

APUNTES DE SPSS

1.- TIPOS DE FICHEROS.

En primer lugar tenemos los ficheros de datos propios de SPSS. Siempre tienen extensión *.sav, y su contenido aparece en la ventana del editor de datos.

El haber incluido el término "propios" en la denominación anterior obedece a dos razones:

Primero, porque SPSS es capaz de leer (importar) ficheros de datos de muy diversas procedencias y formatos.

Asimismo, a la hora de grabar un fichero de datos se puede escoger también el formato en que dicha grabación se va a efectuar de cara a la exportación de la información a otros entornos o programas.

Pero, en cualquier caso, sea vía importación o creación ex profeso mediante el editor de datos del sistema, un fichero propio es aquél cuyo formato es el específico de SPSS y que contiene no sólo los datos sino también la información que el sistema necesita para procesarlos.

A este conjunto de información sobre los datos, en la terminología de SPSS se le asigna el nombre de diccionario de datos (conjunto de características y opciones relativas a la definición de variables y datos).

Un segundo tipo de ficheros son los del navegador de resultados. Su extensión es *.spo y contienen gráficos, tablas, análisis de datos, etc. Su contenido es mostrado en la ventana del navegador de resultados.

Un tercer tipo de ficheros son los de sintaxis, cuya extensión es *.sps. Contienen instrucciones de SPSS escritos en el lenguaje de mandatos propio del sistema. Son tanto ficheros de entrada como de salida en la medida en que, si bien su uso es más habitual es el de ser cargados y leídos por el sistema para ejecutar un programa completo en proceso por lotes, también son generados por el propio sistema cuando en los diferentes cuadros de diálogo, en vez de ejecutar el procedimiento directamente, se pega la instrucción subyacente.

En este último caso se estará empleando SPSS como un interfaz auxiliar de generación de programas. Su contenido se muestra en la ventana de sintaxis.

Por último están los archivos de procesos (también llamados macros en otros programas) cuya extensión es *.sbs.

2.- TIPOS DE VENTANAS.

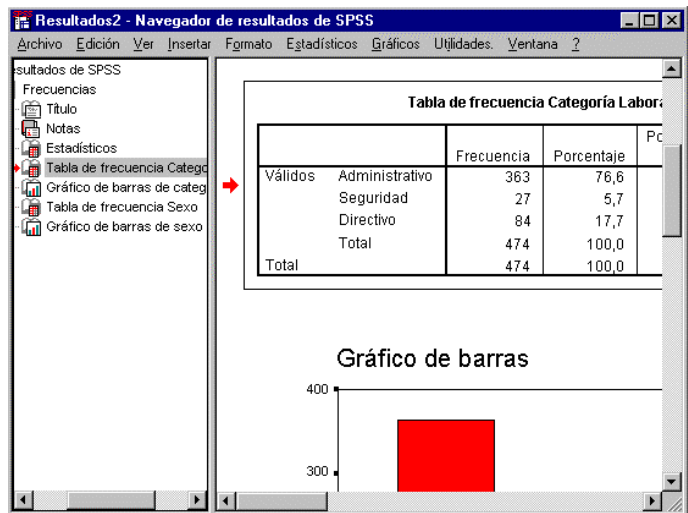
Existen siete tipos de ventanas en SPSS:

Editor de datos: Esta ventana muestra los contenidos del archivo de datos. Con el Editor de datos se pueden crear nuevos archivos de datos o modificar los existentes. La ventana del Editor de datos se abre automáticamente cuando se inicia una sesión de SPSS. No se puede tener más de un archivo de datos abierto al mismo tiempo.

Navegador de resultados: Todos los resultados estadísticos, tablas y gráficos se muestran en el Navegador de resultados. Se pueden editar los resultados y guardarlos para utilizarlos posteriormente. La ventana del Navegador de resultados se abre automáticamente la primera vez que se ejecuta un procedimiento que genera resultados.

	id	sexo	fechnac	educ	catlab	salario	salini	tiemp
1	1	h	02/03/52	15	3	\$57,000	\$27,000	
2	2	h	05/23/58	16	1	\$40,200	\$18,750	
3	3	m	07/26/29	12	1	\$21,450	\$12,000	
4	4	m	04/15/47	8	1	\$21,900	\$13,200	
5	5	h	02/09/55	15	1	\$45,000	\$21,000	
6	6	h	08/22/58	15	1	\$32,100	\$13,500	

Editor de tablas pivote: Los resultados que se muestran en las tablas pivote se pueden modificar de muchas formas con el Editor de tablas pivote. Se puede editar texto, intercambiar datos en filas y columnas, añadir color, crear tablas multidimensionales y ocultar y mostrar resultados de manera selectiva.

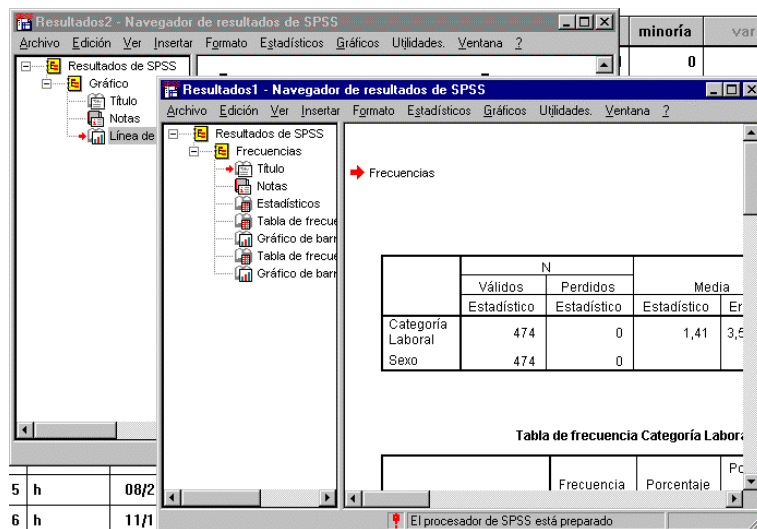


Editor de gráficos: Se pueden modificar los gráficos de alta resolución y los diagramas en las ventanas de gráficos. Se pueden cambiar los colores, seleccionar diferentes tipos de fuentes y tamaños, intercambiar los ejes horizontal y vertical, rotar diagramas de dispersión 3-D e incluso cambiar el tipo de gráfico.

Editor de resultados de texto: Los resultados de texto que no se muestran en las tablas pivote se pueden modificar con el Editor de resultados de texto. Se pueden editar los resultados y cambiar las características de las fuentes (tipo, estilo, color, tamaño).

Editor de sintaxis: Se pueden pegar las elecciones del cuadro de diálogo en una ventana de sintaxis, donde las selecciones aparecerán en forma de sintaxis de comandos. Luego se puede editar la sintaxis de comandos para utilizar funciones especiales de SPSS que no están disponibles mediante cuadros de diálogo. Se pueden guardar estos comandos en un archivo para utilizarlos en sesiones de SPSS posteriores.

Editor de procesos: El procesamiento y la automatización OLE le permiten personalizar y automatizar muchas tareas en SPSS. Utilice el Editor de procesos para crear y modificar procesos en Basic.



VENTANA DESIGNADA FRENTE A VENTANA ACTIVA

Si tiene abierta más de una ventana del Navegador de resultados, los resultados se dirigen hacia la ventana designada del Navegador de resultados. Si tiene abierta más de una ventana del Editor de sintaxis, la sintaxis de comandos se pega en la ventana designada del Editor de sintaxis. Las ventanas designadas se indican por un signo de exclamación (!) en la barra de estado. Puede cambiar las ventanas designadas en cualquier momento.

La ventana designada no debería confundirse con la ventana activa, que es la ventana actualmente seleccionada. Si tiene ventanas superpuestas, la ventana activa aparecerá en primer plano. Si abre una nueva ventana del Editor de sintaxis o del Navegador de resultados, esa ventana se convertirá automáticamente en la ventana activa y la ventana designada.

Para cambiar la ventana designada del Navegador de resultados o del Editor de sintaxis:

Convierta la ventana que desee designar en la ventana activa (pulse en cualquier parte de la ventana).

Pulse en la herramienta Ventana de designación en la barra de herramientas (la que tiene el signo de exclamación).

O Elija en los menús: Utilidades / Designar ventana

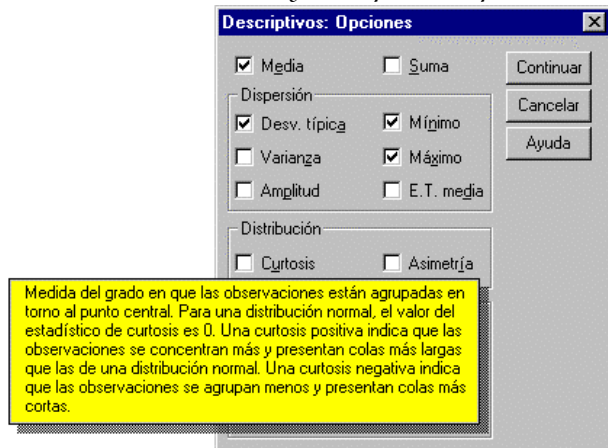
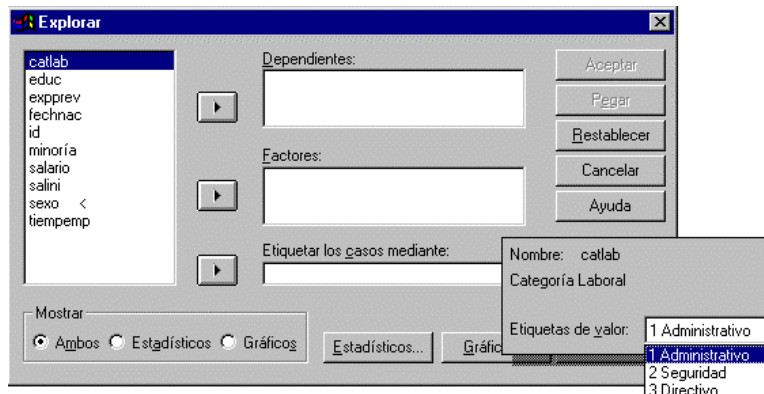
CUADROS DE DIÁLOGO

Los cuadros de diálogo se utilizan para seleccionar variables y opciones de estadísticos y gráficos.

Las variables para análisis se seleccionan en la lista de origen. Y se utiliza el botón de flecha para mover las variables a la lista de destino.

Para obtener más información sobre una variable, pulse con el botón izquierdo del ratón en el nombre de la variable para seleccionarla. A

continuación, pulse con el botón derecho del ratón en cualquier lugar de la lista y seleccione Información sobre la variable en el menú contextual emergente. Una ventana emergente muestra el nombre de la variable y cualquier etiqueta descriptiva definida para la variable y los valores.



Los botones de cuadros de diálogos con puntos suspensivos (...) abren subcuadros de diálogo para selecciones opcionales.

Para obtener información sobre cualquiera de los controles de un cuadro de diálogo, pulse con el botón derecho del ratón sobre el control que desea conocer.

Seleccione ¿Qué es esto? en el menú contextual emergente. Una ventana emergente muestra información sobre el control.

3.- DEFINICIÓN DE LAS VARIABLES.

Si en el menú de Archivo abrimos un nuevo fichero de datos (Archivo / Nuevo / Datos) en la pantalla se abre la ventana del editor de datos.

La parte fundamental de esta ventana es la tabla de doble entrada que aparece y que es el fichero de datos vacío.

Este tipo de ficheros, en SPSS siempre tienen la estructura de una matriz rectangular. En la misma las filas representan los casos o sujetos, y las columnas las variables; o si se quiere emplear una terminología más propia de la informática, hállese de registros y campos, respectivamente.

Las dimensiones de la matriz vienen determinadas por el número de filas y columnas, y dentro de estos límites no hay casillas vacías.

El afirmar que no pueden existir casillas vacías en la matriz no significa que para todos los sujetos y todas las variables haya que disponer de valores reales. De hecho, es relativamente frecuente que en algunos casos tal carencia de información se produzca.

Lo que ocurre es que SPSS, automáticamente, asigna a los blancos en la matriz el código específico de los valores ausentes definidos por el sistema en el caso de las variables numéricas.

Para las variables alfanuméricas, los espacios en blanco que haya en una casilla son considerados como un valor más de la variable y no como carencia de información.

Antes de proceder a rellenar esa matriz con los datos, es preciso definirla, comunicar al sistema la estructura de la misma. Por ejemplo, de cuántas variables va a constar, el nombre de éstas, el tipo y la longitud o número máximo de caracteres de cada una de ellas, etc... Y para tal fin debe procederse a lo que SPSS denomina definición de las variables.

DEFINIR VARIABLE asigna información sobre las definiciones de datos a las variables. Se pueden definir variables nuevas o cambiar las definiciones de variables existentes. La información sobre la definición de datos incluye:

- Nombre de variable
- Tipo de datos (numérico, de cadena, fecha, etc.)
- Etiquetas descriptivas de variable y de valor
- Códigos especiales para valores perdidos

DEFINIR TIPO DE VARIABLE especifica los tipos de datos de cada variable. Por defecto, se asume que todas las variables nuevas son numéricas. Se puede utilizar Definir tipo de variable para cambiar el tipo de datos. El contenido del cuadro de diálogo Definir tipo de variable depende del tipo de datos seleccionado. Para algunos tipos de datos, hay cuadros de texto para el ancho y el número de decimales; para otros, simplemente puede seleccionar un formato de una lista desplegable de ejemplos.

Los tipos de datos disponibles son numéricos, de coma, de punto, notación científica, fecha, dólar, moneda personalizada y de cadena.

Variables numéricas: Son aquellas que admiten cualquier número como valor válido, pudiendo ir precedido del signo más o menos. Su longitud máxima es de 40 caracteres, de los cuales 16 pueden ser cifras decimales. La coma debe teclearse siempre que haya parte decimal.

Variables con coma: Este formato de entrada de datos añade al anterior la coma como separador de los miles y el punto como separador de la parte entera y de la decimal.

Variables con punto: En este caso se emplea el punto como separador de los miles y la coma como separador de la parte entera y de la decimal.

Si, como es previsible, la versión de Windows que usted tiene está en castellano, probablemente mediante el panel de control se haya establecido la configuración del formato numérico, empleando el punto como separador de los miles y la coma como separador decimal.

Variables numéricas en notación científica: En este caso, son valores aceptables todos los numéricos más las letras "D" o "E" y los signos más o menos. Por ejemplo 347D2, 347E2, 347+2, 347-2.

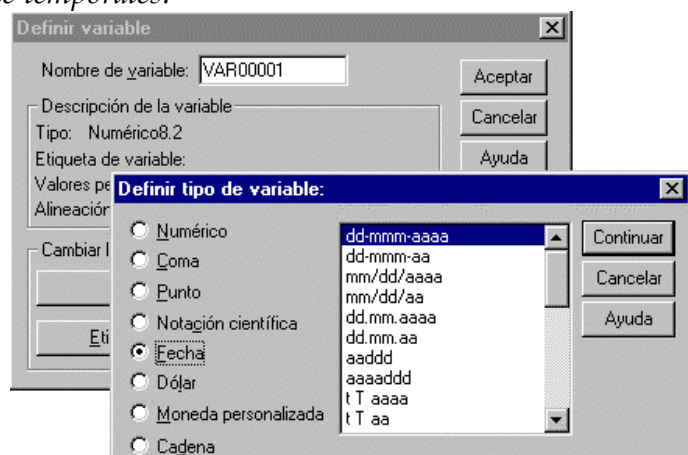
Variables tipo fecha: Mediante este formato es posible introducir las variables temporales. Si este tipo es seleccionado en el cuadro de diálogo, se abre una ventana, con a su vez, todos los formatos aceptables.

Si se observa la lista, es posible hacer una diferenciación importante: hay formatos para variables de fecha y otros para las estrictamente temporales.

En este caso nos encontramos con la diferencia entre el valor que SPSS muestra en función de un determinado formato elegido, y el valor que realmente almacena para el cálculo.

Dado un valor cualquiera, SPSS lo convierte a segundos y de esta manera lo almacena.

Variables con formato de dólar: En este caso, a un valor numérico dado, se le añade el símbolo dólar en los diferentes



formatos que aparecen en la ventana que se despliega en caso de elegir esta opción. El símbolo dólar y las comas como separadores de los miles son introducidos automáticamente por el sistema en caso de que no sean teclados por el usuario.

Variables con formato de moneda personalizada: Si a través del cuadro de diálogo Opciones del menú Edición se han creado formatos específicos para este tipo de variables, mediante esta opción se puede elegir uno de ellos.

Variables alfanuméricas: Admiten en su codificación cualquier carácter. En su definición debe especificarse únicamente su longitud máxima.

Para definir una variable

Active la ventana Editor de datos.

Pulse dos veces en el nombre de la variable situado en la parte superior de la columna o pulse en cualquier parte de la columna de la variable y elija en los menús:

Datos / Definir variable...

Introduzca un nombre de variable.

Pulse en Tipo para cambiar el formato de los datos.

Pulse en Etiquetas para asignar etiquetas descriptivas de variable y de valor.

Pulse en Valores perdidos para especificar códigos para los valores perdidos.

Pulse en Formato de columna para cambiar la alineación o el ancho de columna.

Para definir el tipo de variable:

Active la ventana Editor de datos.

Pulse dos veces en el nombre de la variable situado en la parte superior de la columna o pulse en cualquier parte de la columna de la variable y elija en los menús:

Datos / Definir variable...

Pulse en Tipo en el cuadro de diálogo Definir variable.

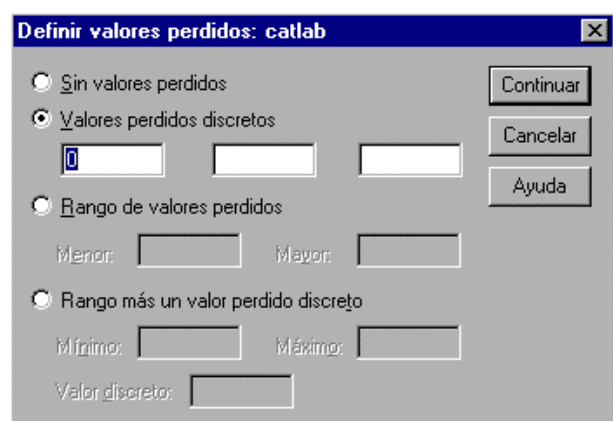
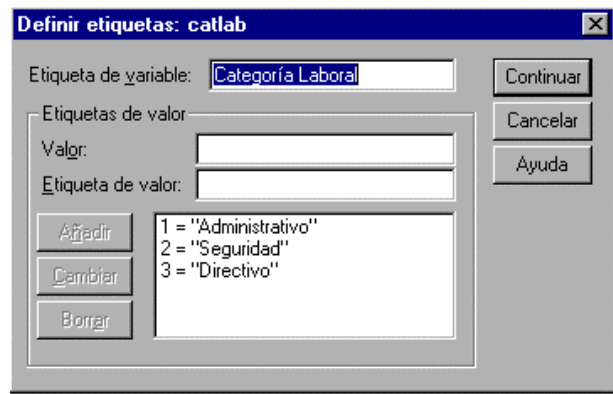
Seleccione el tipo de datos en el cuadro de diálogo Definir tipo de variable.

DEFINIR ETIQUETAS proporciona etiquetas descriptivas de variable y de valor. Aunque los nombres de variable pueden ser sólo de 8 caracteres, las etiquetas de variable pueden tener hasta 256 caracteres, y estas etiquetas descriptivas se muestran en los resultados.

Se pueden asignar etiquetas descriptivas de valor a cada valor de una variable. Esto es particularmente útil si el archivo de datos utiliza códigos numéricos para representar categorías no numéricas (por ejemplo, los códigos 1 y 2 para hombre y mujer). Las etiquetas de valor pueden tener hasta 60 caracteres. Las etiquetas de valor no están disponibles para variables de cadena larga (variables de cadena de más de 8 caracteres).

También se pueden modificar o eliminar etiquetas de valor o de variable.

DEFINIR VALORES PERDIDOS define los valores de datos especificados como perdidos por el usuario. A menudo es útil para saber por qué se pierde información. Por ejemplo, puede desear distinguir entre datos perdidos porque un entrevistado se niega a responder y entre datos perdidos porque la pregunta no afectaba a dicho



entrevistado. Los valores de datos especificados como perdidos por el usuario aparecen marcados para un tratamiento especial y se excluyen de la mayoría de los cálculos.

Se pueden introducir hasta tres valores perdidos (individuales) de tipo discreto, un rango de valores perdidos o un rango más un valor de tipo discreto.

Los rangos sólo se pueden especificar para variables numéricas.

No se pueden definir los valores perdidos para variables de cadena larga (variables de cadena de más de 8 caracteres).

Para definir valores perdidos de una variable

Active la ventana Editor de datos.

Pulse dos veces en el nombre de la variable situado en la parte superior de la columna o pulse en cualquier parte de la columna de la variable y elija en los menús:

Datos / Definir variable...

Pulse en Valores perdidos en el cuadro de diálogo Definir variable.

Introduzca los valores o rango de valores que representan los datos perdidos.

NORMAS DE DENOMINACIÓN DE VARIABLES

A los nombres de variable se les aplican las siguientes reglas:

El nombre debe comenzar por una letra. El resto de los caracteres pueden ser letras, dígitos, puntos o los símbolos @, #, _ o \$.

Los nombres de variable no pueden terminar en punto.

Se deben evitar los nombres de variable que terminan con subrayado (para evitar conflictos con variables creadas automáticamente por varios procedimientos).

La longitud del nombre no debe tener más de ocho caracteres.

Los espacios en blanco y los caracteres especiales (por ejemplo, !, ?, ' y *) no se pueden utilizar.

Cada nombre de variable debe ser exclusivo; no se permiten duplicados. Los nombres de variable no distinguen a las mayúsculas y las minúsculas. Los nombres NEWVAR, NEWVAR y NEWVAR son iguales en SPSS.

Las siguientes palabras claves reservadas no se pueden utilizar como nombres de variable:

ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH

FORMATOS DE ENTRADA FRENTE A FORMATOS DE VISUALIZACIÓN

Dependiendo del formato, la visualización de valores en el Editor de datos puede ser diferente del valor real que se ha introducido y almacenado internamente. Aquí le ofrecemos algunas normas generales:

Para formatos numéricos, de coma y de punto, se pueden introducir valores con cualquier número de dígitos decimales (hasta 16) y el valor completo se almacena internamente. El Editor de datos muestra sólo el número definido de dígitos decimales y redondea los valores con más decimales. Sin embargo, el valor completo se utiliza en cualquier cálculo.

Para variables de cadena, todos los valores se rellenan por la derecha hasta el ancho máximo. Para una variable de cadena con un ancho de 6, un valor de 'No' se almacena internamente como 'No ' y no es equivalente a ' No '.

Para formatos de fecha, se pueden utilizar guiones, barras, espacios, comas o puntos como separadores entre valores de día, mes y año y se pueden introducir números, abreviaciones de tres letras o nombres completos para el valor de mes. Las fechas del formato general dd-mmm-aa se muestran con guiones como separadores y abreviaciones de tres letras para el mes. Las fechas del formato general dd/mm/aa y mm/dd/aa se muestran con barras como separadores y números para el mes. Internamente, las fechas se almacenan como el número de segundos desde el 14 de octubre de 1582.

Para formatos de hora, puede utilizar punto y coma, puntos o espacios como separadores de horas, minutos y segundos. Las horas se muestran con punto y coma como separador. Internamente, las horas se almacenan como el número de segundos.

Definir formato de columna controla el ancho de columnas en el Editor de datos y de la alineación de los valores de datos. Los anchos de columna también se pueden cambiar en el Editor de datos pulsando y arrastrando los bordes de la columna.

Los formatos de columna afectan sólo a la visualización de valores en el Editor de datos. Al cambiar el ancho de columna no se cambia el ancho definido de una variable. Si el ancho real y definido de un valor es más ancho que la columna, aparecerán asteriscos (*) en la ventana Editor de datos.

Plantilla proporciona un método para crear y aplicar plantillas de definición de variables. Se puede asignar la misma información de definición de variables a múltiples variables con plantillas de variables. Por ejemplo, si tiene un grupo de variables en el que todas utilizan los códigos numéricos 1 y 2 para representar las respuestas "sí" y "no" y el 9 para representar las respuestas perdidas, se puede crear una plantilla que contenga esas etiquetas de valor y las especificaciones de valores perdidos y aplicarla al grupo entero de variables.



Para crear una plantilla de variable

Active la ventana Editor de datos.

Elija en los menús: Datos / Plantilla...

Pulse en Definir en el cuadro de diálogo Plantilla.

Seleccione los atributos que desea definir.

Introduzca un nombre de plantilla.

Y pulse en Añadir.

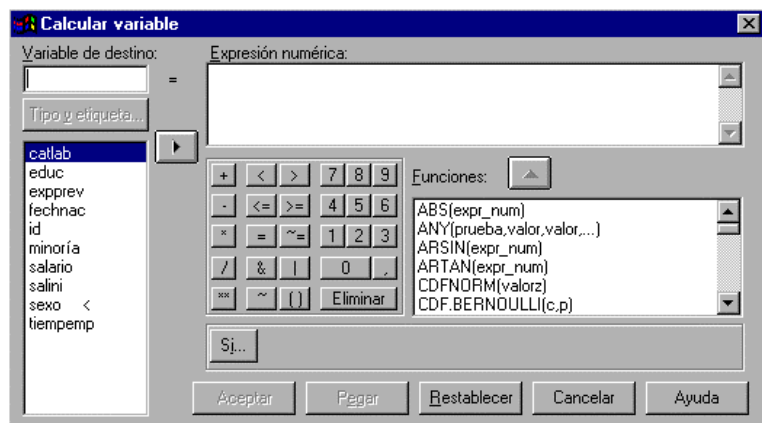
4.- MODIFICACIÓN DE VARIABLES.

GENERACIÓN DE NUEVAS VARIABLES

Se pueden generar nuevas variables mediante transformaciones numéricas efectuadas sobre los valores de las pre-existentes.

Para ello, del menú Transformar, debe escogerse la opción Calcular..., con lo que se abre el cuadro de diálogo de la derecha.

Por defecto, las nuevas variables calculadas son numéricas. Si se quiere especificar otro tipo y añadir



etiquetas, debe emplearse el botón correspondiente. Para calcular una nueva variable de cadena, deberá especificar el tipo de datos y el ancho.

Etiqueta. Etiqueta de variable descriptiva y opcional de hasta 120 caracteres de longitud. Se puede introducir una etiqueta o bien utilizar como la etiqueta los primeros 110 caracteres de la expresión Calcular.

Tipo. Las variables calculadas pueden ser numéricas o de cadena (alfanuméricas). Las variables de cadena no se pueden utilizar en los cálculos.

Una vez que se ha asignado un nombre a esta variable, el siguiente paso es definir la expresión numérica que va a permitir calcular los valores de la misma.

Tal expresión puede constar de los siguientes elementos: nombres de variables del fichero original, constantes, operadores y funciones. Y se escribe en el cuadro de texto titulado Expresión numérica.

Modificación condicional de variables: El cuadro de diálogo *Si los casos* le permite aplicar transformaciones de los datos en subconjuntos de casos seleccionados utilizando expresiones condicionales. Una expresión condicional devuelve un valor verdadero (true), falso (false) o perdido (missing) para cada caso.

Si el resultado de una expresión condicional es true, la transformación se aplicará al caso.

Si el resultado de una expresión condicional es false o missing, la transformación no se aplicará al caso.

La mayoría de las expresiones condicionales utilizan uno o más de los seis operadores relacionales (<, >, <=, >=, =, ~=) en el teclado de calculadora.

Las expresiones condicionales pueden incluir nombres de variables, constantes, operadores aritméticos, funciones numéricas y otras funciones, variables lógicas y operadores relacionales.

RECODIFICACIÓN DE VALORES.

a.- Recodificar en las mismas variables

Recodificar en las mismas variables reasigna los valores de variables existentes o agrupa rangos de valores existentes en nuevos valores. Por ejemplo, podría agrupar los salarios en categorías de rango de salarios.

Se pueden recodificar las variables numéricas y de cadena. Si se seleccionan múltiples variables, todas deben ser del mismo tipo. No se pueden recodificar juntas las variables numéricas y de cadena.



Para recodificar los valores de una variable

Elija en los menús: Transformar / Recodificar / En las mismas variables...

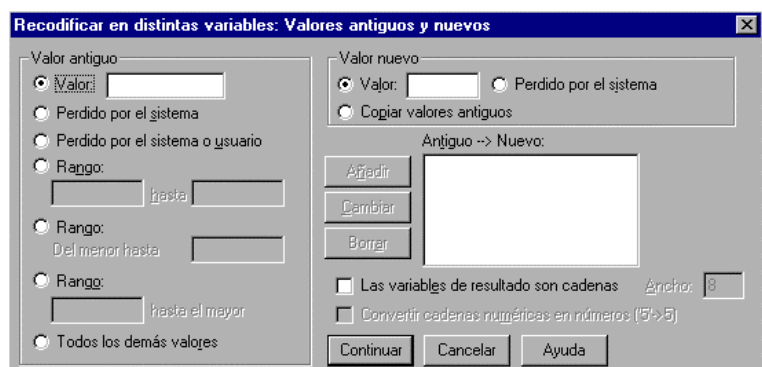
Seleccione las variables que desea recodificar. Si se seleccionan múltiples variables, deberán ser del mismo tipo (numéricas o de cadena).

Pulse en Valores antiguos y nuevos y especifique cómo recodificar los valores.

Si se desea, puede definir un subconjunto de casos que se van a recodificar.

b.- Recodificar en distintas variables

Recodificar en distintas variables reasigna los valores de variables existentes o agrupa rangos de



valores existentes en nuevos valores para una nueva variable. Por ejemplo, podría agrupar los salarios en una nueva variable que contenga categorías de rango de salarios.

Se pueden recodificar las variables numéricas y de cadena.

Las variables numéricas se pueden recodificar en variables de cadena y viceversa.

Si se seleccionan múltiples variables, todas deben ser del mismo tipo. No se pueden recodificar juntas las variables numéricas y de cadena.

Para recodificar los valores de una variable en una nueva variable

Elija en los menús: Transformar / Recodificar / En distintas variables...

Seleccione las variables que desea recodificar. Si se seleccionan múltiples variables, deberán ser del mismo tipo (numéricas o de cadena).

Introduzca el nombre de una (nueva) variable de los resultados para cada nueva variable y pulse en Cambiar.

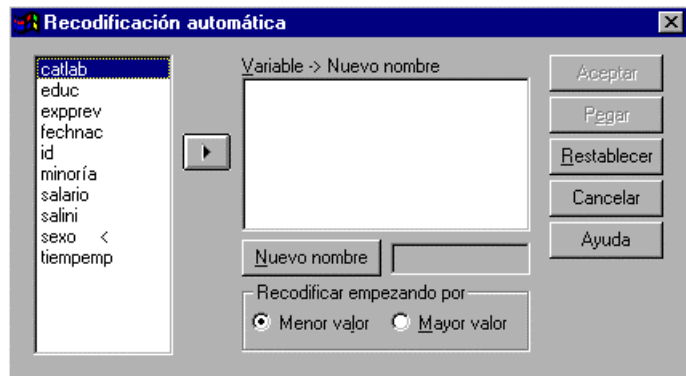
Pulse en Valores antiguos y nuevos y especifique cómo recodificar los valores.

Si se desea, puede definir un subconjunto de casos que se van a recodificar.

c.- Recodificación automática.

Recodificación automática convierte los valores numéricos y de cadena en valores enteros consecutivos. Cuando los códigos de categoría no son secuenciales, las casillas vacías resultantes reducen el rendimiento y aumentan los requisitos de memoria para muchos procedimientos de SPSS. Además, algunos procedimientos no pueden utilizar variables de cadena y otros requieren valores enteros consecutivos para los niveles de los factores.

La nueva variable o variables creadas por Recodificación automática conservan cualquier variable definida y las etiquetas de valor de la variable antigua. Para aquellos valores sin una etiqueta de valor definida, el valor original se utilizará como etiqueta para el valor recodificado. Una tabla muestra los valores antiguos y nuevos y las etiquetas de valor.



Los valores de cadena se recodifican por orden alfabético, con las mayúsculas antes que las minúsculas. Los valores perdidos se recodifican en valores perdidos mayores que cualquier valor no perdido, conservando el orden. Por ejemplo, si la variable original posee 10 valores no perdidos, el valor perdido mínimo se recodificaría en 11 y el valor 11 sería un valor perdido para la nueva variable.

Los valores de cadena se recodifican por orden alfabético, con las mayúsculas antes que las minúsculas. Los valores perdidos se recodifican en valores perdidos mayores que cualquier valor no perdido, conservando el orden. Por ejemplo, si la variable original posee 10 valores no perdidos, el valor perdido mínimo se recodificaría en 11 y el valor 11 sería un valor perdido para la nueva variable.

Para recodificar valores numéricos o de cadena en valores enteros consecutivos

Elija en los menús: Transformar / Recodificación automática...

Seleccione una o más variables que se vayan a recodificar.

Para cada variable seleccionada, introduzca un nombre para la nueva variable y pulse en Nuevo nombre.

5.- DEFINICIÓN Y USO DE CONJUNTOS DE VARIABLES.

El cuadro de diálogo Definir conjuntos de variables crea subconjuntos de variables que se muestran en las listas de origen de los cuadros de diálogo.

Los conjuntos de variables de tamaño reducido hacen que sea más fácil encontrar y seleccionar las variables para su análisis, así como mejorar el rendimiento de SPSS. Si el archivo de datos contiene un elevado número de variables y cuadros de diálogo que se abren con lentitud, restringir las listas de origen de un cuadro de diálogo a subconjuntos de variables más pequeños reduciría la cantidad de tiempo empleado en abrirlos.

Nombre del conjunto. Los nombres de los conjuntos tienen una longitud máxima de 12 caracteres. Puede emplearse todo tipo de caracteres, incluidos los espacios en blanco. Los nombres de los conjuntos no distinguen mayúsculas de minúsculas.

Variables del conjunto. El conjunto puede estar compuesto de cualquier combinación de variables numéricas, de cadena corta y cadena larga. El orden de las variables del conjunto no tiene ningún efecto en el orden de visualización de las variables en las listas de origen del cuadro de diálogo. Una variable puede pertenecer a múltiples conjuntos.

Para definir conjuntos de variables

Elija en los menús: Utilidades / Definir conjuntos...

Seleccione las variables que desee incluir en el conjunto.

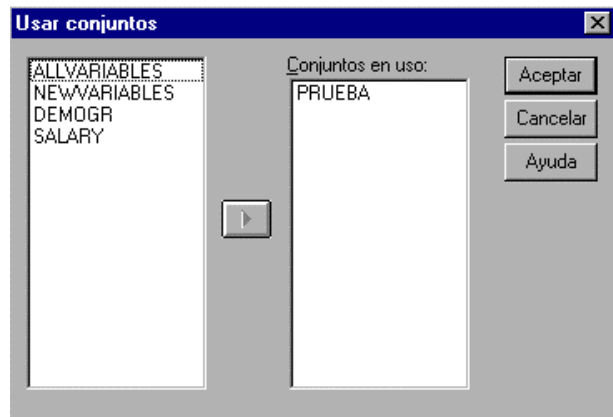
Introduzca un nombre para el conjunto (hasta 12 caracteres).

Pulse en Añadir.



El cuadro de diálogo Usar conjuntos restringe las variables mostradas en las listas de origen del cuadro de diálogo a los conjuntos seleccionados que haya definido.

Los conjuntos de variables de tamaño reducido hacen que sea más fácil encontrar y seleccionar las variables para su análisis, así como mejorar el rendimiento de SPSS. Si el archivo de datos contiene un número elevado de variables y los cuadros de diálogo se abren con lentitud, restringir las listas de origen de un cuadro de diálogo a subconjuntos más pequeños reduciría la cantidad de tiempo empleado en abrirlos.



Conjuntos en uso. Muestra los conjuntos empleados en la creación de listas de variables de origen en los cuadros de diálogo. Las variables aparecen ordenadas en las listas de origen por nombre o por archivo. El orden de conjuntos y de variables en un conjunto no tiene ningún efecto en el orden de las variables de la lista de origen. Por defecto, se utilizan dos conjuntos definidos por el sistema:

ALLVARIABLES. Este conjunto contiene todas las variables del archivo de datos, incluidas las nuevas variables creadas durante una sesión.

NEWVARIABLES. Este conjunto contiene sólo las nuevas variables creadas durante la sesión.

Puede eliminar estos conjuntos de la lista y seleccionar otros, si bien debe haber al menos un conjunto en la lista. Si no elimina el conjunto ALLVARIABLES de la lista Conjuntos en uso, cualquier otro conjunto que incluya no tendrá efecto alguno.

Como mínimo debe haber un conjunto en uso, pero, puede haber más de uno.

6.- PONDERACIÓN DE CASOS.

Cuando cada registro de los que componen un fichero de trabajo representa más de un caso, SPSS permite especificar el coeficiente de ponderación correspondiente.

Es requisito que en el fichero exista una variable cuyos valores van a ser empleados como coeficientes de ponderación.

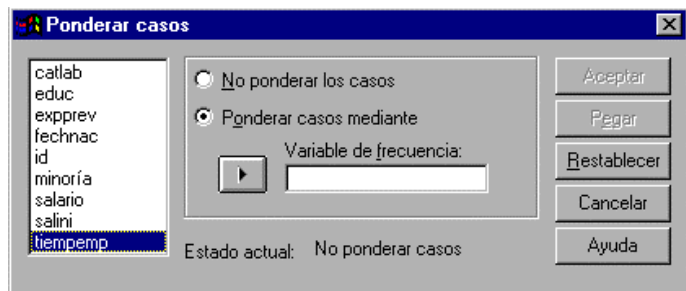
Ponderar casos proporciona a los casos diferentes ponderaciones (mediante una réplica simulada) para el análisis estadístico.

Los valores de la variable de ponderación deberían indicar el número de observaciones representado por los casos únicos del archivo de datos.

Los casos con valores perdidos, negativos o cero para la variable de ponderación se excluyen del análisis.

Los valores fraccionarios son válidos; se usan exactamente cuando tienen sentido y, con mayor probabilidad, cuando se tabulan los casos (como en las tablas de contingencia).

Una vez aplicada una variable de ponderación, permanece activada hasta que se seleccione otra variable de ponderación o se desactive la ponderación. Si guarda un archivo de datos ponderado, la información de ponderación se guarda con el archivo de datos. Puede desactivar la ponderación en cualquier momento, incluso después de haber guardado el archivo de forma ponderada.



Ponderaciones en diagramas de dispersión e histogramas. Los diagramas de dispersión y los histogramas tienen una opción para activar y desactivar las ponderaciones de los casos, pero esto no afecta a los casos con valor negativo, 0, o un valor perdido para la variable de ponderación. Estos casos permanecen excluidos del gráfico incluso si desactiva la ponderación desde dentro del gráfico.

Para ponderar casos

Elija en los menús: Datos / Ponderar casos...

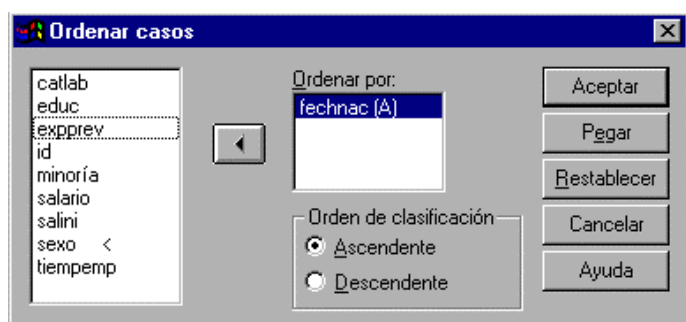
Seleccione Ponderar casos por.

Seleccione una variable de frecuencia.

Los valores de la variable de frecuencia se utilizan como ponderaciones de los casos. Por ejemplo, un caso con un valor de 3 para la variable de frecuencia representará tres casos en el archivo de datos ponderado.

7.- ORDENACIÓN DE CASOS.

Este cuadro de diálogo ordena casos (filas) del archivo de datos basándose en los valores de uno o más tipos de variables. Se



pueden ordenar los casos en orden ascendente o descendente.

Si se selecciona más de una variable de ordenación, los casos serán ordenados por cada variable dentro de las categorías de la variable anterior en la lista Ordenar. Por ejemplo, si selecciona GÉNERO como la primera variable de ordenación y MINORÍA como la segunda variable de ordenación, la clasificación minoría ordenará los casos dentro de cada categoría de género.

Para variables de cadena, las letras mayúsculas preceden a las minúsculas correspondientes en orden. Por ejemplo, el valor de cadena "Sí" precede a "sí" en orden.

Para ordenar casos

Elija en los menús: Datos / Ordenar casos...

Seleccione una o más variables de ordenación.

El archivo de datos se ordena basándose en los valores de las variables de ordenación. Si selecciona múltiples variables de ordenación, los casos serán ordenados por los valores de cada variable dentro de las categorías de la variable anterior en la lista Ordenar.

8.- SELECCIÓN DE CASOS.

Seleccionar casos proporciona varios métodos para seleccionar un subgrupo de casos basados en los criterios que incluyen variables y expresiones complejas. También se puede seleccionar una muestra aleatoria de casos. Los criterios usados para definir un subgrupo pueden incluir:

- Valores y rangos de las variables
- Rangos de fechas y de horas
- Números de caso (filas)
- Expresiones aritméticas
- Expresiones lógicas
- Funciones

Casos no seleccionados. Puede filtrar o eliminar casos que no reúnen los criterios de selección.

Los casos filtrados permanecen en el archivo de datos pero se excluyen del análisis. SPSS crea una variable de filtro, FILTER_\$, para indicar el estado del filtro. Los casos seleccionados tienen un valor de 1; los casos filtrados tienen un valor de 0. Los casos filtrados también están indicados con una barra transversal sobre el número de fila en el Editor de datos. Para desactivar el filtrado e incluir todos los casos en el análisis, seleccione Todos los casos.

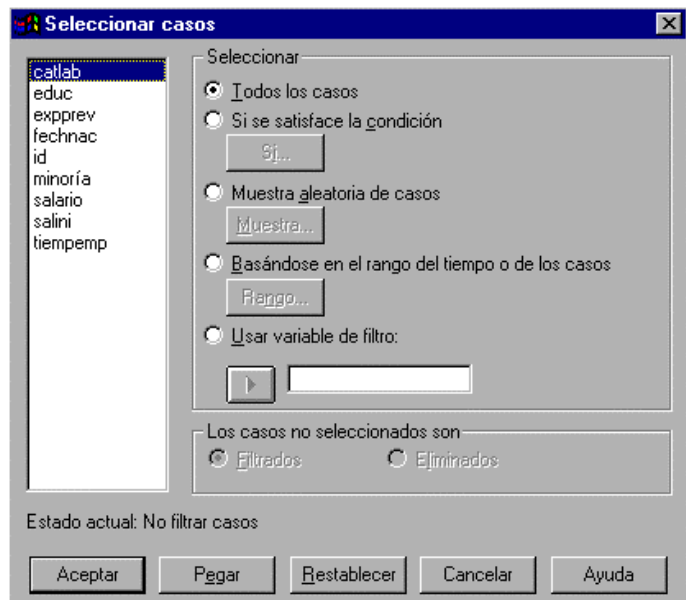
Si después de haber eliminado los casos guarda los cambios del archivo de datos de trabajo (con el mismo nombre), entonces no podrá recuperar los casos eliminados.

Para seleccionar subconjuntos de casos

Elija en los menús: Datos / Seleccionar casos...

Seleccione uno de los métodos de selección de casos.

Especifique los criterios para la selección de casos.

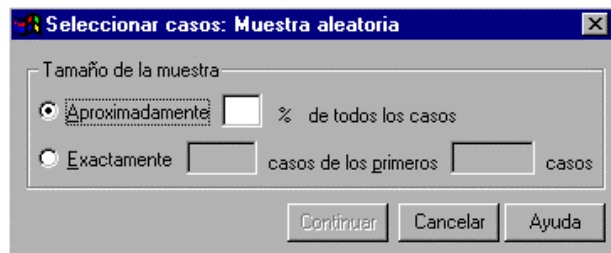


SELECCIONAR CASOS: MUESTRA ALEATORIA

Este cuadro de diálogo le permite seleccionar una muestra aleatoria basándose en un porcentaje aproximado o en un número exacto de casos.

Aproximadamente. SPSS genera una muestra aleatoria con el porcentaje aproximado de casos especificado. Dado que SPSS toma una decisión pseudo-aleatoria para cada caso, el porcentaje de casos seleccionados únicamente se puede aproximar al porcentaje especificado. Cuantos más casos haya en el archivo de datos, más se acercará el porcentaje de casos seleccionados al porcentaje especificado.

Exactamente. Un número de casos especificado por el usuario. También debe especificar el número de casos a partir de los que hay que generar la muestra. Este segundo número debe ser menor o igual que el número total de casos en el archivo de datos. Si este número excede el número total de casos del archivo de datos, la muestra contendrá proporcionalmente un número de casos menor que el número solicitado.



Para seleccionar una muestra aleatoria de casos

Elija en los menús: Datos / Seleccionar casos...

Seleccione Muestra aleatoria de casos.

Pulse en Muestra.

Seleccione el método de muestreo e introduzca el porcentaje o el número de casos.

SELECCIONAR CASOS: RANGO

Este cuadro de diálogo selecciona los casos basándose en un rango de números de caso o en un rango de fechas u horas.

Los rangos de casos se basan en el número de fila tal y como se visualiza en el Editor de datos.

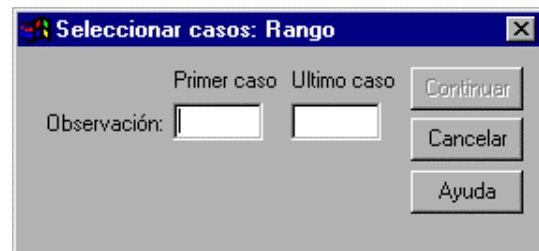
Los rangos de fechas y horas sólo están disponibles para los datos de la serie temporal con variables de fecha definida (menú Datos, Definir fechas).

Elija en los menús: Datos / Seleccionar casos...

Seleccione Basándose en el rango del tiempo o de los casos.

Pulse en Rango.

Introduzca los números de caso, hora o fecha iniciales y finales.

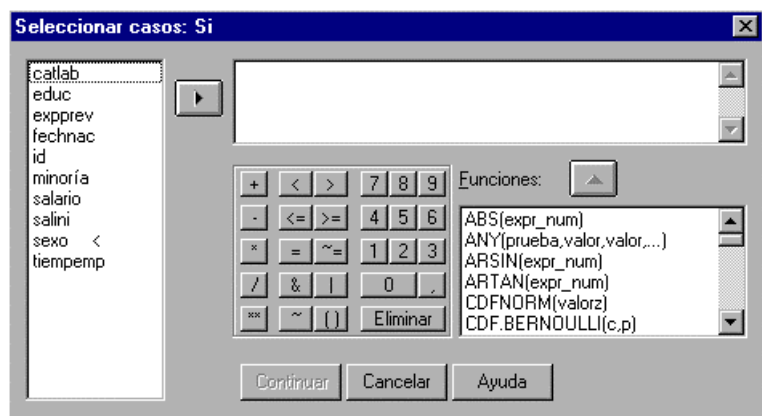


SELECCIONAR CASOS: SI

Este cuadro de diálogo le permite seleccionar subconjuntos de casos utilizando expresiones condicionales. Una expresión condicional devuelve un valor verdadero, falso o perdido para cada caso.

Si el resultado de una expresión condicional es verdadero, se selecciona el caso.

Si el resultado de una expresión condicional es falso o perdido, entonces no se selecciona el caso.



La mayoría de las expresiones condicionales utilizan uno o más de los seis operadores relacionales (<, >, <=, >=, =, ~=) del teclado de la calculadora.

Las expresiones condicionales pueden incluir nombres de variables, constantes, operadores aritméticos, funciones numéricas y otras, variables lógicas y operadores relacionales.

Para seleccionar subconjuntos de casos basados en una expresión condicional

Elija en los menús: Datos / Seleccionar casos...

Seleccione Si se satisface la condición.

Pulse en Si.

Introduzca la expresión condicional.

Para construir una expresión, pegue los componentes en el campo Expresión o teclee directamente en el campo Expresión.

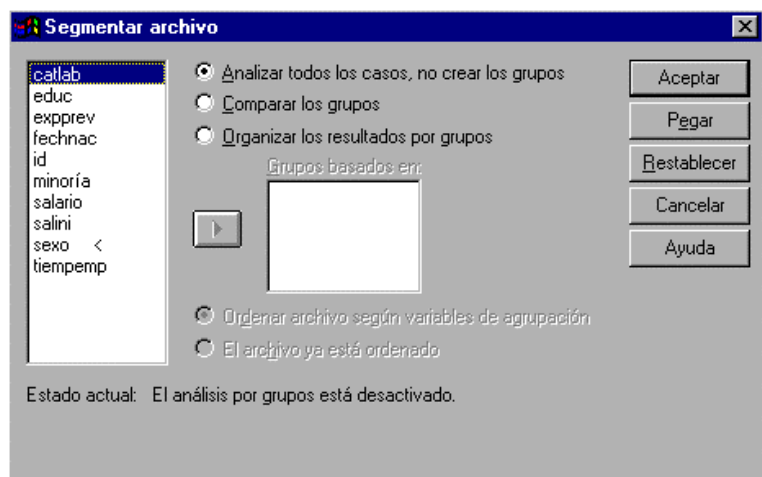
Pegue las funciones de la lista de funciones y rellene los parámetros indicados por signos de interrogación.

Las constantes de cadena deben ir entre comillas o apóstrofes.

Las constantes numéricas deben teclearse en formato americano, con el punto (.) como separador decimal.

9.- PARTICIÓN O SEGMENTACIÓN DEL ARCHIVO DE TRABAJO

Segmentar archivo divide el archivo de datos en grupos diferentes para el análisis, basándose en los valores de una o más variables de agrupación. Si se seleccionan varias variables de agrupación, los casos serán agrupados por cada variable dentro de las categorías de la variable anterior en la lista Grupos basados en. Por ejemplo, si selecciona GÉNERO como la primera variable de agrupación y MINORÍA como la segunda variable de agrupación, los casos serán agrupados por la clasificación minoría dentro de cada categoría de género.



Se pueden especificar hasta ocho variables de agrupación.

Cada ocho caracteres de una variable de cadena larga (las variables de cadena cuya longitud es superior a 8 caracteres) cuenta como una variable en el límite de ocho variables de agrupación.

Los casos deberían ordenarse según los valores de las variables de agrupación, en el mismo orden en el que aparecen las variables en la lista Grupos basados en. Si el archivo de datos todavía no está ordenado, seleccione Ordenar archivo según las variables de segmentación.

Comparar los grupos. Los grupos de segmentación del archivo se presentan juntos para poder compararlos. Para las tablas pivote, se crea una única tabla pivote y cada variable de segmentación del archivo puede desplazarse entre las dimensiones de la tabla. Para los gráficos, se crea un gráfico diferente para cada grupo de segmentación del archivo y todos los gráficos se muestran juntos en el Navegador de resultados.

Organizar los resultados por grupos. Los resultados de cada procedimiento se muestran por separado para cada grupo de segmentación del archivo.

Segmentar un archivo de datos para el análisis

Elija en los menús: Datos / Segmentar archivo...

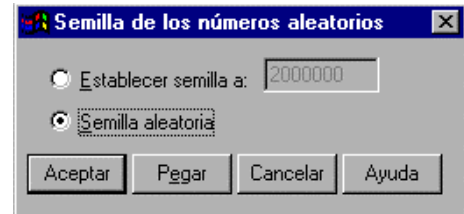
Seleccione Comparar los grupos u Organizar los resultados por grupos.
Seleccione una o más variables de agrupación.

Si el archivo de datos todavía no está ordenado según los valores de las variables de agrupación, seleccione Ordenar archivo según las variables de segmentación.

10.- SEMILLA DE ALEATORIZACIÓN.

Semilla de aleatorización establece la semilla utilizada por el generador del número pseudo-aleatorio en un valor específico de modo que se pueda reproducir una secuencia de números pseudo-aleatorios.

La semilla de aleatorización cambia cada vez que SPSS genera un número aleatorio para su uso en transformaciones (como por ejemplo las funciones UNIFORM y NORMAL), muestreo aleatorio o ponderación de casos. Para replicar una secuencia de números aleatorios, utilice este cuadro de diálogo para restablecer la semilla en un valor específico anterior a cada análisis que utilice los números aleatorios.



La semilla de aleatorización se restablece automáticamente en 2.000.000 cada vez que se inicia una nueva sesión de SPSS.

Para establecer la semilla de aleatorización

Elija en los menús: Transformar / Número de aleatorización...

Seleccione Establecer semilla a.

Introduzca un entero positivo entre 1 y 2.000.000.000

11.- APLICAR EL DICCIONARIO DE DATOS.

Este cuadro de diálogo aplica la información del diccionario SPSS (etiquetas, valores perdidos, formatos) de un archivo de datos externo de SPSS al archivo de datos de trabajo. La información del diccionario se aplica basándose en los nombres de las variables de concordancia. Las variables no tienen que estar en el mismo orden en los dos archivos y las variables que no estén presentes en ambos archivos no se verán afectadas. Se aplicarán las siguientes reglas:

Si el tipo de variable (numérica o de cadena) es el mismo en ambos archivos, se aplica toda la información del diccionario.

Si el tipo de variable no es el mismo en ambos archivos, o si se trata de una cadena larga (de más de ocho caracteres), sólo se aplicará la etiqueta de variable.

Los formatos numéricos, de dólar, punto, coma, fecha y hora se consideran todos como numéricos y se aplicará toda la información del diccionario.

Los anchos de variable no se verán afectados por el diccionario aplicado.

Para las variables de cadena corta (de ocho caracteres o menos), los valores perdidos y los valores especificados para las etiquetas de valor se truncan si exceden el ancho definido de la variable en el archivo de datos de trabajo.

Cualquier información del diccionario aplicada sobrescribe la información del diccionario existente.

12.- SINTAXIS.

El método más sencillo para generar un archivo de sintaxis de comandos consiste en hacer selecciones en cuadros de diálogo de SPSS y pegar la sintaxis de las selecciones en una ventana de sintaxis. Si pega la sintaxis en cada paso de un análisis largo, podrá generar un archivo de trabajo

que le permitirá repetir el análisis con posterioridad o ejecutar un trabajo automatizado con la Unidad de producción de SPSS.

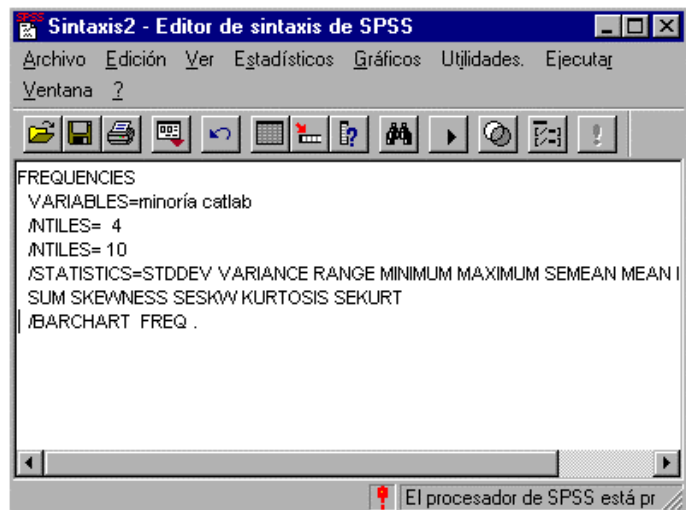
En la ventana de sintaxis, puede ejecutar la sintaxis pegada, editarla y guardarla en un archivo de sintaxis.

Para pegar la sintaxis de comandos desde un cuadro de diálogo

Abra el cuadro de diálogo y realice las selecciones que desee.

Pulse en Pegar.

La sintaxis de comandos se pega en la ventana de sintaxis designada. Si no tiene abierta una ventana de sintaxis, SPSS abrirá una ventana nueva y pegará la sintaxis en ella.



Nota: Si abre un cuadro de diálogo desde los menús de la ventana de proceso, el código necesario para ejecutar la sintaxis de un proceso se pegará en la ventana de proceso.

USO DE LA SINTAXIS DESDE EL REGISTRO DE RESULTADOS

Puede construir un archivo de sintaxis copiando la sintaxis de comandos desde el registro de SPSS que aparece en el Navegador de resultados. Para usar este método debe seleccionar Mostrar los comandos en el registro en la pestaña Navegador del cuadro de diálogo Opciones de SPSS antes de ejecutar el análisis. Cada comando aparecerá en el Navegador de resultados junto con los resultados del análisis.

En la ventana de sintaxis, puede ejecutar la sintaxis pegada, editarla y guardarla en un archivo de sintaxis.

USO DE LA SINTAXIS DESDE EL ARCHIVO DIARIO

Por defecto, SPSS registra todos los comandos que se han ejecutado durante una sesión en un archivo diario denominado SPSS.JNL (definido con las Opciones del menú Edición). Puede editar el archivo diario y guardarlo como archivo de sintaxis que podrá utilizar para repetir la ejecución de un análisis anterior, o bien puede ejecutarlo en un trabajo automatizado con una Unidad de producción de SPSS.

El archivo diario es un archivo de texto que puede editarse como cualquier otro archivo de texto. Dado que los mensajes de error y las advertencias también se registran en el archivo diario junto con la sintaxis de comandos, deberá editar los mensajes de error y las advertencias que aparezcan antes de guardar el archivo de sintaxis. No obstante, tenga en cuenta que los errores deben resolverse o el trabajo no podrá ejecutarse de forma satisfactoria.

Guarde el archivo diario editado con un nombre diferente. Puesto que SPSS añade o sobrescribe automáticamente el archivo diario en cada sesión, todo intento de utilizar el mismo nombre para un archivo de sintaxis y un archivo diario puede producir resultados inesperados.

Para abrir el archivo diario, elija en los menús:

Archivo / Abrir...

Localice y abra el archivo diario de SPSS (por defecto es SPSS.JNL, situado en el directorio TEMP).

Seleccione Todos los archivos (*.*) en Archivos de tipo o escriba *.jnl en el cuadro Nombre de archivo para mostrar los archivos diarios de la lista de archivos. Si tiene

alguna dificultad para localizar el archivo, use Opciones del menú Edición para ver dónde se guarda el diario en su sistema.

Edite el archivo para eliminar los mensajes de error o advertencias que contenga, indicados por el signo >.

Guarde el archivo diario editado con un nombre diferente. (Se recomienda usar un nombre de archivo con la extensión .SPS, la extensión por defecto de los archivos de sintaxis).

Para ejecutar la sintaxis de comandos

Resalte los comandos que desee ejecutar en la ventana de sintaxis.

Pulse en el botón Ejecutar (el triángulo que apunta hacia la derecha) en la barra de herramientas de la ventana de sintaxis.

o bien

Seleccione una de las opciones del menú Ejecutar.

Todo. Ejecuta todos los comandos de la ventana de sintaxis.

Selección. Ejecuta los comandos seleccionados. Se incluyen los comandos resaltados parcialmente.

Actual. Ejecuta el comando donde se encuentra el cursor.

Hasta el final. Ejecuta todos los comandos incluidos desde la posición actual del cursor hasta el final del archivo de sintaxis de comandos.

El botón Ejecutar de la barra de herramientas del Editor de sintaxis ejecuta los comandos seleccionados o el comando donde se encuentra el cursor si no hay nada seleccionado.

Puede acceder a la mayoría de comandos de SPSS desde los menús y cuadros de diálogo de SPSS. No obstante, algunos comandos y opciones sólo están disponibles mediante el uso del lenguaje de comandos de SPSS. El lenguaje de comandos también permite guardar los trabajos en un archivo de sintaxis, lo que permite repetir los análisis posteriormente o ejecutarlos en un trabajo automatizado con la Unidad de producción de SPSS.

Un archivo de sintaxis es sencillamente un archivo de texto que contiene comandos de SPSS. Aunque es posible abrir una ventana de sintaxis y escribir comandos, es más sencillo permitir que SPSS le ayude a construir un archivo de sintaxis mediante uno de los siguientes métodos:

Pegando la sintaxis de comandos desde cuadros de diálogo

Copiando la sintaxis desde el registro de resultados

Copiando la sintaxis desde el archivo diario

En la Ayuda en pantalla de un procedimiento determinado de SPSS, pulse en el botón Sintaxis para saber qué opciones del lenguaje de comandos están disponibles (si hay alguna) para ese procedimiento y acceder al diagrama de sintaxis del comando de SPSS pertinente.

13.- DISTRIBUCIONES DE FRECUENCIAS.

Con el fin de obtener las tablas de frecuencias, las representaciones gráficas correspondientes y los estadísticos asociados, se entra en el cuadro de diálogo:

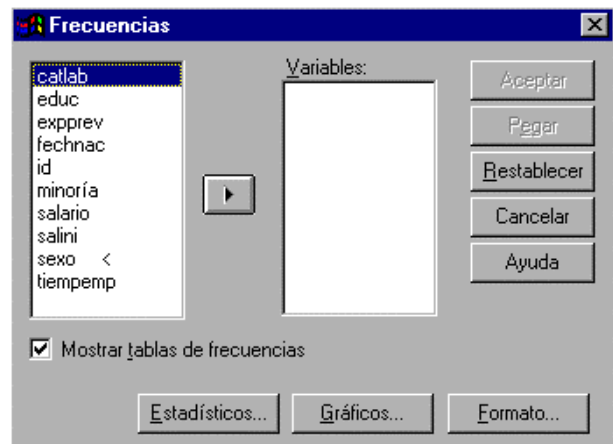
Estadísticos / Resumir / Frecuencias.

CONSIDERACIONES SOBRE LOS DATOS DE FRECUENCIAS

Datos. Utiliza códigos numéricos o cadenas cortas para codificar variables categóricas (medidas nominales u ordinales).

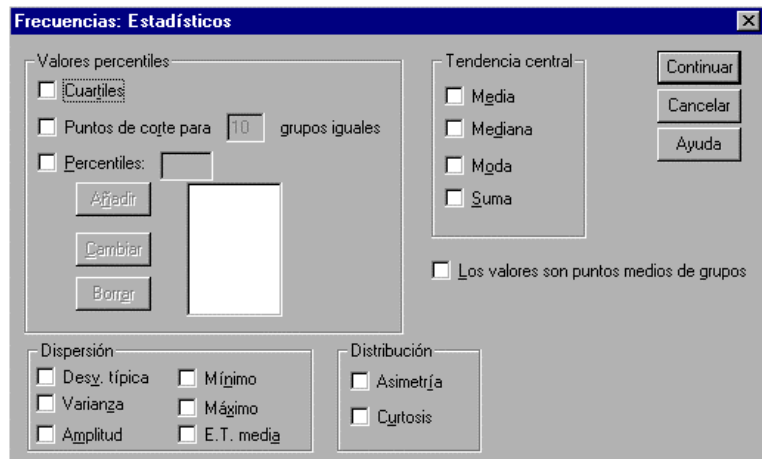
Supuestos. Las tabulaciones y los porcentajes proporcionan una descripción útil para los datos de cualquier distribución, especialmente para las variables con categorías ordenadas o desordenadas. Muchas de los estadísticos de resumen opcionales, tales como la media y la desviación típica, se basan en la teoría normal y son apropiados para variables cuantitativas con distribuciones simétricas. Los estadísticos robustos, tales como la mediana, los cuartiles y los percentiles son apropiados para variables cuantitativas que pueden o no cumplir el supuesto de normalidad.

Procedimientos relacionados. Si desea calcular los estadísticos de resumen para cada uno de los diversos grupos de casos (por ejemplo, desea estadísticos diferentes para mujeres y hombres o para gente que vive en cuatro regiones de un país), utilice Segmentar archivo en el menú Datos, o utilice los procedimientos Explorar, Medias o Resumir. Utilice el procedimiento Explorar para un grupo de histogramas que utilicen todos la misma escala.



FRECUENCIAS: ESTADISTICOS

Valores percentiles. Valores de una variable cuantitativa que dividen los datos ordenados en grupos de forma que un porcentaje se encuentre por encima y otro porcentaje por debajo. Los cuartiles (los percentiles 25, 50 y 75) dividen las observaciones en cuatro grupos de igual tamaño. Si desea un número igual de grupos que no sea cuatro, seleccione Puntos de corte para n grupos iguales. También se pueden especificar percentiles individuales (por ejemplo, el percentil 95, el valor por debajo del cual se encuentran el 95% de las observaciones).



Tendencia central. Los estadísticos que describen la localización de la distribución incluyen la media (aritmética), la mediana, la moda y la suma de todos los valores.

Dispersión. Los estadísticos que miden la cantidad de variación o de dispersión en los datos incluyen la desviación típica (estimación insesgada de la desviación típica: Medida de dispersión en torno a la media. Viene expresada en las mismas unidades de medida que las observaciones, e igual a la raíz cuadrada de la varianza. En una distribución normal, el 68% de los casos se encuentra dentro de una DT respecto a la media y el 95% de los casos se encuentra dentro de 2DT respecto a la media.), la varianza (estimación insesgada de la varianza o cuasivarianza: Medida de dispersión en torno a la media, igual a la suma de los cuadrados de las desviaciones respecto a la media dividida por el número de casos menos 1. La varianza se mide en unas unidades que son el cuadrado de las de la propia variable.), el rango (amplitud total), el mínimo, el máximo y el error típico de la media (Medida de cuánto puede variar el valor de la media de una muestra a otra, extraídas éstas de la misma distribución. Puede utilizarse para comparar de forma aproximada la media observada con un valor hipotetizado (es decir, podremos concluir que dos valores son distintos si la razón de la diferencia respecto al error típico es menor que -2 o mayor que +2)).

Distribución. Asimetría y curtosis son estadísticos que describen la forma y la simetría de la distribución. Estos estadísticos se muestran con sus errores típicos.

Los valores son puntos medios de grupo. Si los valores en sus datos son puntos medios de grupo (por ejemplo, edades de todas las personas entre treinta y cuarenta años se codifican como 35), se debe seleccionar esta opción para estimar la media y los percentiles para los datos originales, no agrupados.

Para obtener estadísticos para Frecuencias

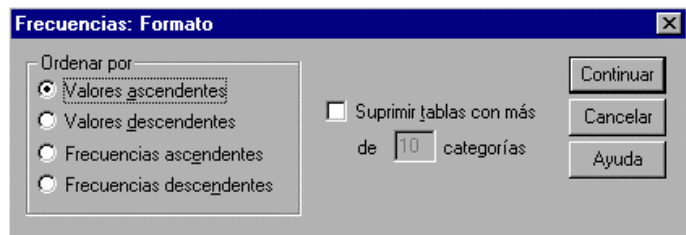
Elija en los menús: Estadísticos / Resumir / Frecuencias...

En el cuadro de diálogo Frecuencias, pulse en Estadísticos.

Seleccione los estadísticos que desee.

FRECUENCIAS: FORMATO

Ordenar por. La tabla de frecuencias se puede organizar de acuerdo con los valores actuales en los datos o de acuerdo con el recuento (frecuencia de aparición) de esos valores, y tanto en orden ascendente como descendente. Sin embargo, si se solicita un histograma o percentiles, SPSS asume que la variable es cuantitativa y mostrará sus valores en orden ascendente.



Suprimir las tablas con más de n categorías. Esta opción impide que se muestren tablas con más valores que el número especificado.

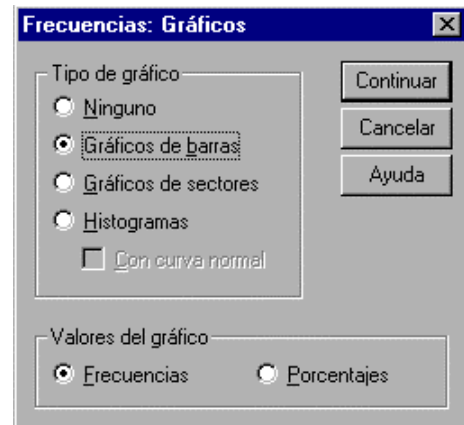
Para especificar formatos para Frecuencias

Elija en los menús: Estadísticos / Resumir / Frecuencias...

En el cuadro de diálogo Frecuencias, pulse en Formato.

FRECUENCIAS: GRÁFICOS

Tipo de gráfico. Un gráfico de sectores muestra la contribución de las partes a un todo. Cada sector de un gráfico de sectores corresponde a un grupo definido por una variable única de agrupación. Un gráfico de barras muestra la frecuencia de cada valor o categoría distinta como una barra diferente, permitiendo comparar las categorías de forma visual. Un histograma también tiene barras, pero se representan a lo largo de una escala de intervalos iguales. La altura de cada barra es el recuento de valores de una variable cuantitativa que están dentro del intervalo. Un histograma muestra la forma, el centro y la dispersión de la distribución. Una curva normal superpuesta en un histograma ayuda a juzgar si los datos están normalmente distribuidos.



Valores del gráfico. Para gráficos de barras, se puede etiquetar el eje de escala con frecuencias o porcentajes.

Para obtener gráficos para Frecuencias

Elija en los menús: Estadísticos / Resumir / Frecuencias...

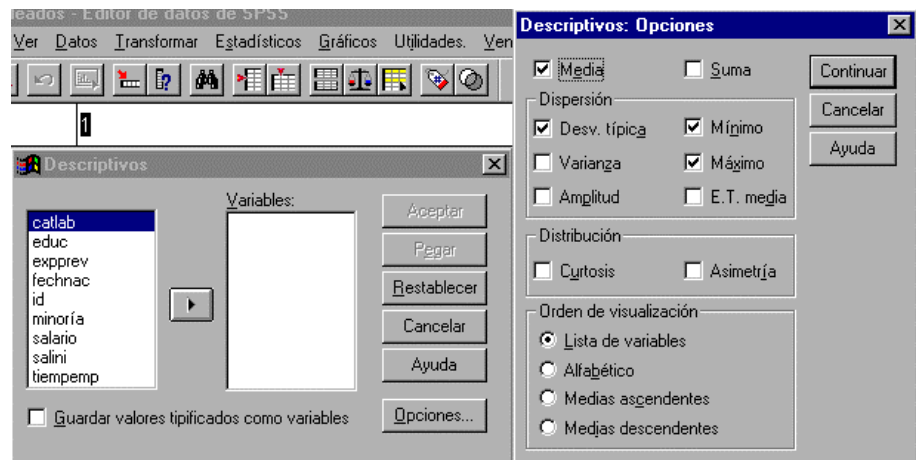
En el cuadro de diálogo Frecuencias, pulse en Gráficos.

14.- ANÁLISIS EXPLORATORIO DE DATOS.

A.- PROCEDIMIENTO: DESCRIPTIVOS.

El procedimiento Descriptivos muestra estadísticos de resumen univariados 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 las variables (el valor por defecto).

Cuando se guardan las puntuaciones z , se añaden a los datos del Editor de datos y están disponibles para gráficos de SPSS, listas de datos y análisis. Cuando las variables se registran en unidades diferentes (por ejemplo, producto interno



bruto per cápita y porcentaje de alfabetización), una transformación de puntuación z sitúa las variables en una escala común para una comparación visual más fácil.

Ejemplo. Si cada caso de los datos contiene los totales de ventas diarias de cada vendedor (por ejemplo, una entrada para Bob, una para Kim, una para Brian, etc.) recogidas cada día durante varios meses, el procedimiento Descriptivos puede calcular la media diaria de ventas para cada vendedor y ordenar los resultados del promedio de ventas de mayor a menor.

Estadísticos. Tamaño de muestra, media, mínimo, máximo, desviación típica, varianza, rango, suma, error típico de la media y curtosis y asimetría con sus errores típicos.

Para obtener frecuencias y estadísticos

Elija en los menús: Estadísticos / Resumir / Frecuencias...

Seleccione una o más variables categóricas o cuantitativas.

Si se desea es posible:

Pulsar en Estadísticos para obtener estadísticos descriptivos para las variables cuantitativas.

Pulsar en Gráficos para gráficos de barras, gráficos de sectores e histogramas.

Pulse en Formato para el orden en el que se muestran los resultados.

Consideraciones sobre los datos de descriptivos:

Datos. Utilice variables numéricas después de haberlas inspeccionado gráficamente para registrar errores, valores atípicos y anomalías de distribución. El procedimiento Descriptivos es muy eficaz para archivos grandes (miles de casos).

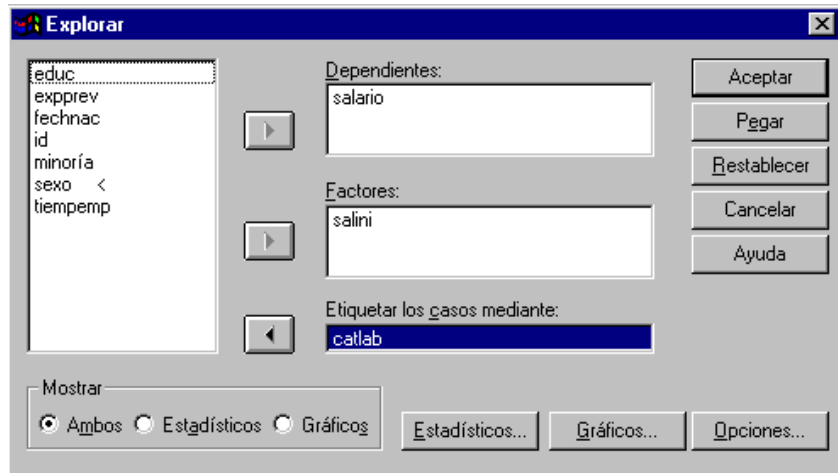
Supuestos. La mayoría de los estadísticos disponibles (incluyendo las puntuaciones z) se basan en la teoría normal y son adecuados para variables cuantitativas (medidas a nivel de razón o de intervalo) con distribuciones simétricas (se deben evitar variables con categorías no ordenadas o distribuciones asimétricas). La distribución de puntuaciones z tiene la misma forma que la de los datos originales; por tanto, el cálculo de puntuaciones z no es un remedio para los datos con problemas.

Procedimientos relacionados. Utilice el procedimiento Frecuencias para la mediana, la moda, los cuartiles, los percentiles y un histograma. Para calcular estadísticos de resumen para cada uno

de los distintos grupos de casos (por ejemplo, si quiere separar los estadísticos para hombres y mujeres o personas que viven en cuatro regiones diferentes de un país), utilice el procedimiento Explorar o Medias. También puede utilizar Segmentar archivo en el menú Datos.

B.- PROCEDIMIENTO: EXPLORAR.

El procedimiento Explorar produce estadísticos de resumen y representaciones gráficas, bien para todos los casos o bien de forma separada para grupos de casos. Existen numerosas razones para utilizar el procedimiento Explorar: inspección de datos, identificación de valores atípicos, descripción, comprobación de supuestos y caracterización de diferencias entre subpoblaciones



(grupos de casos). La inspección de los datos puede mostrar que existen valores inusuales, valores extremos, intervalos en los datos u otras peculiaridades. La exploración de datos puede ayudar a determinar si las técnicas estadísticas que está teniendo en consideración para el análisis de datos son adecuadas. La exploración puede indicar que necesita transformar los datos si la técnica necesita una distribución normal. O bien, el usuario puede decidir que necesita más pruebas no paramétricas.

Estadísticos y gráficos. Media, mediana, media recortada al 5%, error típico, varianza, desviación típica, mínimo, máximo, amplitud, amplitud intercuartil, asimetría y curtosis y sus errores típicos, intervalo de confianza para la media (y nivel de confianza especificado), percentiles, estimador-M de Huber, estimador en onda de Andrews, estimador-M redescendente de Hampel, estimador bponderado de Tukey, los cinco valores mayores y los cinco menores, el estadístico de Kolmogorov-Smirnov con el nivel de significación de Lilliefors para contrastar la normalidad y el estadístico de Shapiro-Wilk. Diagramas de caja, gráficos de tallo y hojas, histogramas, diagramas de normalidad, y diagramas de dispersión por nivel con la prueba de Levene y transformaciones.

Consideraciones sobre los datos de explorar:

Datos. El procedimiento Explorar se puede utilizar para las variables cuantitativas (nivel de medida de razón o de intervalo). Una variable de factor (utilizada para dividir los datos en grupos de casos) debe tener un número razonable de valores distintivos (categorías). Estos valores son de cadena corta o numéricos. La variable de etiquetas de caso, utilizada para etiquetar valores atípicos en diagramas de caja, pueden ser cadenas cortas, cadenas largas (los primeros 15 caracteres) o numérico.

Supuestos. La distribución de los datos no tiene que ser simétrica o normal.

Procedimientos relacionados. Si desea obtener pocos estadísticos de resumen simples para grupos de casos, utilice el procedimiento de medias. Para obtener recuentos, porcentajes de casos o percentiles específicos, utilice el procedimiento de frecuencias. Para calcular puntuaciones z, utilice el procedimiento Descriptivos.

Para explorar los datos

Elija en los menús: Estadísticos / Resumir / Explorar...

Seleccione una o más variables dependientes.

Si se desea, es posible:

Seleccionar una o más variables de factor, cuyos valores definirán grupos de casos.

Seleccione una variable de identificación para etiquetar casos.

Pulse en Estadísticos para estimadores robustos, valores atípicos, percentiles y tablas de frecuencia.

Pulse en Gráficos para histogramas, pruebas y gráficos de probabilidad normal y diagramas de dispersión por nivel con el estadístico de Levene.

Pulse en Opciones para el tratamiento de valores perdidos.

Explorar: estadísticos

Descriptivos. Estas medidas de dispersión y tendencia central se muestran por defecto. Las medidas de tendencia central indican la localización de la distribución; incluyen la media, mediana y la media recortada al 5%. Las medidas de dispersión muestran la disimilaridad de los valores; estos incluyen errores típicos, varianza, desviación típica, mínimo, máximo, amplitud y amplitud intercuartil. Los estadísticos descriptivos también incluyen medidas de la forma de la distribución; la asimetría y la curtosis se muestran con sus errores típicos. También se muestra el intervalo de confianza a un nivel del 95%; se puede especificar un nivel de confianza diferente.

Estimadores-M. Alternativas robustas a la mediana y a la media muestral para estimar el centro de la localización. Los estimadores calculados difieren a los pesos que aplican a los casos. Se visualizan el estimador-M de Huber's, el estimador en onda de Andrew, el estimador-M redescendente de Hampel y el estimador bponderado de Tukey.

Valores atípicos. Muestra los cinco valores mayores y los cinco menores, con etiquetas de caso.

Percentiles. Muestra los valores de los percentiles 5, 10, 25, 50, 75, 90 y 95.

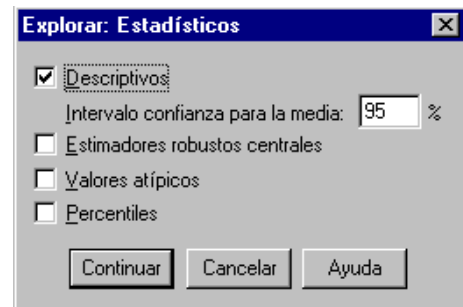
Pulse en Opciones para el tratamiento de valores perdidos.

Para especificar estadísticos para Explorar

Elija en los menús: Estadísticos / Resumir / Explorar...

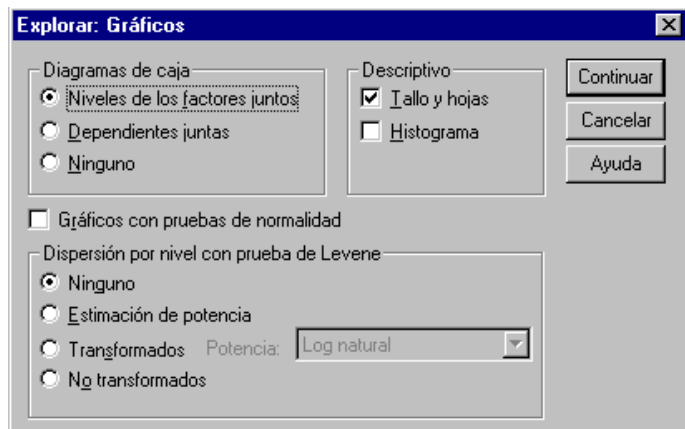
En el cuadro de diálogo Explorar, seleccione Ambos o Estadísticos en el grupo Mostrar para activar el botón Estadísticos.

Pulse en Estadísticos y seleccione uno o más estadísticos.



Explorar: gráficos

Diagramas de caja. Estas alternativas controlan la visualización de diagramas de caja cuando existe más de una variable dependiente. Niveles de los factores juntos genera una visualización diferente para cada variable dependiente. En una visualización, los diagramas de caja se muestran para cada grupo definido por una variable de factor. Dependientes juntas genera una visualización diferente para cada grupo definido por una variable de factor. En una visualización, se muestran los diagramas de



caja de cada variable dependiente. Esta visualización es particularmente útil cuando las diferentes variables representan una característica única medida en momentos distintos.

Descriptivo. El grupo Descriptivo permite seleccionar gráficos de tallo y hojas e histogramas.

Gráficos con pruebas de normalidad. Muestra la probabilidad normal y los gráficos de probabilidad normal sin tendencia. Se muestra el estadístico de Kolmogorov-Smirnov, con un nivel de significación de Lilliefors para contrastar la normalidad. El estadístico de Shapiro-Wilk se calcula para las muestras con 50 o menos observaciones.

Dispersión por nivel con prueba de Levene. Controla la transformación de datos para los diagramas de dispersión por nivel. Para todos los gráficos de dispersión por nivel, se muestra la pendiente de la línea de regresión y la prueba de Levene para la homogeneidad de varianzas. Si se selecciona una transformación, la prueba de Levene se basa en los datos transformados. Si no se selecciona una variable de factor, no se producen diagramas de dispersión por nivel. Estimación de potencia produce un gráfico de logaritmos naturales de las amplitudes intercuartiles frente a los logaritmos naturales de las medianas de todas las casillas, así como una estimación de la transformación de potencia para conseguir varianzas iguales en las casillas. Un diagrama de dispersión por nivel ayuda a determinar la potencia para una transformación para estabilizar (igualar) las varianzas de los grupos. Transformados le permite seleccionar una de las alternativas de potencia, quizás siguiendo las recomendaciones de la estimación de potencia y produce gráficos de datos transformados. Se trazan la amplitud intercuartil y la mediana de los datos transformados. No transformados produce gráficos de los datos brutos. Es equivalente a una transformación con una potencia de 1.

Para especificar los gráficos para Explorar

Elija en los menús: Estadísticos / Resumir / Explorar...

En el cuadro de diálogo Explorar, seleccione Ambos o Gráficos en el grupo Mostrar para activar el botón Gráficos.

Pulse en Gráficos y seleccione al menos un gráfico.

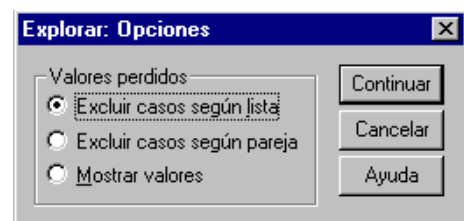
Explorar: opciones:

Valores perdidos. Controla el tratamiento de los valores perdidos.

Excluir casos según lista. Los casos con valores perdidos para variables de factor o dependientes se excluyen de todos los análisis. Es el valor por defecto.

Excluir casos según pareja. Los casos que no tengan valores perdidos para variables en un grupo (casilla) se incluyen en el análisis de ese grupo. El caso puede tener valores perdidos para variables utilizadas en otros grupos.

Mostrar valores. Los valores perdidos por variables de factor se tratan como una categoría diferente. Todos los resultados se generan para esta categoría adicional. Las tablas de frecuencia incluyen las categorías para los valores perdidos. Los valores perdidos para una variable de factor se incluyen pero se etiquetan como perdidos.



Para especificar las opciones de Explorar

Elija en los menús: Estadísticos / Resumir / Explorar...

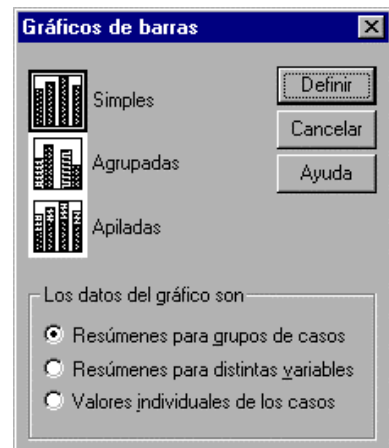
En el cuadro de diálogo Explorar, pulse en Opciones.

15.- REPRESENTACIONES GRÁFICAS.

A.- DIAGRAMAS DE BARRAS, LÍNEAS, ÁREAS Y SECTORES.

Los diagramas de este apartado suponen una representación gráfica de la distribución de frecuencias de una variable. Seleccionamos de **Gráficas** en la barra de menús, el procedimiento correspondiente: Barras..., Líneas..., Áreas..., Sectores...

En los gráficos de barras aparece el cuadro de diálogo de la derecha, en el que las alternativas que se ofrecen se refieren a dos aspectos; en el primero se puede establecer la forma general del diagrama:



Simple: las barras representan elementos del mismo conjunto.

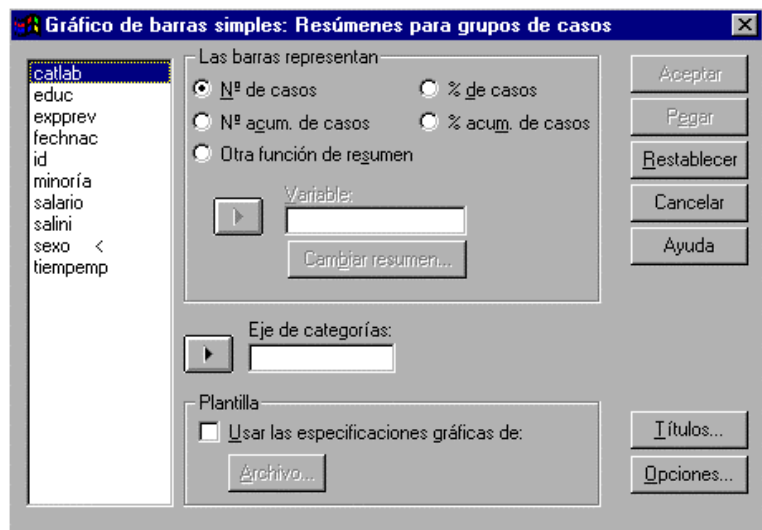
Agrupada: las barras, agrupadas, representan elementos de distintos conjuntos.

Apilada: las barras, apiladas, representan elementos de distintos conjuntos.

En el segundo, en **Los datos del gráfico son** se puede seleccionar la estructura de los datos:

Resúmenes para grupos de casos: cada barra representa la frecuencia de cada categoría de una variable. Ya en su cuadro de diálogo, en **Eje de Categorías** se selecciona la variable cuya distribución de frecuencias se desea representar.

(En los diagramas de barras agrupados, **Agrupada**, y apilados, **Apilada**, hay un grupo de barras por cada uno de sus valores). Por defecto, se obtiene la gráfica en que cada barra representa la frecuencia absoluta de cada categoría. Sin embargo se pueden seleccionar otros tipos de frecuencias en **Las barras representan:**



Nº de casos: Por defecto, frecuencia absoluta de cada categoría.

Nº acum. de casos: frecuencia absoluta acumulada de cada categoría.

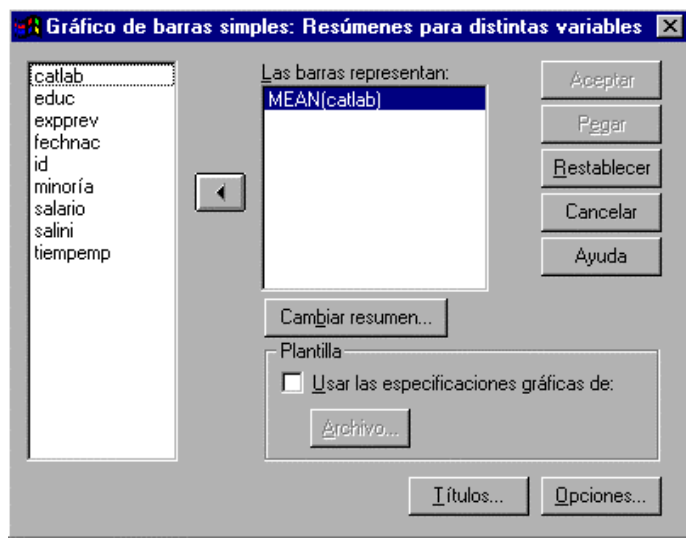
% de casos: % de casos de cada categoría sobre el total.

% acum. de casos: % acumulado de casos de cada categoría.

Otra función de resumen: el eje vertical, así como la altura de cada barra se asocia a otra variable.

Como novedad, los diagramas de barras agrupados y apilados ofrecen la posibilidad de seleccionar una nueva variable para determinar las barras dentro de cada grupo o de cada apilamiento; esto se realiza a través de **Definir grupos por** y de **Definir aplicaciones por** respectivamente, apareciendo tal variable y sus valores en la leyenda adjunta al diagrama ya obtenido.

Resúmenes para distintas variables: cada barra representa una misma característica de dos o más variables especificadas previamente en **Las barras representan**. Por defecto, cada barra refleja la suma de los valores de cada variable según



todos los casos; está claro que las variables por examinar deben ser cuantitativas.

Nuevamente se presenta la posibilidad de utilizar otro estadístico utilizando **Cambiar resumen**.

En los diagramas de barras agrupados y apilados se puede, además, seleccionar en **Eje de categorías** una nueva variable a partir de la cual, y para cada uno de sus valores, se crea un grupo o un apilamiento de barras. Ahora en la leyenda adjunta al diagrama aparecen las variables seleccionadas en **Las barras representan**.

Valores individuales de los casos: cada barra representa un caso.

En **Etiquetas de categorías** se puede especificar la etiqueta de cada barra:

Nº de caso: por defecto, cada barra es etiquetada por el número del caso a que corresponde.

Variable: cada barra es etiquetada por el valor que su correspondiente caso adopta en la variable ahora declarada.

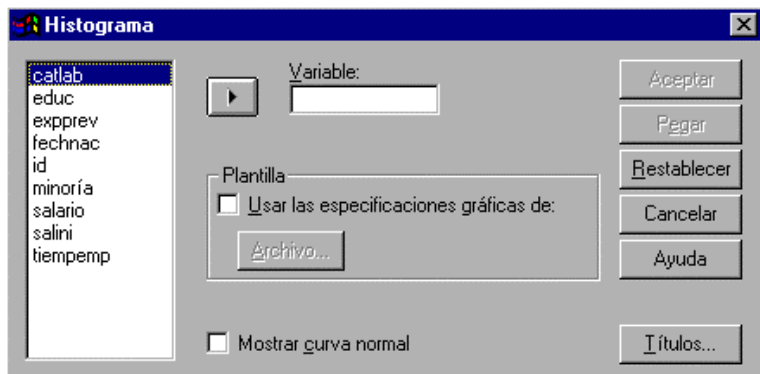
Los diagramas de líneas, áreas y sectores representan, bajo otro aspecto, cuestiones similares a los de barras.

B.- HISTOGRAMAS.

El acceso a este tipo de gráficas se produce a través de: Gráficas / Histograma...

Tras seleccionar una sola variable, desde luego cuantitativa, en **Variable**, se obtienen los resultados por defecto.

A la gráfica se puede superponer la correspondiente curva normal, es decir, aquella con la misma media y varianza que la distribución empírica en cuestión; ello activando **Ver curva normal**.



EDICIÓN DE GRÁFICOS

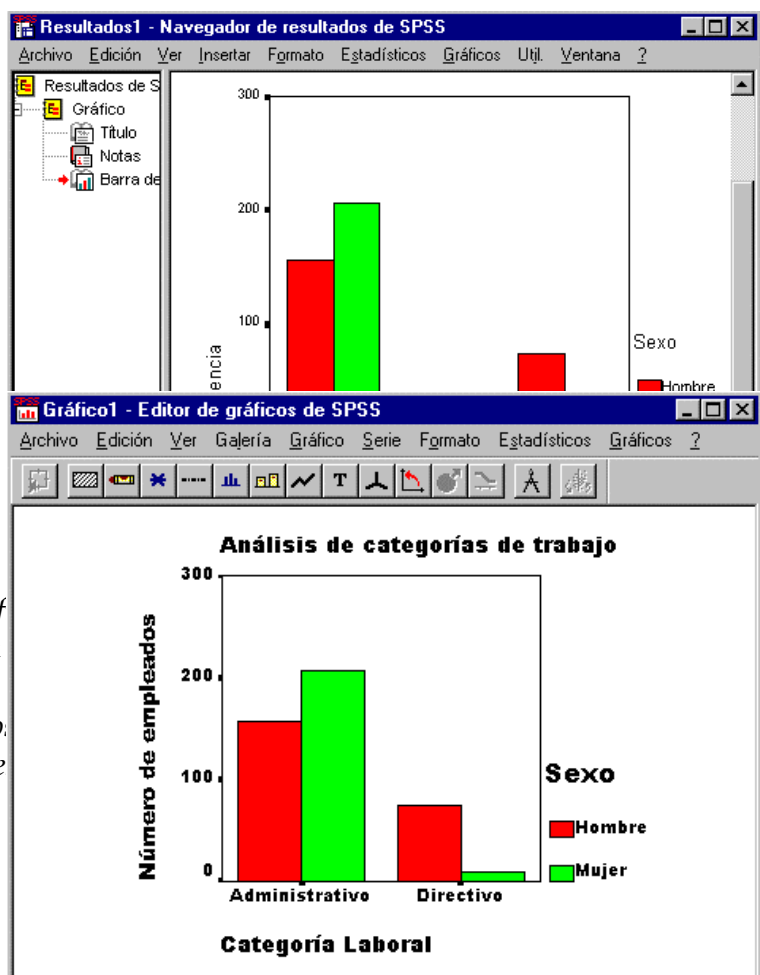
Una vez que ha creado un gráfico, hay muchos atributos que se pueden editar para cambiar su aspecto. Puede cambiar títulos, etiquetas, fuentes, colores, eliminar categorías, cambiar el rango de ejes de escala, intercambiar ejes y cambiar el tipo de gráfico.

Pulse dos veces en el gráfico que desee y éste aparecerá en una ventana de gráficos.

Puede editar el gráfico desde los menús, desde la barra de herramientas o pulsando dos veces en el objeto que desee editar.

Para seleccionar objetos en el gráfico aparecerá un cuadro de selección

Si pulsa dos veces en un objeto correspondiente. Por ejemplo, si pulsa dos veces en un objeto que controla las categorías y agrupaciones



Para eliminar una categoría, pulse dos veces en una de las barras del gráfico. En nuestro ejemplo de barras agrupadas, podemos eliminar un sexo o bien una categoría de empleo. Mueva la categoría que desee eliminar a la lista Omitir.

Para cambiar los intervalos y el rango del eje de escala, pulse dos veces en el eje. Puede cambiar el rango de la muestra y puede cambiar los valores que se muestran en el eje.

Para cambiar una etiqueta de gráfico, pulse dos veces en la etiqueta que desee cambiar, se abrirá el cuadro de diálogo correspondiente, introduzca un nuevo título y pulse Aceptar.

Utilice la herramienta de texto para cambiar las características de las fuentes del objeto seleccionado.

Para insertar un título, seleccione Título en el menú Gráfico. Introduzca un título en el cuadro de diálogo y pulse en Aceptar.

Utilice la herramienta de colores para cambiar los colores de un gráfico. Pulse en un elemento del gráfico. Pulse a continuación en la herramienta de colores de la barra de herramientas. Seleccione un color y pulse en Aceptar. Puede dejar el cuadro de diálogo Colores abierto para cambiar otros colores del gráfico.

Para intercambiar los ejes de un gráfico, pulse en la herramienta intercambiar ejes de la barra de herramientas.

Para cambiar el tipo de gráfico, seleccione un nuevo tipo de gráfico en el menú Galería. Se abre un cuadro de diálogo donde puede seleccionar el tipo específico de gráfico que desea.

16.- USO DEL ASESOR ESTADÍSTICO.

Si no está familiarizado con SPSS o con los procedimientos estadísticos disponibles en SPSS, el Asesor estadístico puede ayudarle a la iniciación mediante preguntas simples, lenguaje no técnico y ejemplos visuales que le ayudan a seleccionar las funciones básicas de estadísticos y gráficos que mejor se adaptan a los datos.

El Asesor estadístico cubre sólo un subconjunto selecto de procedimientos en el sistema base de SPSS. Está diseñado para proporcionar una asistencia general para muchas de las técnicas estadísticas básicas que se utilizan normalmente

Para utilizar el Asesor estadístico seleccione **Asesor estadístico** en el menú **Ayuda**. El Asesor estadístico le pregunta cuestiones sencillas y le ofrece opciones no técnicas. Los ejemplos visuales le ayudan a encontrar el gráfico o procedimiento estadístico apropiado para sus datos. El botón **Ayuda** proporciona instrucciones detalladas sobre cómo usar el cuadro de diálogo correspondiente al procedimiento. Y el

botón *Finalizar* le lleva directamente al cuadro de diálogo correspondiente al procedimiento.

17.- IMPORTACIÓN Y EXPORTACIÓN.

Los archivos de bases de datos son, lógicamente, muy similares a los archivos de datos de SPSS. Las normas generales siguientes se aplican a los archivos de dBASE:

Los nombres de campos se traducen automáticamente a nombres de variable de SPSS.

Los nombres de campos deben estar de acuerdo con las convenciones de denominación de las variables de SPSS. Los nombres de campos de más de ocho caracteres se truncan. Si los primeros ocho caracteres del nombre del campo forman un nombre único, el campo será eliminado.

Los dos puntos en los nombres de campos de dBASE se convierten en subrayado.

Se incluyen los registros marcados para ser eliminados que aún no se han purgado. SPSS crea una nueva variable de cadena, D_R, que contiene un asterisco para los casos marcados para ser eliminados.

Un archivo de datos de SPSS es rectangular. Los límites (o dimensiones) de los archivos de datos se determinan por el número de casos (filas) y variables (columnas). No hay casillas vacías dentro de las fronteras del archivo de datos. Todas las casillas tienen un valor, incluso si ese valor es en blanco. Para leer los datos de las hojas de cálculo se aplican las siguientes normas:

Filas y columnas. Las filas son casos y las columnas son variables.

Tipo y ancho de datos. Los tipos y anchos de los datos para cada variable se determina según el ancho de columna y el tipo de datos de la primera casilla de la columna. Valores de otros tipos se convierten en valor perdido del sistema. Si la primera casilla de datos de la columna está en blanco, se utilizará el tipo de datos por defecto global para la hoja de cálculo (normalmente numérico).

Casillas en blanco. Para variables numéricas, las casillas en blanco se convierten en el valor perdido del sistema indicado con un punto. Para variables de cadenas, un espacio en blanco es un valor de cadena válido y las casillas en blanco se tratan como valores de cadenas válidos.

Nombres de variable. Si no se leen los nombres de variable de la hoja de cálculo, SPSS utiliza las letras de las columnas (A, B, C,...) para nombres de variable para archivos de Excel y de Lotus. Para archivos de SYLK y archivos de Excel guardados en formato de visualización R1C1, SPSS utiliza el número de la columna precedido por la letra C para nombres de variable (C1, C2, C3,...).

Puntos que recordar sobre la lectura de archivos de hojas de cálculo:

El tipo de datos para cada variable se define por los valores de la primera fila de la hoja de cálculo.

Si la primera fila contiene cabeceras de columnas, pulse en **Leer los nombres de variable** en el cuadro de diálogo **Opciones** que se abre después de seleccionar un archivo de hoja de cálculo. Los valores de la primera fila de la hoja de cálculo se usarán como nombres de variables en SPSS. Y el tipo de datos se definirá mediante valores de datos en la segunda fila. Si los datos que desea leer de la hoja de cálculo no comienzan en la primera fila ni en la primera columna... Introduzca el rango de casillas que desea leer de la hoja de cálculo, utilizando el mismo formato para especificar el rango de casillas que usaría en el programa de hoja de cálculo.

Si define etiquetas de valor y variable descriptivas para sus datos en SPSS puede aplicar esas etiquetas a archivos de hojas de cálculo similares que se puedan leer en SPSS. Seleccione

Aplicar diccionario de datos en el menú **Archivo**. Y seleccione el archivo de datos de SPSS de la lista de archivos. SPSS compara los nombres de variables entre los dos archivos y aplica etiquetas del archivo guardado a las variables con nombres coincidentes en el nuevo archivo.

SPSS puede enviar datos a otras aplicaciones en los siguientes formatos: SPSS, Excel 4.0, Lotus 1-2-3 versión 3, SYLK, delimitado por tabuladores y dBASE IV.

Los gráficos se pueden exportar en los siguientes formatos: metaarchivo de Windows, mapa de bits de Windows, PostScript encapsulado, JPEG, TIFF, CGM o PICT de Macintosh.

Para copiar resultados de SPSS en otra aplicación como metaarchivo, pulse en el elemento del panel de contenido o titulares del Navegador de resultados para seleccionarlo. Seleccione **Copiar** en el menú **Edición**. En la otra aplicación, seleccione **Pegado especial** en el menú **Edición**. En el cuadro de diálogo Pegado especial, seleccione **Imagen**.

El metaarchivo retiene todas las características de fuentes y los estilos de bordes de los elementos de cuando se copiaron. Sin embargo, un metaarchivo sólo contiene la información visible en el momento de haber copiado los elementos del Navegador de resultados. Los metaarchivos no contienen la información de capas o categorías ocultas.

Para copiar una tabla pivote en otra aplicación como texto sin formato pulse en el elemento del panel de contenido o titulares del Navegador de resultados para seleccionarlo. Seleccione **Copiar** en el menú **Edición**. En la otra aplicación, seleccione **Pegado especial** en el menú **Edición**. En el cuadro de diálogo Pegado especial, seleccione **Texto sin formato**. El texto sin formato contiene tabulaciones entre las columnas. Puede alinear las columnas ajustando las tabulaciones en la otra aplicación.

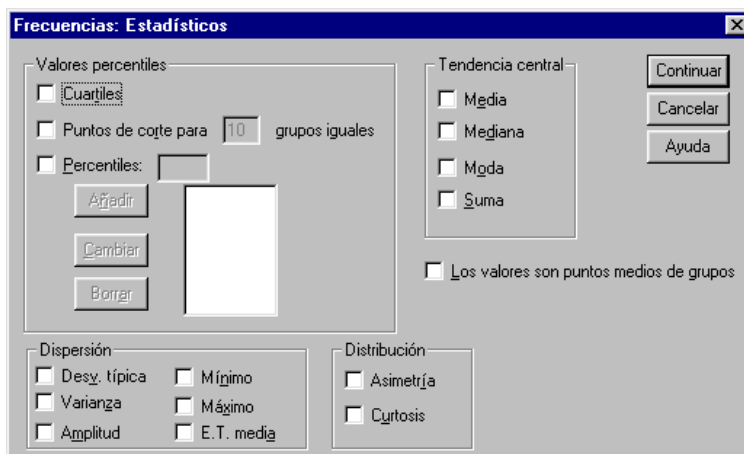
APLICACIONES ESTADÍSTICAS

En lo que sigue se desarrollara la parte de aplicaciones estadísticas siguiendo la estructura de los cursos básicos de estadística. Abordaremos los siguientes casos:

- ESTADÍSTICA DESCRIPTIVA
- EST. DESCRIPTIVA CON UNA VARIABLE
- EST. DESCRIPTIVA CON DOS VARIABLES
- INFERENCIA ESTADÍSTICA
- TEST DE HIPÓTESIS

ESTADÍSTICA DESCRIPTIVA:

El menú del SPSS para la realización de estadística descriptiva se encuentra en el menú de descriptiva en la parte de frecuencias pudiendo elegir entre los siguientes parámetros estadístico:



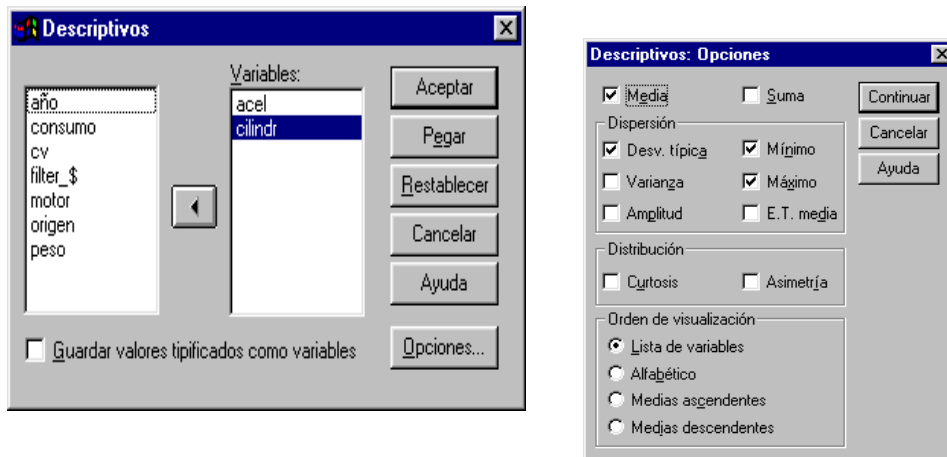
Se pueden elegir los percentiles que se quieran sin más que añadir en la casilla respectiva el valor deseado.

Debe de tenerse en cuenta que para el calculo de la desviación típica y varianza el programa utiliza la cuasidesviación o cuasivarianza, esto es, divide por $n-1$ y no por n .

*La medida **E.T. media** se refiere al Error Típico de la media, esto es, la desviación típica entre la raíz de n .*

Esta opción presenta la posibilidad de realizar gráficas, incluso con ajuste del histograma a la distribución Normal de media la media de la muestra y desviación la cuasidesviación típica de la muestra.

La opción Resumir/descriptivos viene a ser lo misma que la anterior, pero sin la opción de gráficos. En este caso se puede, definidas las variables que se van a estudiar configurar en Opciones



los estadísticos a tratar.

En el menú Resumir/explorar se puede hacer un análisis descriptivo de la variable, veamos el caso de estudiar la variable aceleración del fichero coches:

Descriptivos:

Aceleración 0 a 100 km/h (segundos) Media 15,50

Intervalo de confianza para la media al 95%

Límite inferior 15,22

Límite superior 15,77

Media recortada al 5% 15,46

Mediana 15,50

Varianza 7,958

Desv. típ. 2,82

Mínimo 8

Máximo 25

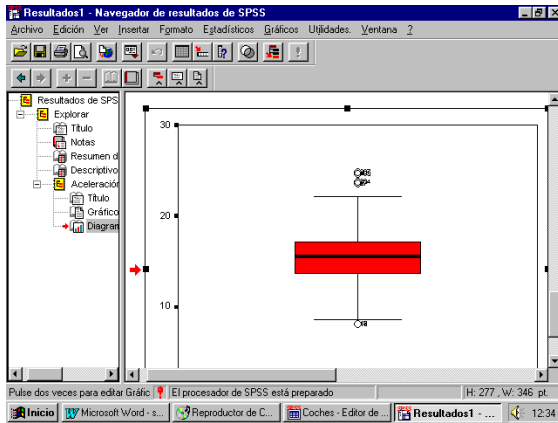
Rango 17

Amplitud intercuartil 3,53

Asimetría ,211 ,121

Curtosis ,389 ,242

Teniendo la posibilidad de calcular intervalos de confianza para la media, tablas de frecuencias agrupadas (agrupar una variable en intervalos de clase), medias recortadas, diagramas de caja y bigotes.



Puede observarse que cuando se realiza un proceso estadístico aparece una ventana a la izquierda de los resultados que permite moverlos por ellos.

En el caso de la figura de arriba se observa un diagrama de cajas, donde aparecen tres valores atípicos, estos diagramas muestran la forma de la distribución, estando limitados por el primer y tercer cuartil, y marcando el valor de la mediana, extendiéndose los bigotes hasta 1.5 veces la longitud de la caja.

Los valores anómalos o atípicos pueden resaltarse identificándolos en la base de datos.

Análisis de 2 variables. Regresión y correlación.

Dada una variable dependiente y un conjunto de una o más variables independientes, todas cuantitativas, la regresión trata de obtener una función que permita explicar o predecir el valor de la variable dependiente.

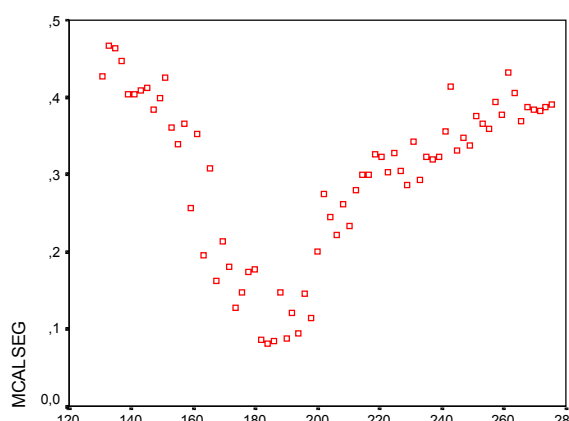
El SPSS estudia estos problemas en el menú: Estadística/correlación/bivariadas y en Estadística/regresión/curvilínea. También se hace un estudio en el menú: Gráficas/dispersión.

Se pueden calcular varios coeficientes de correlación, por defecto calcula el coeficiente de Pearson.

Calcula la matriz de **varianzas-covarianzas** entre las variables a estudio. Teniendo en cuenta que cuando aparecen valores perdidos en un determinada variable, se puede excluir los casos según pareja o excluir casos según lista, eliminando los casos de los valores perdidos en la lista o la pareja.

El coeficiente de correlación r mide el grado de dependencia lineal entre dos variables, tomando valores entre -1 y 1 . Valores próximos a 1 o -1 indican dependencia lineal y próximos a 0 independencia lineal.

Veamos un ejemplo donde la variable independiente es la temperatura y la dependiente es el nivel de injerto de un material compuesto de polipropileno, medida en mcal/seg. Se tiene un total de 72 datos muestrales los cuales se pueden representar mediante el siguiente diagrama de dispersión:

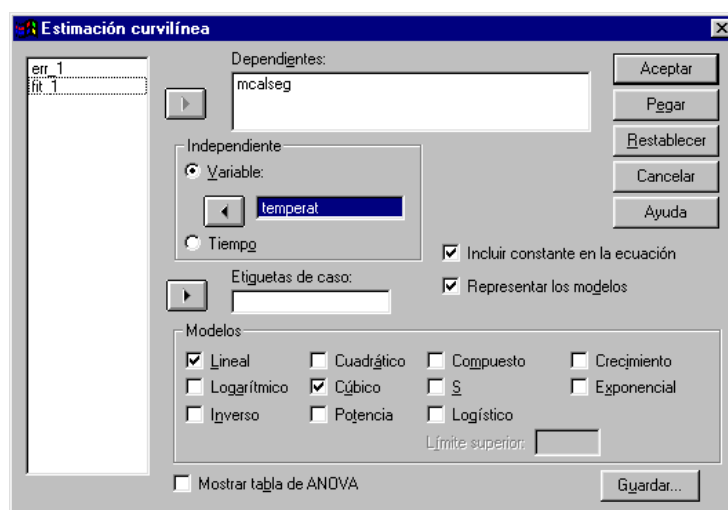
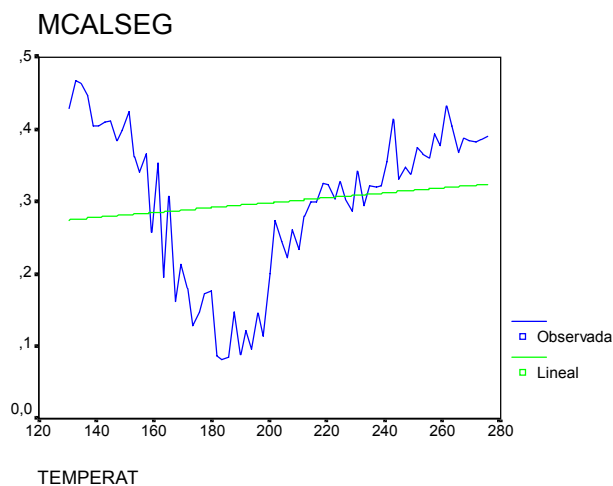


Puede observarse que la correlación lineal es muy baja

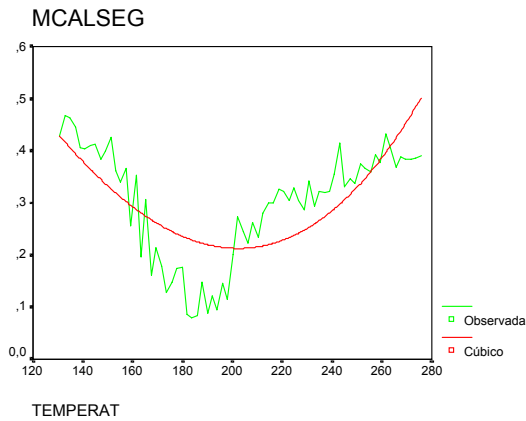
Coefficientes^a

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	,230	,062		3,738	,000
	TEMPERAT	3,402E-04	,000	,136	1,147	,255

a. Variable dependiente: MCALSEG



En estimación curvilínea permite los modelos de regresión indicados arriba, veamos el ajuste a un polinomio de grado 3.



Este caso es necesario el ensayo con métodos de estimación no paramétrico tipo método núcleo, pues sale un mal ajuste curvilíneo.

Veamos finalmente como se puede predecir el salario actual en función de las variables independientes salario inicial y antigüedad en la empresa en el fichero de empleados.

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,886 ^a	,785	,784	\$7,936.14

a. Variables predictoras: (Constante), Meses desde el contrato, Salario inicial

Test de Hipótesis Paramétricos.

Comparar medias:

El programa SPSS ofrece el siguiente menú para la comparación de medias:

- Comparar medias/Medias
- Comparar medias/Prueba T para una muestra
- Comparar medias/Prueba T para muestras independientes
- Comparar medias/Prueba T para muestras relacionadas
- Comparar medias/ANOVA de un factor

Estas cinco opciones del submenú **Comparar medias** ofrecen contrastes de hipótesis sobre las diferencias entre las medias de variables cuantitativas.

El submenú **Medias** proporciona una variedad de estadísticos descriptivos para variables cuantitativas en casillas definidas por una o más variables de agrupación (por ejemplo, los ingresos medios en casillas definidas por el sexo y el nivel cultural). La relación distribución por edad y sexo:

Informe

Fecha de nacimiento

Hombre	Media	04/28/57
	N	257
	Desv. tıp.	*****
Mujer	Media	02/11/56
	N	216
	Desv. tıp.	*****
Total	Media	10/08/56
	N	473
	Desv. tıp.	*****

La prueba T de Student para una muestra comprara la media de una única variable respecto a un valor constante. Ha de verificarse que la variable que se compare sea Normal. La hipótesis a contrastar es: $H_0: \mu = m$. Si el p-valor asociado al estadístico es menor que el nivel de significación del contraste se rechaza la hipótesis Nula. En caso contrario no se rechazaría. Por ejemplo en un fichero de empleados de una empresa donde se tiene la variable edad, podemos contrastar si esta tiene media 45, esto es:

$H_0: \mu = 45$

Dentro de las opciones se permite el cálculo del intervalo de confianza para la media, en nuestro ejemplo la media muestral es 37.18, por lo que el resultado del contraste nos proporciona un p-valor = 0, lo que lleva a rechazar de plano la hipótesis nula.

Otro contraste es suponer el salario actual en 35.000 \$.

Prueba para una muestra

	Valor de prueba = 35000					
	t	gl	Sig. (bilateral)	Diferencia de medias	Intervalo de confianza para la diferencia	
					Inferior	Superior
Salario actual	-,740	473	,460	-\$580.43	-\$2,121.60	\$960.73

Como el nivel de significación del test da 0.46 aceptaríamos el valor 35.000 como valor estimado del salario medio.

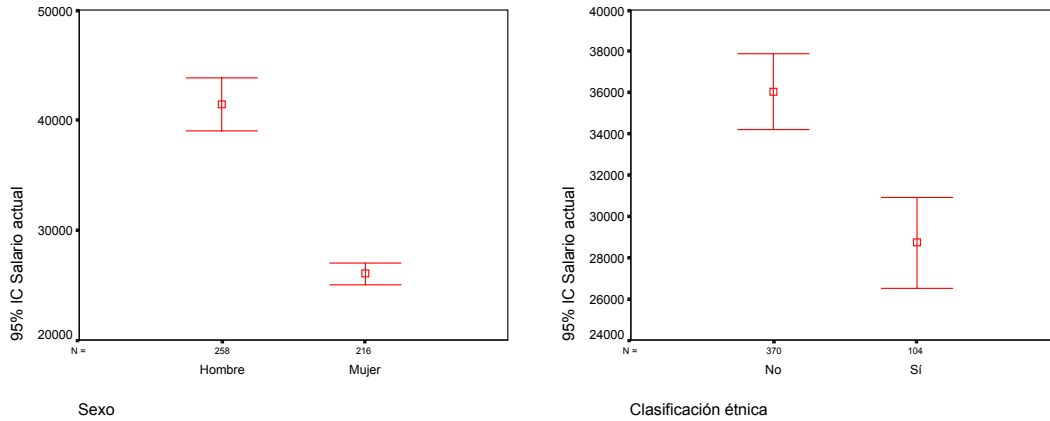
La prueba T para muestras independientes (o prueba t para dos grupos) compara la media de una única variable para sujetos de un grupo con la de otro grupo. Responde a preguntas del tipo: Dadas dos muestras independientes de datos normales, ¿tienen la misma media?, esto es:

$H_0: \mu_x = \mu_y$

Por ejemplo con el fichero de empleados podría contrastar la variable Salario actual según el sexo. Para lo cual podría dividir la muestra según la variable sexo, utilizando esta como variable de agrupación, definiendo los grupos asignando a varones 0 y a mujeres 1. Se realiza el test, para lo cual calcula primero un test de igualdad de varianzas, test de Levene, como en este caso sale el p-valor 0, las varianzas son distintas por lo que en el test de las medias uso los resultados de varianzas distintas. La conclusión es de rechazo de igualdad de salarios por sexo.

Otro contraste que podríamos llevar a cabo con la misma variable es ver si hay diferencias significativas del salario para mayores de 45 años. Rompiendo la variable edad en dos grupos, en Definir grupos utilizamos punto de corte 45.

Un procedimiento gráfico para este test consiste en usar las Barras de error, proporcionando intervalos de confianza para la media al 95%.



En ambos casos se aprecia diferencia significativa en el salario tanto por sexo como por pertenecer o no a minoría étnica.

La prueba T para muestras relacionadas comprueba si la media de diferencias por casos entre dos variables difiere de 0. El diseño del estudio para esta prueba implica medir cada sujeto dos veces, antes y después de una intervención. Por ejemplo, se puede realizar un test a alumnos antes de asistir a un curso y después de haberlo recibido. Comparando si se ha mejorado con la realización de dicho curso.

En el caso de la variable salario inicial y salario actual, podríamos comparara si el salario inicial medio es igual al salario actual medio.

Prueba de muestras relacionadas

		Diferencias relacionadas				t	gl	Sig. (bilateral)	
		Media	Desviación típ.	Error típ. de la media	Intervalo de confianza para la diferencia				
					Inferior				Superior
Par 1	Salario actual - Salario inicial	*****	\$10,814.62	\$496.73	*****	*****	35,036	473	,000

En este caso nos proporciona un p-valor igual a 0, por lo que rechazaría la igualdad de medias de salarios inicial y actual.

La prueba t de Student para muestras relacionadas compara la dispersión de los datos, estudiando la diferencia entre las medias de las dos variables a estudio.

significativos.

*Cualquiera de estos test lo que hace es cruzar un grupo con otro indicando mediante un * los grupos estadísticamente significativas.*