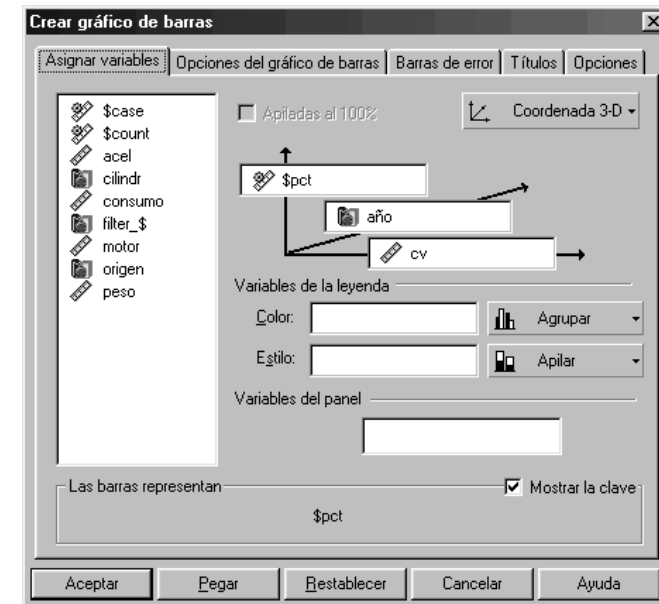
Los gráficos disponibles en su configuración inicial son:



Para acceder a la *Galería*, hemos de buscar dicha opción dentro de los temas de la ayuda. Después, seleccionamos uno de los tipos y nos aparecen todos los subtipos de gráficos posibles. Por último, pinchando en uno de los subtipos nos aparece otra ventana de ayuda que austeramente nos indica cómo hacerlo.

4.4. Creación de gráficos interactivos

Cuando accedemos a un gráfico de este tipo a través del menú *Gráficos / Interactivos*, se nos muestra un cuadro de diálogo distinto a los que veníamos utilizando hasta ahora. Este nuevo cuadro posee varias pestañas:



Estas pestañas nos permiten acceder a los distintas configuraciones del gráfico. En la parte inferior se nos muestran los típicos botones de los cuadros de diálogo para aceptar o cancelar el procedimiento, etc. Los cuadros del resto de gráficos son similares.

La primera pestaña es la de selección de variables. Nos permite asignar variables a los ejes, designar agrupaciones, cambiar a 3D (si es posible), aplicar funciones para representar las variables, etc. Para asignar variables hay que arrastrarlas con el ratón hasta la casilla correspondiente. Para quitarla la arrastraremos de nuevo a su sitio o bien situaremos otra en su lugar. Con la tecla [Ctrl] podemos arrastrar varias.

La lista de origen (en la parte izquierda) presenta una relación de las variables disponibles para el tipo de gráfico seleccionado. Se puede cambiar el orden de las variables en la lista con el menú contextual. También se puede elegir entre mostrar las etiquetas de las variables o sus nombres.

El tipo de variable se indica mediante un símbolo a su izquierda:

- Función de escala: usada para el eje dependiente; resume casos o variables.
- Variable de escala.
- Variable categórica.

La segunda pestaña es la de configuración del gráfico en sí. Dependiendo del tipo que sea aparecerán unas opciones u otras. En el caso de ejemplo, que es un gráfico de barras, se nos mostrarían opciones para cambiar la forma de las barras, las etiquetas del gráfico, etc.

El gráfico puede tener más pestañas de configuración propia del gráfico, o de barras de error (si son posibles), en cuyo caso también podríamos modificar el formato de estas barras.

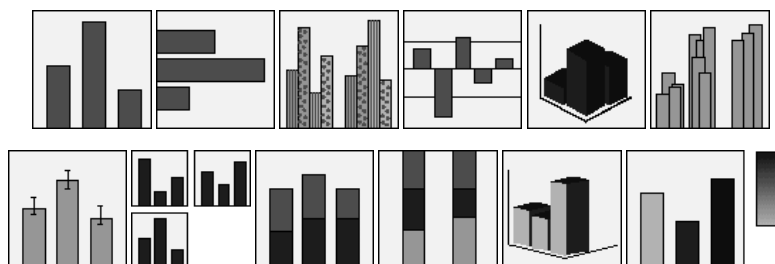
La penúltima pestaña nos ofrece tres paneles para introducir el título, subtítulo y el pie del gráfico.

La última pestaña es la de opciones generales (independientes del tipo de gráfico) como los órdenes categóricos, las escalas, la longitud de los ejes, el aspecto, etc.

4.4.1. Gráficos de barras interactivos

Presenta estadísticos de resumen de una o más variables, la mayor parte de las veces respecto a los grupos definidos por una o dos variables categóricas. La longitud de las barras representa habitualmente la frecuencia de casos de cada categoría, un porcentaje del número total de casos o una función de otra variable (por ejemplo, el valor medio para cada categoría).

Las posibilidades gráficas son muy extensas:



Pestaña *Asignar variables:*

La función se puede cambiar en la parte inferior del cuadro de diálogo, y se puede mostrar una clave para identificar la función.

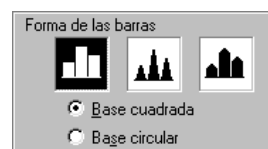
Permite cambiar la orientación, de vertical a horizontal, y la representación en 3-D para añadir otra variable, así como simplemente generar un efecto 3-D.

Si se asigna una variable categórica a *Color* o *Estilo*, las barras de las categorías se podrán agrupar unas junto a otras o apilarse. En las barras apiladas, las categorías de la variable de la leyenda se apilan unas sobre otras. La parte superior de un segmento conforma la base del siguiente. Este tipo de gráficos destaca la suma de las categorías; por tanto, sólo se deberán seleccionar funciones adecuadas para la suma en *Las barras representan*. Si se ha seleccionado una leyenda de apilado de color o de estilo, podremos cambiar el eje dependiente para que represente el 100%. En ese caso, los segmentos representarán el porcentaje con el que cada categoría contribuye al total.

Si se asigna la variable preincorporada *CASO* a un eje, aparecerá otra opción disponible en la que se podrá asignar una variable para etiquetar los casos. Por ejemplo, las etiquetas pueden ser días de la semana o meses.

Pestaña *Opciones del gráfico de barras:*

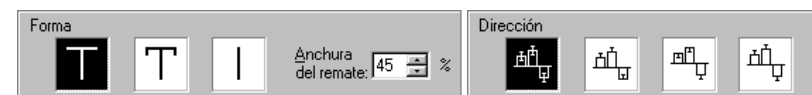
Existen varias opciones para mejorar los gráficos de barras, como modificar su forma, etiquetar las barras o establecer el valor de la línea base de las barras. En el ejemplo mostrado vemos las formas posibles de las barras y la selección de base, cuadrada o circular.



Pestaña *Barras de error:*

Las barras de error, que indican la variabilidad de la medida mostrada, se encuentran disponibles cuando la variable dependiente es una variable de escala y las barras o las líneas representan la función de resumen *Medias*. La longitud de las barras de error se muestra como los límites inferior y superior de un intervalo de confianza (para un nivel porcentual especificado) o en desviaciones típicas o en errores típicos de la media.

Se puede especificar el nivel del intervalo de confianza, la dirección y la forma de las barras de error y cambiar su aspecto. En el ejemplo vemos las diferentes formas que pueden tomar los "bigotes", el porcentaje de anchura del remate y la dirección de las barras:



Pestaña *Títulos:*

Se puede añadir un título y un subtítulo al gráfico, así como establecer un pie.

Pestaña *Opciones:*

Por último, en esta ficha se muestran otros métodos para personalizar los resultados. Se puede determinar el orden de las variables categóricas y el rango de las variables de escala. También se puede elegir un *Aspecto de gráfico* y el tamaño de la región de datos.

Las variables categóricas se pueden ordenar por una propiedad (valores, etiqueta, apariciones o frecuencia) o por una función de resumen de otra variable. Además, se puede determinar que el orden sea ascendente o descendente, y excluir categorías sin datos.

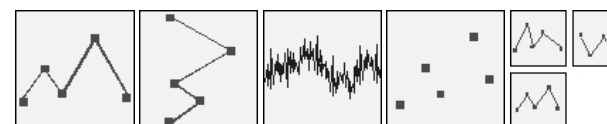
Se puede establecer el valor mínimo y el máximo de todas las variables de escala, utilizar los valores predeterminados del conjunto de datos o crear una amplitud personalizada.

La lista *Aspectos de gráfico* contiene varios aspectos almacenados en el directorio "Looks" (por defecto). Estos aspectos controlan las propiedades visuales iniciales del gráfico, tales como los colores y los símbolos.

El *Tamaño* de la región de datos viene determinada por la longitud de cada eje. Si se asigna una variable de panel, la longitud de los ejes Y y X1 se utilizará para determinar la relación de aspectos de los paneles.

4.4.2. Gráficos de puntos interactivos

Su objeto es el mismo que el de los gráficos de barras. Resumen categorías de una o más variables. Representa puntos en los puntos de datos, pero también líneas que conecten los puntos o tanto líneas como puntos. Ejemplos:



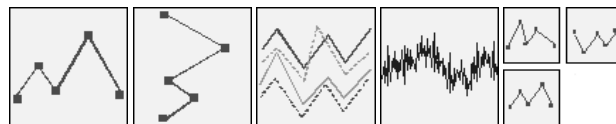
Las pestañas son las mismas que en el caso anterior, a excepción de las opciones propias del gráfico que podemos usar para modificar las especificaciones particulares de este tipo de gráfico. Esta pestaña se emplea también en líneas, bandas y líneas verticales.

Por ejemplo, una de estas opciones propias es la de interpolación, que controla el modo de dibujar las líneas entre los puntos de datos. Las elecciones posibles serían:



4.4.3. Gráficos de líneas interactivos

Realmente este tipo de gráfico funciona al unísono con el anterior y con el de líneas verticales, sólo debemos seleccionar en las opciones de *puntos y líneas* si queremos mostrar puntos, líneas o líneas verticales. Ejemplos:



Las líneas, en concreto, tienden a enfatizar el flujo o los movimientos, y no los valores individuales. Se suelen utilizar para representar datos a lo largo del tiempo y, por tanto, pueden usarse para observar tendencias. Los gráficos de líneas son especialmente eficaces para representar un gran número de valores en un espacio compacto.

4.4.4. Gráficos de bandas interactivos

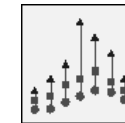
Un gráfico de bandas es similar a uno de líneas, pero las líneas se representan como bandas en una tercera dimensión. Los gráficos de bandas pueden ser bidimensionales (con un efecto 3-D) o tridimensionales. Ejemplo:



El resto es idéntico a los de líneas y puntos.

4.4.5. Gráficos de líneas verticales interactivos

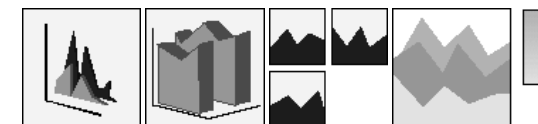
En los gráficos de líneas múltiples o de puntos, en los que se asigna una variable de categoría a *Color* o a *Estilo*, se pueden dibujar líneas verticales entre los puntos de color o de estilos del mismo eje de categoría. Ejemplo:



En cuanto al resto, nos remitimos a los gráficos de puntos y líneas.

4.4.6. Gráficos de áreas interactivos

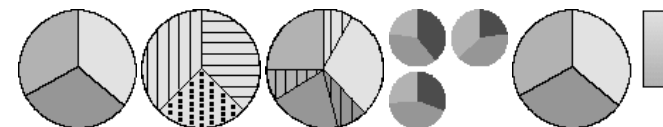
En un gráfico de áreas simple los puntos se representan y se unen con líneas, y el área comprendida entre la línea de datos y la línea base se rellena. Uno de los usos más comunes de estos gráficos consiste en mostrar la continuidad de una variable de resumen (dependiente) en el tiempo (variable independiente). Posibilidades:



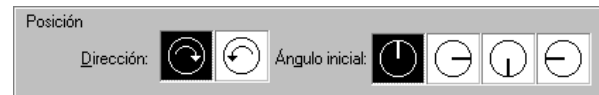
Lo único que varía respecto a los anteriores tipos de gráficos son las opciones propias, que en este caso sólo nos permiten configurar las etiquetas y los valores perdidos.

4.4.7. Gráficos de sectores simples interactivos

Resume las categorías definidas por una sola variable o por un grupo de variables relacionadas. En cada categoría, el tamaño de cada sector representa la frecuencia, el porcentaje o una función de resumen de una variable. También se puede crear un gráfico de sectores con paneles. Ejemplos:



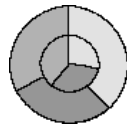
Las opciones nos permiten establecer las etiquetas y la posición. Por ejemplo, vemos cómo cambiar la dirección y el ángulo inicial:



El resto de opciones no difiere de las de los tipos de gráficos anteriores.

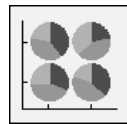
4.4.8. Gráficos de sectores agrupados interactivos

Contiene una agrupación de gráficos de sectores simples, todos ellos apilados como categorías de una misma variable. Los gráficos poseen diferentes tamaños y aparecen apilados unos sobre otros. La agrupación contiene tantos gráficos de sectores como categorías hay en la variable de agrupación. Veamos un ejemplo:



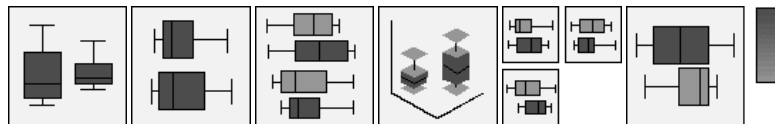
4.4.9. Gráficos de sectores dispersados interactivos

Contiene un conjunto de gráficos de sectores situados en puntos definidos por coordenadas rectangulares. Es similar a un diagrama de dispersión con sectores como marcadores gráficos. Todos los gráficos están divididos en sectores por la misma variable categórica. Ejemplo:



4.4.10. Diagramas de caja interactivos

Muestran la mediana, los cuartiles, los valores atípicos y los valores extremos para una variable de escala. Pueden dividirse en categorías y agrupaciones. Posibilidades:



En las opciones podemos mostrar los valores atípicos, los extremos y las líneas de la mediana de cada caja. Los valores atípicos se encuentran a una distancia del final de la caja de

entre 1,5 y 3 veces la longitud de ésta. Los extremos se encuentran a una distancia del final de la caja de más de 3 veces la longitud de la caja.

En los diagramas de caja 2-D se puede elegir entre varias formas de remate para cada bigote. En las cajas 3-D, la forma de la base puede ser circular o cuadrada.

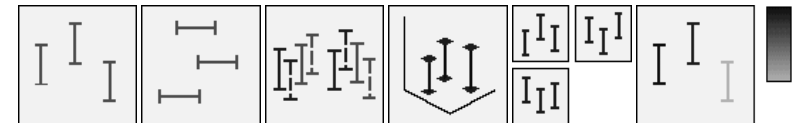
Por último, si *Mostrar etiquetas de frecuencia* está seleccionado, mostrará el número de casos incluidos en cada categoría.

4.4.11. Barras de error interactivas

Las barras de error facilitan la comprobación en el gráfico de las distribuciones y la dispersión indicando la variabilidad de la medida mostrada. La media de una variable de escala se representa para un conjunto de categorías; la longitud de una barra de error a cada lado del valor medio indica un intervalo de confianza o un número especificado de errores típicos o de desviaciones típicas. Las barras de error pueden extenderse en una o en ambas direcciones desde la media.

A veces, en el mismo gráfico se muestran las barras de error junto a otros elementos gráficos como, por ejemplo, barras.

Las opciones nos permiten modificar las etiquetas, la dirección o la forma, así como si queremos ver una marca en el valor medio de cada barra de error. Ejemplos:



4.4.12. Histogramas interactivos

Reúne los valores de una variable en grupos espaciados regularmente (intervalos o clases) y representa el recuento del número de casos dentro de cada grupo. Este recuento, o frecuencia, se puede expresar en forma de porcentaje, el cual es especialmente útil para comparar conjuntos de datos de diferentes tamaños. La frecuencia, o el porcentaje, también pueden acumularse a lo largo de los grupos.

El histograma permite detectar los valores atípicos y las desviaciones de la simetría. Dichas desviaciones pueden indicar que una variable no es adecuada para ser analizada mediante un procedimiento que asuma una distribución normal.

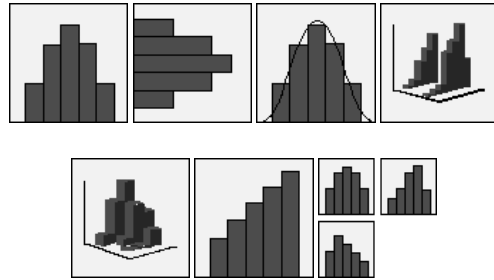
A través de la pestaña *histograma*, se puede superponer una curva normal (sólo disponible para 2-D) con la misma media y varianza que las de los datos.

Permite situar intervalos y punto de inicio de la variable sobre un eje determinado.

Podemos establecer el número de intervalos (máximo 250) o su anchura, de lo contrario se ajustará automáticamente.

El punto inicial del primer intervalo puede establecerse en cualquier lugar desde el valor mínimo de escala hasta el valor mínimo más la anchura de un intervalo.

Ejemplos de distintas configuraciones:



4.4.13. Diagramas de dispersión interactivos

Estos diagramas destacan la relación existente entre dos o tres variables cuantitativas mediante la representación de los valores reales a lo largo de dos o tres ejes. A menudo es posible observar relaciones, un patrón curvilíneo por ejemplo, que los estadísticos descriptivos no revelan. También es posible descubrir valores atípicos bivariados.

Asignando una variable de leyenda, se pueden estudiar las características de los subgrupos. Además, pueden utilizarse los valores de una variable para etiquetar los puntos, de lo contrario se emplearán los números de caso como etiquetas.

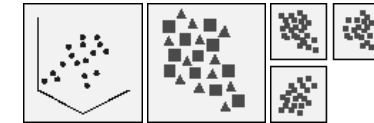
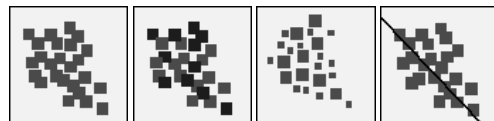
Para ajustar líneas a los puntos de un diagrama de dispersión disponemos de los métodos Regresión, Media, Suavizador, o ninguno. Con el método Regresión, se puede indicar si queremos incluir una constante en la ecuación. Con el Suavizador se puede indicar el tipo de kernel, los multiplicadores del ancho de banda, etc.

Con los métodos media o de regresión se pueden mostrar las líneas del pronóstico promedio o bien las líneas del pronóstico individual para ilustrar el nivel de confianza especificado.

Además de una línea de ajuste para la nube total, si existe una división categórica, se pueden seleccionar las líneas de ajuste para los subgrupos.

Los trazos de unión son líneas dibujadas entre una nube de símbolos y un punto, un eje, una línea o una superficie. Estos trazos pueden resultar útiles para leer los valores de los ejes o para comparar distancias mediante la longitud de las líneas. Los trazos que se dibujan hasta un eje horizontal se denominan también líneas verticales.

Algunos ejemplos son:



4.5. Creación de gráficos según la estructura de los datos

Para la creación de un gráfico es fundamental tener claro lo que se desea obtener y el tipo de gráfico adecuado a nuestra estructura de datos.

A partir de aquí se establece una clasificación de los gráficos que podremos utilizar según las variables de las que consten nuestros datos.

4.5.1. Gráficos para variables categóricas

- ♦ **Para una única variable categórica:** Podemos utilizar barras simples, líneas simples, áreas simples, sectores y Pareto simple.
Siempre hemos de seleccionar resúmenes para grupos de casos ya que tratamos con una única variable.
- ♦ **Categorías de una variable dentro de categorías de otra variable:** Barras agrupadas o apiladas, líneas múltiples o verticales, áreas apiladas, Pareto apilados, máximo-mínimo-cierre simples, barras de rango simples y líneas de diferencia.
El análisis para todos ellos por grupos de casos.

4.5.2. Gráficos para variables continuas

- ♦ **Variable continua única:** Resúmenes para distintas variables en diagramas de caja simples o en barras de error simples, histogramas, P-P, Q-Q, los casos son subgrupos en p, np y en gráficos c,u, secuencias, autocorrelaciones y espectral.
- ♦ **Variable continua única dentro de categorías de otra variable:** Resúmenes para distintas variables en diagramas de caja simples o en barras de error simples, los casos son unidades en X-barra, R, s; p, np y c,u.
- ♦ **Varias variables continuas:** Resúmenes para distintas variables en barras simples, líneas simples, sectores, diagramas de caja simples, barras de error simples; diagrama de dispersión simple, superpuesto matricial y 3-D; sumas para distintas variables en Pareto simples, los casos son subgrupos en X-barra, R, s; secuencia, autocorrelaciones, correlaciones cruzadas y espectral.
- ♦ **Varias variables continuas dentro de categorías de otra variable:** Resúmenes para distintas variables en barras agrupadas o apiladas, líneas múltiples y verticales, áreas apiladas, diagramas de caja agrupados, barras de error agrupadas; diagrama de

dispersión simple con variable de control, Pareto apiladas, máximo-mínimo-cierre simples, barras de rango simples, líneas de diferencia simples; resúmenes para grupos de casos en máximo-mínimo-cierre agrupados y en barras de rango agrupadas.

4.5.3. Gráficos para casos individuales

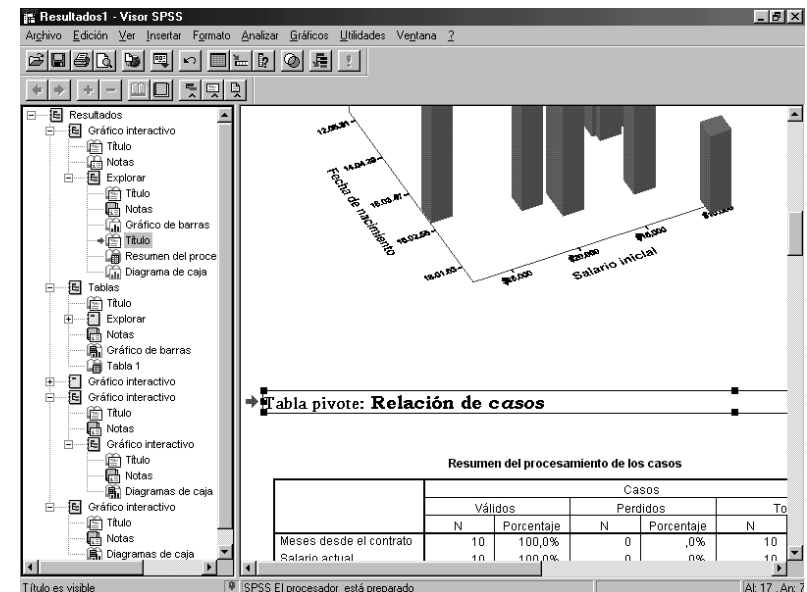
- ♦ **Casos individuales (variable única):** Valores de los casos individuales en barras simples, líneas simples, áreas simples, sectores, Pareto simples; y los casos son unidades en Individuo, rango móvil.
- ♦ **Casos individuales (dos o más variables):** Barras agrupadas y apiladas, líneas múltiples y verticales, áreas y Pareto apilados, máximo-mínimo-cierre simples y agrupados, barras de rango simples y agrupadas, y líneas de diferencias.
Siempre con la opción *Valores de los casos individuales*.

4.5.4. Gráficos especializados

- ♦ **Variable única de máximo-mínimo-cierre dentro de categorías:** Resúmenes para grupos de casos en máximo-mínimo-cierre simples, barras de rango simples y líneas de diferencia.
- ♦ **Variables distintas de máximos y mínimos dentro de categorías:** Resúmenes para distintas variables en máximo-mínimo-cierre simples y agrupados, barras de rango simples y agrupadas, líneas de diferencia. Resúmenes para grupos de casos en máximo-mínimo-cierre agrupados y barras de rango agrupadas.
- ♦ **Distintas variables en máximos y mínimos por caso individual:** Valores de los casos individuales en máximo-mínimo-cierre simples y agrupados, barras de rango simples y agrupadas, líneas de diferencia.
- ♦ **Control de procesos. Variable de medida única:** Los casos son unidades en Individuo, rango móvil. Los casos son subgrupos en p,np; c,u.
- ♦ **Control de procesos. Variable única dentro de categorías:** Los casos son unidades en X-barra, R, s; p, np; c,u. Los casos son subgrupos en X-barra, R, s.
- ♦ **Series temporales:** Secuencia, autocorrelaciones, correlaciones cruzadas y espectral.

5. EXAMEN DE RESULTADOS

Los resultados de los análisis, gráficos, tablas o textos, se van acumulando como objetos en el *Visor*, que se abre automáticamente la primera vez que ejecutemos cualquier procedimiento que genere resultados.



5.1. Uso del visor

Como podemos observar en la ilustración anterior, el *Visor* nos muestra dos paneles. El de la izquierda es el panel de titulares que nos presenta el índice de los objetos de resultados que tenemos en el panel de la derecha. Los objetos de resultados pueden ser textos (en formato RTF), gráficos o tablas pivote.


Con el *Visor* se nos muestran dos barras de herramientas. La propia del *Visor* y la del panel de titulares. La del *Visor* es similar a la del *Editor de datos*, se han extraído algunos botones y se han añadido otros como el de seleccionar el último resultado obtenido o el indicador de ventana activa.

Para editar un objeto basta con pulsar dos veces sobre él con el ratón, elegir la opción *Editar* del menú contextual o del menú *Edición*. En ese momento el *Visor* se transformará en un editor, con barras de botones distintas y nuevas opciones en la barra de menús. Si por el contrario seleccionamos la opción *Abrir*, el objeto pivote aparecerá en una nueva ventana distinta de la del *Visor*, que será su editor.

En el caso de los gráficos, no podremos editarlos en el mismo *Visor*, sólo tenemos la opción de abrirlos para modificarlos en una ventana distinta, que será el *Editor de gráficos*.

Por el contrario, con los gráficos interactivos el *Visor* permanece inalterado y las opciones de edición del gráfico aparecen alrededor de él mismo.

En el menú *Archivo* podemos encontrar las opciones que vimos en el *Editor de datos*, pero ligeramente ajustadas al *Visor*. Podemos crear una nueva ventana de *Visor*, abrir resultados almacenados en disco (*.spo) y, por supuesto, guardarlos.

Cuando usemos varias ventanas de *Visor* debemos tener en cuenta un símbolo de admiración existente en la barra de estado. La ventana designada para representar los resultados mostrará este símbolo en rojo. Si queremos que los resultados aparezcan a partir de ese momento en otra ventana, nos desplazaremos a ella y pulsaremos el botón .

Continuando con el menú *Archivo*, existe la opción de *Exportar*, que nos da tres posibilidades: exportarlo todo, todo sin los gráficos, o sólo los gráficos. A la vez podemos exportar sólo los objetos visibles o los seleccionados. Según el tipo de objeto que queramos exportar, el formato puede ser texto o HTML (resultados en general), o diversos formatos gráficos (jpg, bmp, wmf, tif, eps, etc.). Y según este tipo de objeto, las opciones que nos mostrará serán distintas.


Por último, comentar que las opciones de *Presentación preliminar* e *Imprimir* sólo muestran e imprimen, respectivamente, los objetos seleccionados, a no ser que le indiquemos lo contrario. Además, a través de *Preparar página* podemos configurar la página antes de imprimir y añadir cabeceras y pie generales para todas las páginas.


5.1.1. Panel de titulares y panel de resultados


Los titulares están estructurados en un árbol jerárquico. Existen elementos individuales que se corresponden con elementos gráficos (objetos) en el panel de la derecha, y existen elementos de encabezamiento (nombres de procedimiento) que agrupan elementos individuales o a otros encabezamientos.

El elemento seleccionado se indica mediante una flecha roja ➡, tanto en el panel de titulares como en el de objetos de resultado. Además, podemos seleccionar varios elementos, incluso aunque no sean consecutivos, y cambiarlos de sitio. Esto último producirá cambios en el panel de la derecha, observaremos cómo los elementos gráficos también varían su posición. Para seleccionar varios elementos en el panel de objetos podemos describir un rectángulo arrastrando el puntero del ratón.


Cualquier elemento del panel de titulares es susceptible de ser renombrado, aunque esto no conlleva ningún cambio en los objetos de texto. Además, en el menú *Ver* se encuentra la posibilidad de cambiar el tamaño de los titulares así como la fuente empleada para visualizarlos.

Los titulares de elementos tienen forma de libro que puede estar abierto o cerrado. Si está abierto, el elemento gráfico correspondiente se mostrará en el panel de la derecha. Para abrir o cerrar un libro podemos hacer doble click sobre él o usar las opciones del menú *Ver*, *Mostrar* y *Ocultar*, o usar los botones de la barra del panel de titulares: .

Asimismo, los encabezados poseen forma de libreta que puede estar abierta o cerrada. Si está abierta, se muestran todos los elementos gráficos supeditados a él. Para abrir o cerrar usaremos los símbolos a la izquierda del encabezado, las opciones *Expandir* o *Contraer* del menú *Ver*, o los botones de la barra: .

Un titular se puede ascender o degradar en la estructura jerárquica. Para ello podemos desplazarlos, utilizar la opción *Titular* del menú *Edición*, o bien usar los botones .

Además, cualquier elemento se puede copiar, cortar, pegar, eliminar, etc. Todo esto, como es natural, se encuentra accesible rápidamente mediante el menú contextual.

Mediante el menú *Insertar*, o los botones de la barra , podemos añadir encabezados, títulos, textos, gráficos, objetos OLE externos a SPSS, etc. La inserción de saltos de página es conveniente para distribuir los objetos por grupos ya que esto no se realiza de forma automática en la impresión.

El menú *Formato* únicamente nos permite alinear a la izquierda, derecha, o centrar cualquier objeto del panel de resultados.

Al seleccionar un objeto de tipo gráfico (no interactivo) o texto, aparecen unos manejadores en los puntos medios y extremos del rectángulo que lo rodea. Con ellos podremos modificar el tamaño del objeto.

Por último, cualquier objeto o grupo de objetos seleccionado es susceptible de ser copiado y pegado en otras aplicaciones (por ejemplo, Word) en forma de imagen o texto.

5.1.2. Textos

Los textos en SPSS están definidos según el formato RTF, lo que les da una gran capacidad para ser exportados o importados sin pérdidas de formato.

Ya hemos comentado como acceder al *Editor de resultados de texto* o acceder a la edición en el mismo *Visor*. En ambos casos tenemos las mismas opciones.

La edición consiste básicamente en el cambio de formato, de la fuente, del tamaño y del color, así como la aplicación de los estilos negrita, subrayado y cursiva. Podemos alinear el texto a un lado o centrarlo. Además se encuentran disponibles las funciones típicas de cualquier elemento, como copiar, cortar, pegar, seleccionar, etc.

La barra de herramientas del *Editor de resultados de texto* es la siguiente:



El editor dispone de botones de búsqueda, de salto entre páginas y de una opción para insertar un salto de página manual.

5.1.3. Tablas pivote

Las tablas pivote son las tablas que usa SPSS para la presentación de resultados. Se caracterizan por su capacidad de personalización, manejo y edición (aunque no siempre es sinónimo de sencillez). Un ejemplo sería:

Estadísticos descriptivos					
	Mínimo	Máximo	Media	Desv. típ.	Varianza
Nivel educativo	8	16	13,50	2,46	6,056
Categoría laboral	1	3	1,50	,85	,722
Salario actual	\$21,450	\$57,000	*****	*****	1,4E+08
Salario inicial	\$9,750	\$27,000	*****	\$5,239.66	2,7E+07
Meses desde el contrato	50	80	66,70	14,10	198,900

Para modificar una tabla hemos de acceder al *Editor de tablas pivote* que nos permite, básicamente, insertar títulos y notas al pie, seleccionar un grupo de casillas (o la tabla completa) para aplicar formatos, ocultar y mostrar columnas y otros elementos, arrastrar casillas para copiar, ver cuadrículas, agrupar y desagrupar conjuntos de casillas, modificar el ancho de las casillas, autoajustar ese ancho, cambiar la fuente y sus propiedades, rotar etiquetas, y algunas otras funciones usuales. Además podemos editar el texto de una casilla pulsando dos veces sobre ella, pero cuidado, al cambiar el valor de una casilla no se modificarán los valores del resto de casillas dependientes de ella (como totales, por ejemplo).

Las propiedades de una casilla o grupo de casillas son extensas. Podemos acceder a ellas desde el menú *Formato* o desde el menú contextual. Nos permite establecer el formato del dato (fecha, numérico, etc.), su alineación horizontal y vertical, sus márgenes y los colores de fondo usados. Asimismo podemos usar determinadas funciones a través de la barra de herramientas:



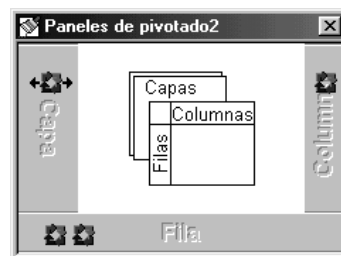
La barra nos proporciona las funciones más usuales como los formatos o el acceso directo a la ayuda (icono de signo de interrogación).



Las propiedades de la tabla completa nos permiten controlar los aspectos generales de la tabla, y los aspectos particulares aplicados a todos los elementos. Podemos establecer las dimensiones de las columnas, las notas al pie, el formato de todas las casillas, el borde y color de la tabla y de sus diferentes elementos, y algunos parámetros de impresión.

SPSS posee varios aspectos de tabla predefinidos que podemos usar para dar formato rápidamente a las tablas que generemos. Sólo es necesario elegir un aspecto de los disponibles en el menú *Formato*. En cambio, si el diseño de la tabla es nuestro, podemos almacenarlo en disco (con extensión *.tlo) para posteriores usos. Además, es posible editar los aspectos almacenados en disco.

Otra posibilidad es la de generar un gráfico a partir de la tabla. El comando se encuentra en el menú *Edición*, o bien pulsando el botón . Esto nos crea de forma automática el gráfico interactivo correspondiente.

El menú *Pivotar* es el que ofrece flexibilidad a las tablas. Una tabla extensa puede descomponerse en varias capas dependiendo de una o varias variables. De esta forma sólo veremos la capa que queramos en cada momento. Para realizar estos cambios pinchamos sobre el botón o sobre *Paneles de pivotado* que nos aparecerán de inmediato en pantalla:



Estos paneles nos muestran con símbolos de pivotado () las variables o elementos que determinan las filas, las columnas y las capas de la tabla. Podemos desplazar estos elementos arrastrándolos con el ratón al eje que queramos. Esto producirá los cambios pertinentes de datos en la tabla. Las tablas admiten varios elementos en cualquier posición. En el caso de las capas, para pasar de un valor de la variable o elemento a otro tendríamos que pulsar sobre las flechas del símbolo de pivotado () así se nos mostraría la capa deseada en la tabla. También se puede cambiar de categoría a través de la lista desplegable que nos aparece en la misma tabla:

Sexo	Hombre		
	Hombre	Categoría laboral	
	Mujer	Administrativo	Seguridad

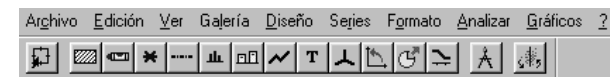
Otras opciones del menú *Pivotar* son los señalizadores, que permiten guardar diferentes presentaciones de una tabla pivote; la transposición de filas y columnas, restablecer pivotes por defecto, ir a una capa determinada, etc.

Para una correcta impresión debemos mencionar que existe la posibilidad de indicarle a la tabla por dónde debe separarse en caso de que no quepa en la hoja debido a su extensión. Una vez seleccionada la etiqueta de la columna situada a la izquierda del lugar donde queremos separarla, pulsamos sobre *Formato / Insertar ruptura*. Por otra parte, si la tabla contiene varias capas, a través del menú *Formato / Propiedades de la tabla / Impresión*, podemos indicar que se impriman todas las capas.

5.1.4. Gráficos

Los gráficos creados por SPSS son fácilmente modificables con su editor. Además incorpora la suficiente potencia para poder personalizar a nuestro gusto cualquier elemento del gráfico. Pero vayamos por partes.

Una vez que estamos en el *Editor de gráficos*, por ejemplo después de hacer doble click sobre el gráfico, se nos presenta una barra de herramientas específica y una barra de menús que difiere poco, aparentemente, de las otras.





Lo primero a tener en cuenta es que, dependiendo del gráfico que estemos manejando, algunas de las opciones permanecerán inactivas. Por ejemplo, el primer botón de la barra nos permite identificar puntos en gráficos de tipo dispersión, el cursor cambiaría de forma y una vez situados encima de un punto, nos mostraría el identificador/es del punto.


Los siguientes botones nos permitirían cambiar algunos aspectos del formato del gráfico: la trama de relleno, el color de las líneas, la forma de los marcadores, el estilo de las líneas y de las barras, y el estilo de las etiquetas de las barras.

Es obvio que para poder ejecutar la mayoría de funciones de formato hemos de seleccionar primero el objeto, o la parte del mismo, al que queremos aplicar el formato. En un

gráfico de barras con efecto 3-D, por ejemplo, podríamos seleccionar la parte frontal de las barras, la parte lateral, la parte superior, el área de representación gráfica, los valores representados en los ejes, los títulos de los ejes, la escala de los ejes, etc. Cuando un elemento o una parte suya es seleccionado, aparecen unos manejadores en los extremos del elemento u objeto.

El estilo de interpolación () es la forma de darle continuidad a un gráfico cuyos valores están representados por puntos a lo largo de una serie temporal. Con los botones siguientes podemos cambiar el formato de cualquier texto que aparezca en el gráfico, rotar los ejes de los gráficos tridimensionales, intercambiar los ejes entre sí, desgajar sectores en gráficos de ese tipo, y romper las líneas cuando encuentren valores perdidos.

El siguiente botón () muestra las opciones generales del gráfico, nos presenta un pequeño cuadro de diálogo para cambiar aspectos tales como la visualización o las etiquetas. Este cuadro es diferente para cada tipo de gráfico, incluso algunos no disponen de él.

El último botón () activa y desactiva los botones de giro en los gráficos tridimensionales. Con ellos podremos rotar los ejes en el sentido que queramos.

En cuanto a la barra de menús, si vamos a la opción *Archivo* veremos que ha cambiado sustancialmente. Básicamente nos permite guardar la plantilla gráfica (*.set) en uso, es decir, el conjunto de títulos, notas al pie, fuentes, colores, y otros formatos que estemos utilizando, para poder recuperarlos posteriormente y aplicarlos a otros gráficos del mismo o similar tipo. Además, podemos exportar el gráfico en varios formatos para su uso en otras aplicaciones.

En el menú *Edición* hemos de observar la inexistencia de la opción *Deshacer*, lo cual nos indica que debemos llevar cuidado con lo que hagamos ya que no podremos volver atrás.

La opción *Galería* nos presenta una lista de los tipos de gráficos en los que podemos transformar el nuestro. Si se precisa algún cambio de variable nos lo solicitará a través de un cuadro de diálogo.

En *Diseño* podemos acceder a los diferentes elementos del gráfico, incluidas sus opciones generales. Podemos modificar títulos, anotaciones, ejes, marcos, y otros elementos. Según el gráfico que tengamos, se nos mostrarán activas unas opciones u otras. Además, como ya hemos comentado, las opciones generales dependerán del tipo en cuestión.

En *Serie Visualizadas* podemos cambiar las variables de series mostradas. Tanto las opciones de *Serie* como las de *Diseño* son accesibles, casi todas, pulsando dos veces sobre el elemento gráfico que queremos modificar, sean los ejes, las variables representadas, las escalas, etc.

Por último, en el menú *Formato* tenemos las opciones representadas en la barra de herramientas que ya hemos descrito.

5.1.5. Gráficos interactivos


La modificación de los gráficos interactivos difiere en muchos aspectos de la de los gráficos normales. Lo primero que observamos es que la edición no se sitúa en otra ventana, como el *Editor de gráficos*, sino que se realiza en el mismo *Visor*. Por una parte, nos muestra una serie de botones situados en la parte superior e izquierda del gráfico y, por otra parte, la barra de menús permanece inalterada, pero sólo aparentemente, ya que los menús *Edición*, *Ver*, *Insertar* y *Formato* contienen opciones totalmente distintas de las del *Visor*.

Aunque parezcan dos, las barras de herramientas en que se dividen los botones son cuatro: *Utilidad*, *Texto*, *Estilo* y *Cursor*. La barra principal es la de *Utilidad*:




El primer botón nos muestra un cuadro similar al que hemos utilizado para crear el gráfico interactivo. Nos permite modificar las variables empleadas en la generación del gráfico en dos o tres dimensiones, etiquetar los casos, establecer una variable de ponderación, un filtro, etc.


El segundo botón sirve para insertar un nuevo elemento, gráfico o texto, que asume las variables establecidas para la creación del gráfico. De esta forma, podemos incrustar en los mismos ejes varios gráficos, a los que será recomendable aplicar distintos formatos para obtener una visualización aceptable. Aún así, si utilizamos muchos elementos gráficos en los mismos ejes podemos llegar a un punto en el que incluso sea difícil seleccionar con el cursor el elemento que queremos.

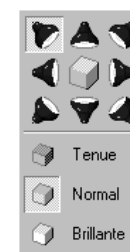
Para la cómoda gestión de los objetos gráficos, SPSS nos proporciona el *Administrador de gráficos*, accesible mediante el tercer botón (). Éste nos muestra una lista con estructura de árbol dividida en partes diferenciadas: el área de dibujo, los objetos gráficos que contiene, los objetos de texto, las leyendas, etc. De esta forma podemos acceder de forma inmediata a todos los objetos, mostrarlos, ocultarlos, eliminarlos y editarlos.

También podemos acceder a la edición de la misma forma que con los gráficos no interactivos, es decir, pulsando dos veces sobre cualquier elemento del gráfico o mediante las propiedades del menú contextual. Además todas las opciones vienen recogidas en el menú *Formato* que veremos después.

Continuando con la barra de *Utilidad*, existe un botón () que sólo se muestra activo si hemos efectuado alguna operación irregular. Si es así, al pulsarlo nos visualiza el administrador de gráficos. Dentro de éste encontraremos algún elemento con un símbolo de admiración que, al seleccionarlo, nos mostrará una advertencia en el cuadro inferior.

A la derecha de este botón disponemos de las opciones normales como deshacer (cuidado con el botón deshacer del *Visor*, pues no es el mismo) o rehacer, o las de intercambiar los ejes (siempre que estemos en 2-D).

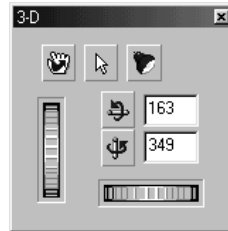
Disponemos de varios controles para el diseño 3-D accesibles mediante el botón (). El primer submenú nos muestra el cuadro de luces:





La iluminación de la imagen se realiza mediante un haz de luz simulado que parte de un foco situado en determinada posición alrededor del gráfico, o bien en el centro (haz plano).

Los controles de la parte inferior modifican la intensidad del brillo de los objetos: simulan el aumento o disminución de la luz.

En cuanto a la rotación del gráfico, disponemos del cuadro de giro 3-D. Podemos hacer girar el gráfico grado a grado de forma exacta, o bien a través de las ruedas de forma visual. Cuando los ejes queden en una posición oculta, automáticamente se situarán en la parte visible. El botón que muestra una mano se utiliza para el giro libre del gráfico, simula que agarramos el gráfico y lo hacemos rotar en cualquier sentido libremente. Con este botón, además, podemos darle un impulso para que el gráfico gire de forma autónoma.




Por último, el botón  permite conmutar entre 2-D y 3-D, o 2-D con efecto 3-D; y el botón  organiza automáticamente todos los objetos, tanto gráficos como texto.

La siguiente barra de herramientas es la de *Texto*. Sirve para cambiar el formato de un texto, sea un título, un subtítulo, los valores de escala de un eje, etc.





En posición vertical encontramos la barra *Cursor*. El primer botón que aparece seleccionado por defecto es la herramienta flecha, que sirve para seleccionar un objeto del gráfico, o varios utilizando la tecla [Ctrl].



El botón  permite introducir un texto en cualquier situación. La mano ya hemos visto para qué servía, y el último botón es el mismo que teníamos en los gráficos normales: se utiliza para identificar puntos en gráficos que los contengan, como el de dispersión.

La última barra de herramientas es la utilizada para cambios de estilo.




Los tres primeros botones permiten cambiar el color, el borde y el estilo de relleno del objeto. Los dos siguientes ( ) se encargan de modificar el estilo del símbolo y su tamaño. Los tres últimos son los que afectan a las líneas, permiten variar el estilo y grosor de la línea así como el conector empleado (cuando se necesite).

Una vez vistas las barras de herramientas, que son las que usaremos normalmente, pasamos a describir las nuevas opciones que tenemos en la barra de menús.

El menú *Edición* tiene funciones similares a las usuales, como son deshacer, rehacer, copiar, pegar, etc. Además podemos acceder a la asignación de variables y al *Administrador de gráficos*, dos opciones que ya hemos visto en la barra de *Utilidad*.

Las últimas opciones que nos presenta son propias de cada tipo de gráfico, con lo que se activarán o no y presentarán determinadas opciones según el que sea. Por ejemplo, si disponemos de un gráfico de tipo media de regresión, al seleccionar el objeto se activará la opción de parámetros y nos permitirá modificarlos. Asimismo ocurrirá con sus propiedades. Estas opciones, como es usual, son accesibles desde el menú contextual del objeto.

El menú *Ver* permite conmutar entre 2-D y 3-D, y visualizar u ocultar las barras de herramientas. El cuadro *Paleta de relleno* contiene distintos efectos y tramas aplicables a cualquier objeto del gráfico.

Mediante el menú *Insertar*, podemos añadir más gráficos superpuestos al que tenemos, así como textos, títulos, etc. Son exactamente las mismas opciones que nos proporcionaba el botón  de la barra de *Utilidades*.

El menú *Formato* nos proporciona el acceso a las propiedades de todos los objetos del gráfico para aplicarles el formato que nos interese. En primer lugar tenemos las propiedades generales de los gráficos. El cuadro que nos muestra tiene multitud de opciones:

- ♦ **Colores:** permite especificar la secuencia de colores de una leyenda categórica o el rango de colores de una leyenda de escala. Se puede reordenar la lista o asignar cualquier color a los que aparecen por defecto.
- ♦ **Estilos:** de la misma forma que antes, podemos especificar la secuencia de estilos de relleno de una leyenda categórica.
- ♦ **Tamaños:** esta pestaña permite modificar la secuencia de los tamaños de los símbolos y las especificaciones de grosor de línea que se aplicarán siempre que existan leyendas de tamaño.
- ♦ **Texto:** a través de la lista desplegable se selecciona el texto del gráfico y se modifica su formato, fuente, tamaño, alineación, etc.
- ♦ **Objetos rellenos:** también a través de una lista, accederemos al objeto susceptible de relleno (es obvio que no se pueden rellenar líneas, pero sí el marco de las leyendas de texto, por ejemplo). Además admite el cambio de formato del borde del objeto.
- ♦ **Ejes:** desde esta pestaña es sencillo cambiar cualquier detalle de los ejes, como el estilo de la línea o las marcas de las categorías.
- ♦ **Líneas y símbolos:** sirve para alterar las propiedades de las líneas, los símbolos y los conectores de las barras de error, de los gráficos de regresión, de las nubes, etc. Sin embargo, los cambios que hagamos no se aplicarán a los objetos que resulten afectados por asignaciones de variables. Por ejemplo, si se ha asignado una variable a la leyenda de estilo, los símbolos o las líneas adoptarán un estilo de acuerdo con la leyenda, en lugar del estilo asignado aquí.
- ♦ **Líneas de cuadrícula:** establece líneas guía según los valores de las escalas o espaciadas uniformemente.

La opción del menú *Aspectos de gráfico* nos muestra un cuadro con los aspectos predefinidos que SPSS mantiene almacenados en disco. Éstos se encuentran en el directorio “\Looks” y su extensión es “*.clo”. Si seleccionamos uno cualquiera y pulsamos en *Editar*, pasaremos al cuadro de propiedades del gráfico y podremos modificarlo. Posteriormente lo almacenaremos en disco para su posterior utilización en otros diseños. También podemos crear uno totalmente nuevo y guardarlo a través de *<Propiedades del gráfico actual>*.

Cuando seleccionamos un texto, su formato se encontrará accesible desde la opción *Texto*. En caso de que se trate de un número, también podremos especificar una plantilla de fecha, de hora, de moneda, general, etc.

Las leyendas y las claves pueden tener un marco, al que le podremos modificar las propiedades del borde, los márgenes y el relleno.

El siguiente bloque del menú *Formato* es el dedicado a la configuración de los elementos que no dependen del tipo de gráfico, sino que son comunes en general. Son muchas las características modificables de los ejes: la escala, el aspecto visual, las etiquetas, el título de los ejes, las líneas de cuadrícula y de referencia, etc. Asimismo podemos cambiar el formato de leyendas, claves, establecer líneas de cuadrícula y establecer el formato de la región de datos. Este último permite cambiar el tamaño y el relleno de la región comprendida entre los ejes.


La última opción del menú *Formato* nos permite acceder rápidamente a cualquier elemento del gráfico.

Antes de terminar con esta parte debemos considerar que los gráficos interactivos son posibles debido a la permanente existencia de los datos. Quiere decir esto que si guardamos el contenido del *Visor*, los datos se desvinculan del gráfico, y cuando recuperemos el archivo de resultados (*.spo) será imposible volver a modificarlos. Para evitarlo, debemos activar la opción *Guardar datos con el gráfico* que se encuentra en la pestaña *Interactivos* del cuadro *Opciones* del menú *Edición*. De esta forma es posible ejecutar la mayoría de funciones interactivas disponibles para los gráficos anexados al archivo de datos que los creó (exceptuando la adición de variables no incluidas en el gráfico original). La única consideración que debemos tener es que el archivo del *Visor* puede aumentar considerablemente de tamaño.

5.2. Visor borrador

SPSS nos ofrece otra forma de obtener resultados, y es a través del *Visor borrador*. Este visor es similar al anterior pero mucho más simple. Es útil cuando queremos obtener resultados que usaremos en otras aplicaciones sencillas, o que sólo utilicen texto, ya que el *Visor borrador* convierte los resultados en textos e imágenes.

Concretando un poco, las posibilidades de alineación de los textos han desaparecido. Las tablas pivote se forman a base de símbolos de texto y por supuesto no son pivotables. Los gráficos de cualquier tipo ahora son imágenes estáticas sin posibilidad de modificación, tan solo podremos cambiar su tamaño. De esta forma el archivo sólo contiene elementos con formato RTF (Rich Text Format) fácilmente exportables a cualquier aplicación.

Si queremos utilizar este visor, debemos abrir una ventana de este tipo en *Archivo / Nuevo / Resultados de borrador*, y a partir de ese momento todos los resultados aparecerán en ella hasta que se designe otra como ventana activa (). Otra forma sería indicar, en el menú de opciones generales, que todos los resultados aparezcan por defecto en el *Visor borrador*.

Conceptualmente no cambia mucho del *Visor*, la diferencia estriba en que es mucho más sencillo. Para empezar, el menú *Archivo* contiene casi las mismas funciones, sólo que ahora al abrir o guardar un archivo, nos indicará la extensión RTF. El menú *Edición* también contiene las funciones usuales, además de otras que nos permiten desplazarnos por los elementos y procedimientos que tenemos en el *Visor borrador* ya que en éste no dispone del panel de titulares. Los menús *Ver*, *Insertar* y *Formato* se han reducido a su mínima expresión, y contienen funciones suficientemente conocidas.

De la misma forma, las barras de herramientas contienen botones de consabido funcionamiento.

Por último, y para mayor comodidad, se ha incluido en el menú contextual la opción de exportar, o bien todo el texto, o bien la parte seleccionada del mismo.

6. PERSONALIZACIÓN DE SPSS. PROCESAMIENTO

6.1. Cuadro general de opciones

Este cuadro se encuentra accesible desde cualquier ventana o editor a través del menú *Edición / Opciones*. Nos posibilita la configuración de SPSS a nivel general y se encuentra distribuido en pestañas de acceso a diferentes fichas. Las modificaciones que realicemos aquí, por norma general, causarán efecto cuando el sistema deba volver a utilizarlas, no antes.

Descripción de las opciones más importantes:

♦ **Ficha General**

El *diario de la sesión* registra todos los comandos ejecutados en una sesión, tanto en ventanas de sintaxis como en cuadros de diálogo. Este archivo se puede editar y reutilizar.

El *sistema de medida* utilizado puede basarse en puntos, pulgadas o centímetros.

Los usuarios avanzados que utilicen comandos usualmente pueden establecer que se abra una ventana de sintaxis automáticamente al inicio de cada sesión.

Las *listas de variables* controlan la presentación de las mismas y su orden en los cuadros de diálogo.

El *tipo de resultados al arrancar* controla el tipo de visor que se abre en primera instancia.

La *notificación de resultados* es la forma de indicar al usuario que el procedimiento ha terminado y los resultados están listos en el visor.

♦ **Ficha Visor**

Controla los elementos que se muestran y se ocultan automáticamente cada vez que se ejecuta un procedimiento, además de la alineación inicial de los elementos, pero sólo en la impresión, no en pantalla.

La *fuentes del título* cambia el estilo, tamaño y color.

El *tamaño de página de los resultados de texto* controla el ancho de y largo de página expresado en caracteres y líneas respectivamente.

La *fuentes de los resultados de texto* también podemos modificarla, pero usando siempre las que posean un ancho de letra homogéneo.

♦ **Ficha Visor de borrador**

Mostrar elementos de resultados determina los elementos que deben mostrarse automáticamente cada vez que se ejecute un procedimiento.

Se pueden insertar saltos de página entre resultados, así como cambiar la fuente por defecto.

Resultados tabulares controla las opciones de los resultados de las tablas pivotadas convertidos a resultados tabulares.

En los *resultados de texto* que no sean de tablas pivotadas convertidos se puede establecer el ancho y la longitud de página.

♦ **Ficha Etiquetas de los resultados**

Controlan la visualización de la información sobre variables y valores de datos en las tablas pivotadas y en los titulares. Puede visualizar nombres de variable, etiquetas de variable definidas y valores de datos reales, etiquetas de valor definidas, o una combinación de éstas.

♦ **Ficha Gráficos**

Los archivos nuevos pueden utilizar las opciones establecidas en *Especificaciones actuales* o las de una plantilla.

La *relación de aspecto de los gráficos* indica la proporción ancho-alto del marco exterior de los nuevos gráficos.

Las *especificaciones actuales* controlan la fuente de los textos en los gráficos, las tramas de relleno y estilos de línea, los marcos internos y externos, y la visualización de las líneas de cuadrícula.

Todo esto no afecta a los gráficos interactivos.

♦ **Ficha Interactivos**

Permite establecer el *aspecto del gráfico* inicial y las *unidades de medida* usadas.

En el apartado de gráficos interactivos ya vimos las ventajas e inconvenientes de guardar datos con el gráfico.

También es posible determinar la *resolución de impresión* de estos gráficos.

♦ **Ficha Tablas pivotadas**

Igual que en los gráficos, permite establecer un *aspecto de tabla* por defecto.

Se puede ajustar automáticamente el ancho de las columnas en las tablas pivotadas.

El *modo de edición por defecto* establece la forma de edición de las tablas, bien en una nueva ventana de edición, bien en la ventana del visor donde se encuentra.

♦ **Ficha Datos**

Las *opciones de transformación y fusión* ya vimos qué ventajas aportaba en el apartado que describía el menú *Transformar*.

El formato de *visualización para nuevas variables numéricas* controla la visualización por defecto del ancho y el número de posiciones decimales de las nuevas variables numéricas. No hay formato por defecto para las de cadena.

Definir rango de siglo para años de dos dígitos hace referencia a las variables con formato de fecha introducidas o mostradas con un año de dos dígitos. Buena solución para mantener dos dígitos aparentes en las fechas tras el pánico del Y2K.

♦ **Ficha Moneda**

Pueden crearse hasta cinco formatos de visualización de moneda personalizados que pueden incluir caracteres de prefijo y sufijo especiales, tratamiento especial para los *valores negativos*, e indicación del *separador decimal* a usar.

♦ Ficha *Procesos*

Especifica el *archivo de procedimientos global*, el *archivo de autoproceso*, y las subrutinas de autoproceso a utilizar.

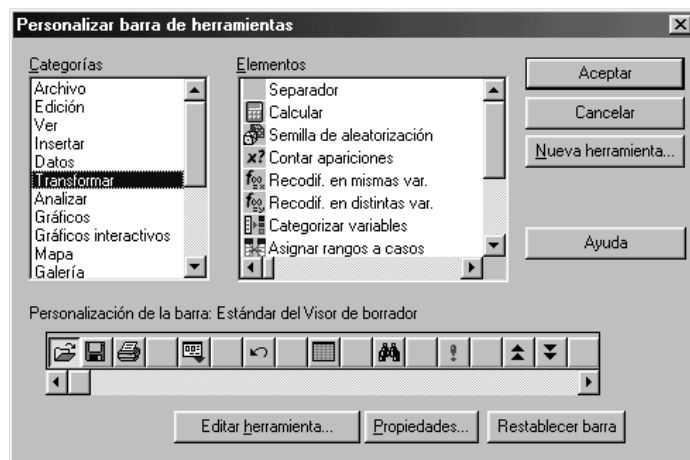
Un *archivo de procedimientos globales* es una biblioteca de subrutinas de proceso y funciones que pueden ser invocadas por archivos de proceso, incluyendo los archivos de autoproceso. Estos términos se explican en el apartado de *Procesamiento*.

En cuanto al resto, se concreta en el apartado *Autoprocesamiento*.

6.2. Personalización de las barras de herramientas

Para personalizar las barras de botones de cualquier ventana hemos de acceder a *Ver / Barras de herramientas*. Dependiendo del tipo de documento que seleccionemos, nos mostrará las barras asociadas a dicha ventana, que son susceptibles de mostrar u ocultar. Podemos crear una nueva barra asignándole un nombre y posteriormente personalizarla. O bien, podemos elegir una de las barras que posee SPSS y modificarla. También podemos cambiar el tamaño de los botones (cuadruplicando su tamaño) o eliminar las pistas que nos ofrecen cuando detenemos el puntero del ratón unos segundos sobre cualquiera de ellos.

Una vez en el cuadro de personalizar, hemos de elegir la categoría global de herramientas. En la lista de la derecha nos mostrará los botones asociados a esa categoría, y únicamente debemos arrastrarlo a la barra modelo que aparece en la parte inferior para que se inserte.



Con *Editar herramienta* se abre una nueva ventana capaz de editar un pequeño icono de tipo BMP de 16x16 puntos con 16 colores (tamaño de un botón estándar). Una vez acabado el diseño seleccionamos *Aplicar y salir*, con lo que ya tendremos el aspecto del botón a nuestro gusto.

La opción *Propiedades* permite asignar la barra que estemos editando a las ventanas que queramos. De esta forma, cuando nos encontremos en una de esas ventanas, dicha barra se mostrará por defecto.

Si queremos crear una nueva herramienta debemos darle un nombre (etiqueta) e indicar la acción a realizar, que puede ser ejecutar una aplicación, un archivo de sintaxis, o un proceso. Por ejemplo, seleccionamos *Aplicación*, indicamos datos de tipo XLS y en el nombre de archivo ponemos "Excel.exe". Nos aparecerá una nueva herramienta con el nombre que le hayamos dado en la categoría *Personales* (que es donde se guardan las creadas por el usuario), y ya sólo queda arrastrar dicho botón a la barra que estemos editando.

Siempre podemos restablecer la barra original si la barra creada no se ajusta a nuestras necesidades o ya le hemos dado el uso que pretendíamos.

6.3. Personalización de las barras de menús

Para crear elementos de menú personales hemos de seleccionar el *Editor de menús* en el menú *Utilidades*. Ahora elegimos la barra de menú que queremos modificar y nos aparecerá en el panel de la izquierda una lista jerárquica con todas las opciones de dicha barra. Seleccionamos la posición donde insertar el nuevo elemento y lo definimos. Este elemento ejecuta una aplicación, una sintaxis o un proceso, de la misma forma que vimos al crear un nuevo botón en una barra de herramientas.

También podemos insertar un nuevo menú o un separador en cualquier posición de la lista, es decir, podríamos crear un menú dentro de otro menú.

El símbolo **&** indica que el carácter siguiente aparecerá subrayado y generará un acceso rápido mediante la combinación de teclas [Alt] y dicho carácter.

6.4. Procesamiento

SPSS dispone de una utilidad de procesamiento para automatizar tareas que se realicen con frecuencia. Tareas como personalización de resultados, apertura y almacenamiento de archivos, trabajo con cuadros de diálogo, transformación de datos, procedimientos estadísticos, exportación de gráficos, etc.

El sistema incluye una serie de procesos de muestra que pueden servir para iniciarse en su creación y edición. Algunos ejemplos son:

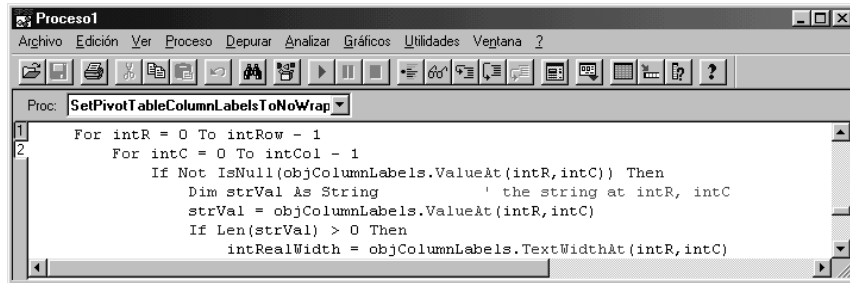
- ♦ **Limpiar navegador:** elimina las tablas *Notas* de un documento de resultados.
- ♦ **Totales en negrita:** aplica negrita y azul a cualquier fila o columna con la etiqueta *Total* en una tabla pivote.
- ♦ **Borrar por etiqueta:** elimina filas o columnas de una tabla pivote basándose en el contenido de las etiquetas de fila o columna.

Los procesos se ejecutan a través del menú *Utilidades*. Dependiendo de la función que realice el proceso, es usual que debamos seleccionar previamente el objeto al que queremos que se aplique dicha función (por ejemplo una tabla pivote).

Para crear un proceso, nos dirigimos al menú *Archivo / Nuevo / Proceso*. Nos aparecen una serie de procesos de inicio (directorio "\\Starter") que podemos usar, o bien pulsamos

cancelar y partiremos de cero. Los procesos tienen la extensión “*.sbs” y suelen contener una breve descripción en el panel de la derecha.

La ventana que nos aparece es el *Editor de procesos*. Nos encontramos en un entorno de programación completo que utiliza el lenguaje Sax Basic, compatible con Visual Basic for Applications™. Incluye un editor de cuadros de diálogo, un visor de objetos, funciones de depuración y ayuda sensible al contexto. Los usuarios habituados a la programación pueden encontrar muy útil la definición de este lenguaje ubicada en la ayuda del editor.



6.5. Autoprocesamiento

El concepto de autoproceso es similar al de proceso con una ligera diferencia: se ejecuta automáticamente siempre que se produzca determinado tipo de objeto de resultado. Por ejemplo, podemos definir un autoproceso que aumente en 2 puntos el tamaño de fuente de la primera fila de una tabla pivote. Cada vez que se genere una tabla pivote, automáticamente se ejecutará el autoproceso y la modificará.

Para usar los autoprocesos hemos de ir al menú de opciones generales (menú *Edición*) y a la pestaña *Procesos*. Existe una casilla denominada *Permitir el autoprocesamiento* que hemos de activar. Elegimos el archivo de autoproceso y nos aparecen en la lista inferior las subrutinas que contiene (cada archivo puede contener varias). Sólo resta activar las subrutinas que queramos para que actúen automáticamente.

Para crear o editar un autoproceso, hemos de situarnos en el *Visor* y seleccionar el objeto al que aplicar el autoprocesamiento, que actuará como desencadenante del mismo. En el menú *Utilidades* tenemos la opción *Crear/Editar autoproceso* (también en el menú contextual), que nos abrirá el *Editor de procesos* para añadir subrutinas al archivo actual de autoprocesos.

7. PRODUCCIÓN AUTOMATIZADA

SPSS posee la capacidad de funcionar de forma totalmente autónoma y automatizada. Para realizar un trabajo de producción automatizado hemos de elaborar un archivo de sintaxis de comandos (es un archivo con formato de texto). Disponemos de un *Editor de sintaxis* que se abre cuando pulsamos en *Archivo / Nuevo / Sintaxis*, o podemos usar el botón pegar de cualquier cuadro de diálogo. Estos archivos poseen la extensión “*.sps”.

Ejemplo de un archivo de sintaxis de comandos:

```
EXAMINE
VARIABLES=salario educ /COMPARE VARIABLE/PLOT=BOXPLOT/STATISTICS=NONE
/NOTOTAL
/MISSING=LISTWISE .
```

Este archivo realiza diagramas de caja simples para las variables *salario* y *educ*.

Una vez disponemos del archivo de sintaxis, arrancamos la *Utilidad de producción* desde Windows, que se encuentra como un módulo aparte del programa.



Sólo es necesario añadir archivos de sintaxis al trabajo de producción y pulsar el botón de ejecución: . Desde este mismo cuadro podemos decidir si queremos que se muestren los resultados en el visor normal o en el de borrador, o bien que se impriman al terminar.

Los archivos de trabajos de producción se pueden guardar y se pueden abrir desde el disco. Poseen extensión “*.spp”.

8. AYUDA

8.1. Temas

La ayuda de SPSS se encuentra en el menú ?. Desde ahí accedemos a los *Temas*, donde se incluye toda la ayuda disponible en formato HLP de Windows. Nos muestra un panel con tres fichas, la de *Contenido*, la de *Índice* y la de *Buscar*.

El *Contenido* se halla distribuido en libros que, al abrirlos, nos muestran más libros o temas de ayuda a exponer.

El *Índice* contiene una lista por la cual se indexa la ayuda. Introduciendo las primeras letras del término que buscamos, nos irán apareciendo objetivos de ayuda en la lista inferior.

Buscar genera una base de datos con todas las palabras existentes en la ayuda, de forma que podamos encontrar los términos más recónditos. Eso sí, para maximizar las capacidades de búsqueda se requiere recopilar una ingente cantidad de datos que puede tener un elevado coste temporal.

Cualquier tema que visualicemos es susceptible de ser copiado, transportado a otra aplicación, impreso, etc.

Uno de los libros de la ayuda es el de *Tutoriales*. Éste nos muestra temas de ayuda en un formato diferente. Se trata de la explicación progresiva de un tema expuesta de una forma simple y didáctica.

8.2. Guía de sintaxis

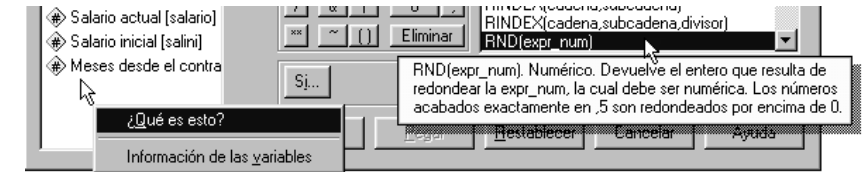
Ya sabemos que SPSS funciona internamente a base de comandos. La mayoría de funciones se encuentran desarrolladas mediante la interfaz de ventanas, pero existen algunas que sólo pueden ser implementadas mediante comandos.

La descripción de la sintaxis de comandos se encuentra en ficheros con formato PDF de Adobe Acrobat. Para poder verlos necesitamos un visor de este formato (Acrobat Reader, de libre distribución). Los ficheros se distribuyen por módulos del sistema: *Base*, *Regression Models*, *Categories*, etc.

8.3. Ayuda contextual

La ayuda contextual podemos obtenerla de varias formas, según donde nos encontremos.

En los cuadros de diálogo debemos pinchar con el botón derecho del ratón sobre cualquier elemento para obtener una descripción del mismo. Dependiendo del elemento del que se trate, nos puede mostrar un pequeño cuadro con información sobre el mismo, o un menú contextual con la opción “¿Qué es esto?”:



En las tablas pivote debemos obtener el menú contextual sobre la cabecera de una columna o fila, y pulsar sobre la opción “¿Qué es esto?” para que nos muestre una descripción de los valores contenidos en esa columna o fila.

Por último, todos los cuadros de diálogo contienen un botón de ayuda que nos ofrece una descripción de la utilidad general de cuadro y particular de cada elemento sito en él.

8.4. Asesor de resultados

Este asesor se encarga de presentar ayuda sobre el tipo de objeto elegido. Por ejemplo, si tenemos una tabla pivote en el visor de resultados, la editamos y mostramos la barra de herramientas, nos aparecerá un botón (🔍) que nos presentará una ventana explicativa a modo de tutorial del tipo de resultados que nos ofrece la tabla.

8.5. Asesor estadístico

Se encuentra en el menú de ayuda. Básicamente nos pregunta qué queremos hacer y nos dice cómo hacerlo.

La primera pantalla que nos presenta es para solicitarnos el objetivo que perseguimos como, por ejemplo, crear informes, resumir datos, identificar relaciones, etc. Seleccionando una de estas opciones nos mostrará un ejemplo de lo que obtendremos, y pulsando en *Más ejemplos*, nos presentará otros resultados posibles. Con el botón de ayuda la descripción de las opciones es más completa.

Cuando tengamos claro lo que queremos obtener, pulsaremos en *Siguiente*. Ahora nos preguntará el tipo de datos de los que disponemos, que pueden ser ordinales, categóricos, numéricos continuos, etc., o su estructura. Según la opción elegida en primera instancia, podrá solicitarnos alguna cosa más como el tipo de representación que deseamos, o tipo de resumen, o tipo de gráfico, etc.

Una vez contestadas todas las preguntas que, como podremos observar no tienen un nivel muy técnico, pulsaremos en *Finalizar* y nos presentará un resumen con lo que debemos hacer para obtener los resultados deseados. Además abrirá automáticamente el cuadro de diálogo correspondiente al procedimiento que debemos emplear para obtener dichos resultados.

◆◆◆ TABLA DE CONTENIDO ◆◆◆

1.	<u>INTRODUCCIÓN</u>	2
1.1.	Descripción del producto	2
1.2.	Ejecución	3
1.3.	Forma de trabajar: pasos básicos	4
1.3.1.	Obtención de datos	4
1.3.2.	Selección de procedimiento	4
1.3.3.	Selección de variables y opciones	4
1.3.4.	Examen de resultados	4
1.4.	Ventanas	5
1.5.	Barra de menús	5
1.6.	Barras de herramientas	6
1.7.	Ayuda	6
2.	<u>OBTENCIÓN DE DATOS</u>	8
2.1.	Matriz de datos	8
2.2.	Definición de variables	8
2.3.	Introducción de datos	9
2.4.	Uso del Editor de datos y variables	9
2.5.	Menú Archivo	10
2.6.	Menú Edición	11
2.7.	Menú Ver	11
2.8.	Uso de los cuadros de diálogo: selección de variables y opciones	11
2.9.	Reclamada de los cuadros de diálogo	13
2.10.	Menú Datos	13
2.11.	Menú Transformar	15
2.12.	Menú Utilidades	18
3.	<u>ANÁLISIS</u>	19
3.1.	Informes	19
3.2.	Estadísticos descriptivos	20
3.3.	Tablas personalizadas	20
3.4.	Comparar medias	22
3.5.	Modelo lineal general	22
3.6.	Correlaciones	24
3.7.	Regresión	25
3.8.	Loglineal	27
3.9.	Clasificar	28
3.10.	Reducción de datos	29
3.11.	Escalas	30
3.12.	Pruebas no paramétricas	31
3.13.	Series temporales	32
3.14.	Supervivencia	33
3.15.	Respuestas múltiples	34
4.	<u>GRÁFICOS</u>	36
4.1.	Galería de gráficos	36
4.2.	Creación de un gráfico	37
4.2.1.	Barras	37
4.2.2.	Líneas	38
4.2.3.	Áreas	39
4.2.4.	Sectores	39
4.2.5.	Máximos y mínimos	39
4.2.6.	Pareto	40
4.2.7.	Control	41
4.2.8.	Diagramas de caja	41
4.2.9.	Barras de error	42

4.2.10.	Dispersión	42
4.2.11.	Histograma	42
4.2.12.	P-P	43
4.2.13.	Q-Q	43
4.2.14.	Secuencia	43
4.2.15.	Curvas COR	43
4.2.16.	Serie temporal: autocorrelaciones	44
4.2.17.	Serie temporal: correlaciones cruzadas	44
4.2.18.	Serie temporal: análisis espectral	44
4.3.	Galería de gráficos interactivos	45
4.4.	Creación de gráficos interactivos	45
4.4.1.	Gráficos de barras interactivos	47
4.4.2.	Gráficos de puntos interactivos	48
4.4.3.	Gráficos de líneas interactivos	49
4.4.4.	Gráficos de bandas interactivos	49
4.4.5.	Gráficos de líneas verticales interactivos	50
4.4.6.	Gráficos de áreas interactivos	50
4.4.7.	Gráficos de sectores simples interactivos	50
4.4.8.	Gráficos de sectores agrupados interactivos	51
4.4.9.	Gráficos de sectores dispersados interactivos	51
4.4.10.	Diagramas de caja interactivos	51
4.4.11.	Barras de error interactivas	52
4.4.12.	Histogramas interactivos	52
4.4.13.	Diagramas de dispersión interactivos	53
4.5.	Creación de gráficos según la estructura de los datos	54
4.5.1.	Gráficos para variables categóricas	54
4.5.2.	Gráficos para variables continuas	54
4.5.3.	Gráficos para casos individuales	55
4.5.4.	Gráficos especializados	55
5.	<u>EXAMEN DE RESULTADOS</u>	56
5.1.	Uso del visor	56
5.1.1.	Panel de titulares y panel de resultados	57
5.1.2.	Textos	58
5.1.3.	Tablas pivote	58
5.1.4.	Gráficos	60
5.1.5.	Gráficos interactivos	61
5.2.	Visor borrador	65
6.	<u>PERSONALIZACIÓN DE SPSS. PROCESAMIENTO</u>	67
6.1.	Cuadro general de opciones	67
6.2.	Personalización de las barras de herramientas	69
6.3.	Personalización de las barras de menús	70
6.4.	Procesamiento	70
6.5.	Autoprocesamiento	71
7.	<u>PRODUCCIÓN AUTOMATIZADA</u>	72
8.	<u>AYUDA</u>	73
8.1.	Temas	73
8.2.	Guía de sintaxis	73
8.3.	Ayuda contextual	73
8.4.	Asesor de resultados	74
8.5.	Asesor estadístico	74