

Departamento Salud Pública y Materno-Infantil

Facultad de Medicina

Universidad Complutense de Madrid



Colección

**“Guías para la iniciación a la investigación en
Ciencias de la Salud”**

Nº 3: Introducción a SPSS para Windows

Editado por la Unidad Docente de Medicina Preventiva y Salud Pública

Departamento de Salud Pública y Materno-infantil

Facultad de Medicina

Universidad Complutense de Madrid

Madrid, 2018

ISSN 2605-5252

Autores David Martínez Hernández¹
Napoleón Pérez Farinós²
Ana Martínez Santos⁴
Juana María Santos Sancho¹
Julia Wörnberg³

¹ Departamento de Salud Pública y Materno-infantil. Universidad Complutense de Madrid. España

² Departamento de Salud Pública y Psiquiatría. Universidad de Málaga. España

³ Departamento de Enfermería. Universidad de Málaga. España

⁴ School of Health Sciences. Salford University. Mánchester. Reino Unido

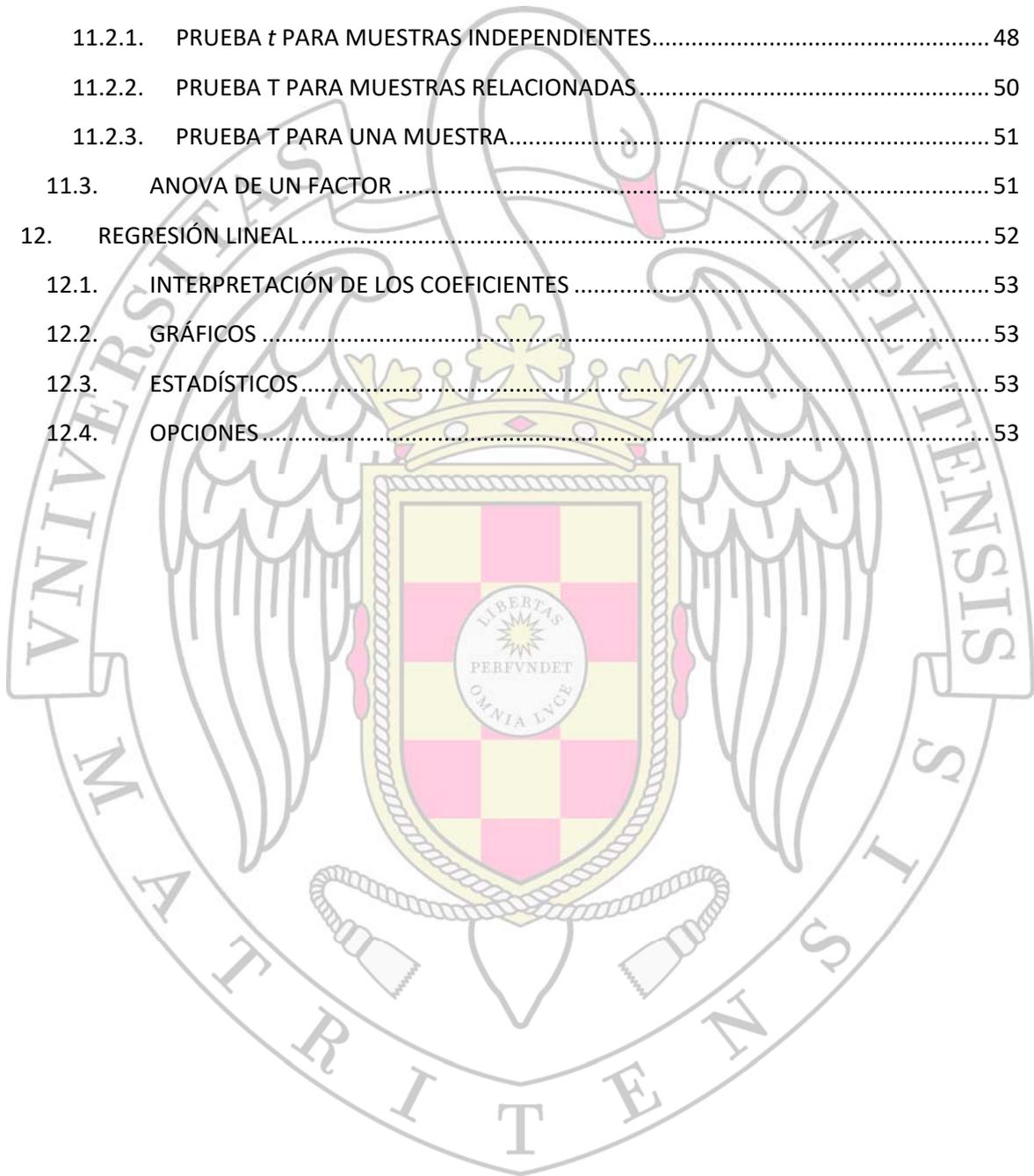
Contenido

1.	CARACTERÍSTICAS BÁSICAS Y ESTRUCTURA DE TRABAJO EN SPSS	5
1.1.	VENTANAS DE SPSS	5
1.1.1.	Editor de datos:	5
1.1.2.	Visor de resultados:.....	5
1.1.3.	Editor de sintaxis:	5
1.1.4.	Editor de gráficos	6
1.2.	BARRA DE ESTADO	6
1.3.	VENTANA DESIGNADA Y VENTANA ACTIVA	6
1.4.	BARRAS DE HERRAMIENTAS.....	6
1.5.	CUADROS DE DIÁLOGO	7
1.5.1.	Subcuadros de diálogo	7
1.5.2.	Ejecución de un procedimiento	8
1.6.	ANÁLISIS DE DATOS CON SPSS	8
2.	EDITOR DE DATOS	9
2.1.	VISTA DE DATOS	9
2.2.	VISTA DE VARIABLES.....	9
2.2.1.	nombre:.....	10
2.2.2.	Tipo de variable	10
2.2.3.	Anchura:	11
2.2.4.	ETIQUETA	11
2.2.5.	ETIQUETAS DE VALOR:	11
2.2.6.	VALORES PERDIDOS:	11
2.2.7.	FORMATO DE COLUMNA	12
2.2.8.	ALINEACIÓN.....	12
2.2.9.	MEDIDA:	12
2.3.	INTRODUCCIÓN DE DATOS.....	12
2.4.	IMPORTACIÓN Y LECTURA DE ARCHIVOS	13
3.	FUSIÓN DE ARCHIVOS	14
3.1.	AÑADIR VARIABLES	14
3.2.	AÑADIR CASOS	17
4.	TRANSFORMACIÓN DE DATOS.....	19

4.1.	CALCULAR VARIABLE	19
4.2.	CONTAR APARICIONES DENTRO DE LOS CASOS.....	21
4.3.	RECODIFICAR VALORES	21
4.3.1.	RECODIFICAR EN LAS MISMAS VARIABLES.....	22
4.3.2.	RECODIFICAR EN DISTINTAS VARIABLES	22
4.4.	ASIGNACIÓN DE RANGOS.....	23
4.5.	CATEGORIZACIÓN DE VARIABLES.....	23
4.6.	RECODIFICACIÓN AUTOMÁTICA	23
5.	GESTIÓN Y TRANSFORMACIÓN DE ARCHIVOS.....	24
5.1.	ORDENAR CASOS.....	24
5.2.	TRANSPONER ARCHIVOS.....	24
5.3.	SEGMENTAR ARCHIVO	25
5.4.	SELECCIONAR CASOS.....	26
5.4.1.	SELECCIONAR UNA MUESTRA ALEATORIA DE CASOS.....	26
5.4.2.	SELECCIONAR CASOS EN FUNCIÓN DE UN RANGO.....	27
5.4.3.	PONDERAR CASOS.....	27
6.	VISOR DE RESULTADOS	29
6.1.	MOSTRAR Y OCULTAR RESULTADOS.....	29
6.2.	INSERTAR.....	30
6.3.	FORMATO.....	30
6.4.	CAMBIAR EL NIVEL DE LOS TITULARES.....	30
6.5.	GUARDAR LOS RESULTADOS.....	31
7.	SINTAXIS	32
7.1.	Archivo diario	32
7.2.	Ejecución de comandos desde la sintaxis.....	33
8.	ESTADÍSTICOS DESCRIPTIVOS.....	34
8.1.	FRECUENCIAS	34
8.1.1.	ESTADÍSTICOS.....	34
8.1.2.	GRÁFICOS	35
8.1.3.	FORMATO.....	36
8.2.	DESCRIPTIVOS	36
8.3.	EXPLORAR.....	37
8.3.1.	ESTADÍSTICOS.....	37

8.3.2.	GRÁFICOS	37
8.3.3.	OPCIONES.....	38
8.4.	TABLAS DE CONTINGENCIA.....	38
8.4.1.	ESTADÍSTICOS.....	39
8.4.2.	CASILLAS	39
8.4.3.	FORMATO.....	39
8.5.	RAZÓN	39
9.	REPRESENTACIONES GRÁFICAS.....	40
9.1.	VARIABLES CUANTITATIVAS CONTINUAS:.....	41
9.1.1.	HISTOGRAMA:.....	41
9.1.2.	EDITOR DE GRÁFICOS	41
9.1.3.	BARRAS DE ERROR.....	42
9.1.4.	DIAGRAMA DE CAJA O BOXPLOT	42
9.1.5.	CONSTRUCCIÓN DEL DIAGRAMA DE CAJA.....	42
9.2.	VARIABLES CUALITATIVAS Y CUANTITATIVAS DISCRETAS.....	43
9.2.1.	DIAGRAMA DE BARRAS	43
9.2.2.	DIAGRAMA SIMPLE	43
9.2.3.	DIAGRAMA AGRUPADO	43
9.2.4.	DIAGRAMAS APILADOS	44
9.2.5.	RESÚMENES PARA DISTINTAS VARIABLES	44
9.2.6.	VALORES INDIVIDUALES DE LOS CASOS	44
9.3.	DIAGRAMA DE PARETO	44
9.4.	GRÁFICOS DE LÍNEAS.....	44
9.5.	DIAGRAMA DE SECTORES.....	44
10.	CORRELACIÓN	45
10.1.	DIAGRAMA DE DISPERSIÓN.....	45
10.1.1.	DIAGRAMA SIMPLE	45
10.1.2.	DIAGRAMA SUPERPUESTO	46
10.1.3.	DIAGRAMA MATRICIAL	46
10.1.4.	GRÁFICO 3-D.....	46
10.2.	CORRELACIONES BIVARIADAS	46
10.2.1.	Coeficiente de correlación lineal de Pearson (r):.....	46
10.2.2.	Rho de Spearman y tau- <i>b</i> de Kendall.	47

10.3.	CORRELACIONES PARCIALES	47
11.	COMPARACIÓN DE MEDIAS	48
11.1.	MEDIAS.....	48
11.1.1.	OPCIONES.....	48
11.2.	PRUEBAS t DE STUDENT	48
11.2.1.	PRUEBA t PARA MUESTRAS INDEPENDIENTES.....	48
11.2.2.	PRUEBA T PARA MUESTRAS RELACIONADAS.....	50
11.2.3.	PRUEBA T PARA UNA MUESTRA.....	51
11.3.	ANOVA DE UN FACTOR	51
12.	REGRESIÓN LINEAL.....	52
12.1.	INTERPRETACIÓN DE LOS COEFICIENTES	53
12.2.	GRÁFICOS	53
12.3.	ESTADÍSTICOS.....	53
12.4.	OPCIONES.....	53



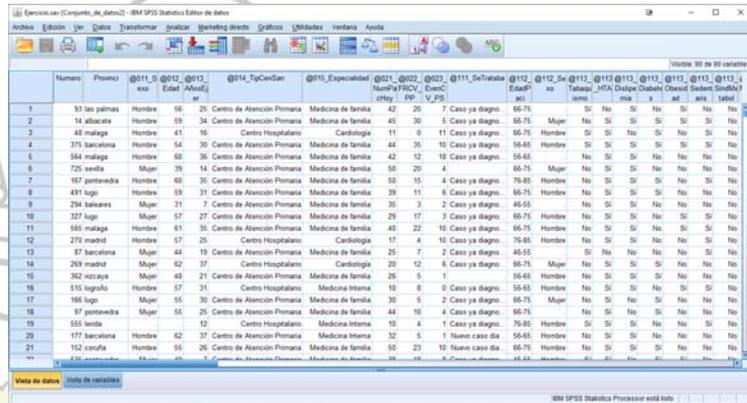
1. CARACTERÍSTICAS BÁSICAS Y ESTRUCTURA DE TRABAJO EN SPSS

1.1. VENTANAS DE SPSS

Existen varios tipos de ventanas en SPSS, las más importantes son:

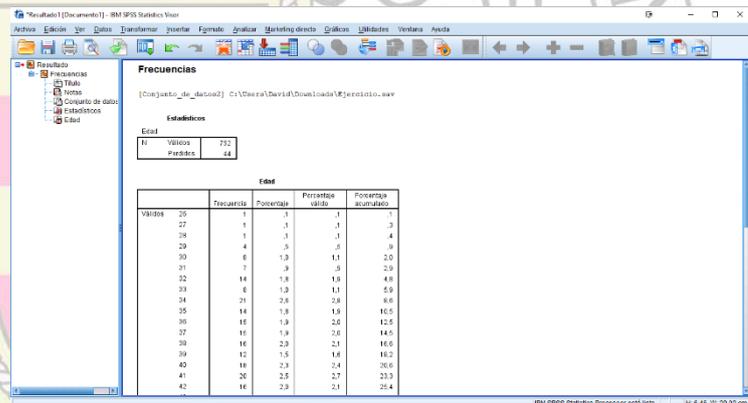
1.1.1. EDITOR DE DATOS:

Presenta el contenido del fichero de datos. Se puede editar un fichero de datos existente (agregar, modificar o borrar datos) o crear un nuevo fichero a través de esta ventana. El Editor de datos se abre automáticamente cuando se inicia una sesión de SPSS. Sólo puede haber un archivo de datos abierto al mismo tiempo.



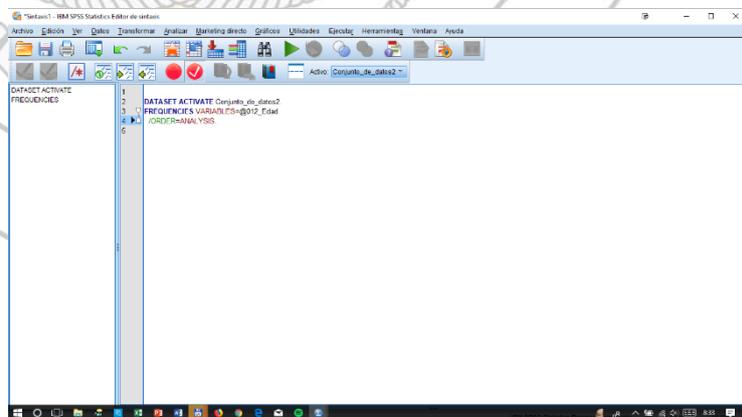
1.1.2. VISOR DE RESULTADOS:

Presenta la salida de las tablas, gráficos y resultados estadísticos realizados. Esta ventana se abre automáticamente la primera vez que se realiza algún procedimiento que genera resultados. Puede haber varias ventanas de resultados abiertas a la vez.



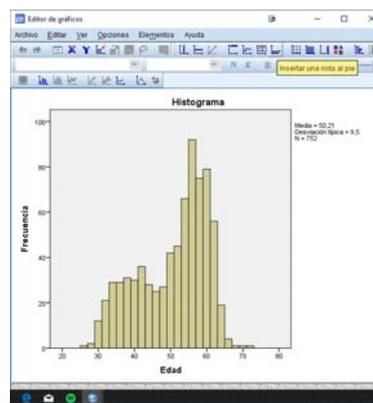
1.1.3. EDITOR DE SINTAXIS:

Cualquier instrucción ejecutada con SPSS mediante los menús, la aplicación de Windows la ejecuta, traduciéndola previamente a instrucciones del lenguaje de programación de SPSS. Estas instrucciones podemos crearlas, modificarlas o borrarlas mediante el Editor de sintaxis si conocemos el lenguaje de SPSS.



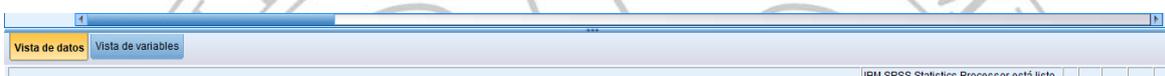
1.1.4. EDITOR DE GRÁFICOS

Permite realizar modificaciones de todo tipo en los gráficos (colores, fuentes, tamaños, intercambiar ejes, etc.).



1.2. BARRA DE ESTADO

Ocupa la línea mas inferior de la pantalla, informa del estado de la aplicación. En ella podemos diferenciar cuatro áreas diferentes de interés: El **área de Estado del Procesador** indica el estado del programa (preparado, no disponible). El **área de Estado de comando** indica el número de casos procesados. El **área de Estado de filtro** informa de la presencia de filtro (selección de casos). El **área de Estado de ponderación** informa del uso de una variable de ponderación para ponderar los casos para el análisis.



1.3. VENTANA DESIGNADA Y VENTANA ACTIVA

Ventana activa: Funciona como cualquier otra ventana en el entorno Windows. Sólo hay una ventana activa, y es la que está coloreada de azul. Cuando hay abierta ya una ventana de Resultados o de Sintaxis, los nuevos resultados o comandos de sintaxis se agregarán al final en la Ventana Designada.

Ventana Designada: se indica con un icono en la barra de estado.

Para señalar una ventana como designada, primero hay que activarla, procedimiento que podremos realizar de dos maneras:

- ✓ Pulsar el botón con un signo de admiración  en la ventana que se quiere señalar como designada.
- ✓ Elegir en el menú **Utilidades** y después **Designar ventana**.

1.4. BARRAS DE HERRAMIENTAS

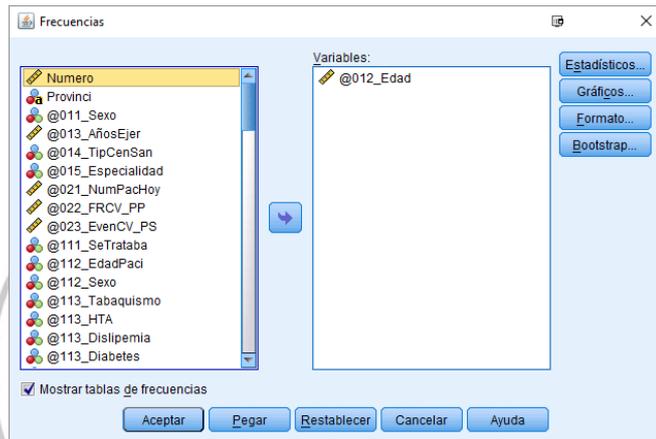
Cada tipo de ventana de SPSS tiene sus propias barras de herramientas específicas, cuya presentación puede controlarse eligiendo en el menú **Ver** y después **Barras de herramientas**. Las barras de herramientas pueden mostrarse u ocultarse, así como situarse en cualquier lugar de la pantalla (incluso fuera de la ventana). También pueden personalizarse.



1.5. CUADROS DE DIÁLOGO

La mayoría de las opciones de los menús acceden a un cuadro de diálogo al ser seleccionadas. Cada cuadro de diálogo tiene, de forma general, unos componentes básicos:

- ✓ **Lista de variables de origen:** son las variables disponibles para realizar el procedimiento seleccionado. Al poner el puntero sobre una variable, se visualiza completamente su etiqueta, y entre corchetes, su nombre. Si hemos definido una variable con características que no permiten su uso en un determinado procedimiento (variables texto en procedimientos de variables numéricas), esta no aparecerá en el listado.
- ✓ **Lista de variables de destino:** una o varias listas con las variables que hemos seleccionado para el análisis.
- ✓ **Botones de comandos:** Permiten seleccionar las instrucciones de los análisis que queremos aplicar a las variables para realizar determinadas acciones:
 - **Aceptar:** ejecuta el procedimiento.
 - **Pegar:** pega en la ventana de sintaxis los comandos necesarios para realizar en análisis en función de lo que hayamos seleccionado en el cuadro de diálogo. Si conocemos la sintaxis del lenguaje SPSS, su modificación permite personalizar la ejecución de los comandos.
 - **Restablecer:** deshace cualquier selección previa realizada.
 - **Cancelar:** cancela los cambios realizados desde la última vez que se abrió el cuadro de diálogo, y lo cierra.
 - **Ayuda:** conduce a una ventana de ayuda con información sobre el cuadro de diálogo.



1.5.1. SUBCUADROS DE DIÁLOGO

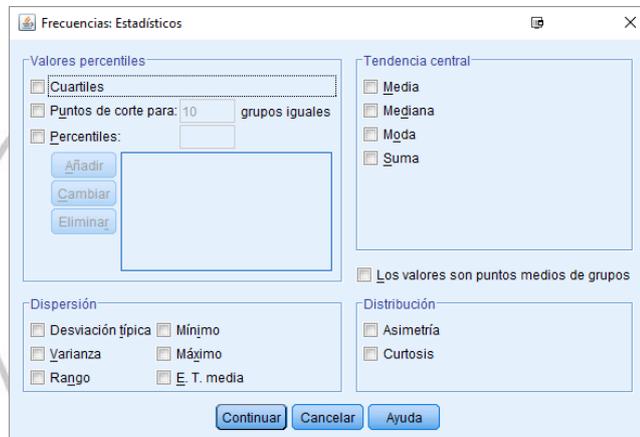
Son cuadros de diálogo que se despliegan desde el cuadro de diálogo principal. Están indicados con puntos suspensivos en el botón correspondiente.

Para **seleccionar una variable**, la pincharemos con el cursor del ratón y su color cambiará; posteriormente pulsaremos el botón que tiene una flecha y que está entre las dos listas para que sea trasladada a la lista de variables seleccionadas. Para **seleccionar múltiples variables contiguas** pincharemos la primera, y manteniendo pulsada la tecla de mayúsculas pincharemos la última, quedando seleccionadas todas las variables relacionadas entre ambas. Si deseamos **seleccionar simultáneamente variables no contiguas**, pulsaremos la tecla *Ctrl* y sin soltar iremos pulsando para seleccionar las diferentes variables que deseamos incluir en el análisis. En estos cuadros se puede obtener información sobre las variables: seleccionar la

variable, pulsar en cualquier parte de la lista con el botón derecho del ratón y pulsar en *Información de las variables* en el menú contextual.

1.5.2. EJECUCIÓN DE UN PROCEDIMIENTO

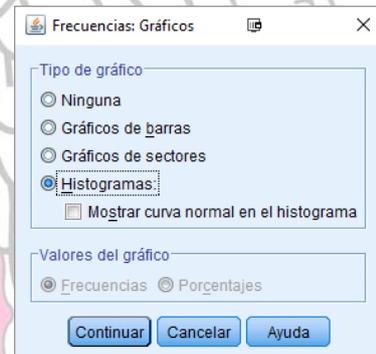
Dada la gran flexibilidad y número de posibles pruebas estadísticas disponibles en SPSS para Windows, del total de posibles métodos debemos indicar cual o cuales de ellos deseamos realizar. Por ejemplo, pulsando el botón Estadísticos que aparece en la caja de diálogo del procedimiento Frecuencias, podremos seleccionar las pruebas estadísticas que deseemos:



Cuando las opciones presentan un "tick" para su selección pueden seleccionarse una o todas de cada recuadro.

En algunos procesos de selección sólo es posible seleccionar una opción de entre varias disponibles, al ser estas mutuamente excluyentes. En este caso siempre hay una selección por defecto realizada. Si la cambiamos, se deshará la selección por defecto y se marcará la opción seleccionada por nosotros.

A veces SPSS ofrece una serie de posibilidades que habremos de seleccionar desplegando el menú previamente. Para ello pulsaremos la flecha existente a su derecha y tras su despliegue realizaremos la selección.



1.6. ANÁLISIS DE DATOS CON SPSS

Hay que seguir unos pasos básicos:

- 1) Seleccionar un fichero de datos. Éste puede crearse introduciendo los datos a través de la pantalla del Editor de datos, o bien abriendo un archivo previamente existente. En ocasiones lo importaremos desde otras aplicaciones.
- 2) Seleccionar el procedimiento deseado en el menú correspondiente.
- 3) Seleccionar la o las variables que queremos incluir en el estudio, así como las características que queramos seleccionar del método estadístico.
- 4) Ejecutar el procedimiento y observar los resultados.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	ID	Cadena	6	0	Número de caso	Ninguna	Ninguna	6	Izquierda	Nominal	Entrada
2	Edad	Numérico	6	2	Edad del paciente	Ninguna	Ninguna	6	Derecha	Desconocido	Entrada
3	Sexo	Numérico	6	2	Sexo del paciente	{1,00, Hombre}...	Ninguna	6	Derecha	Desconocido	Entrada
4	VAR00001	Numérico	6	2		Ninguna	Ninguna	6	Derecha	Desconocido	Entrada

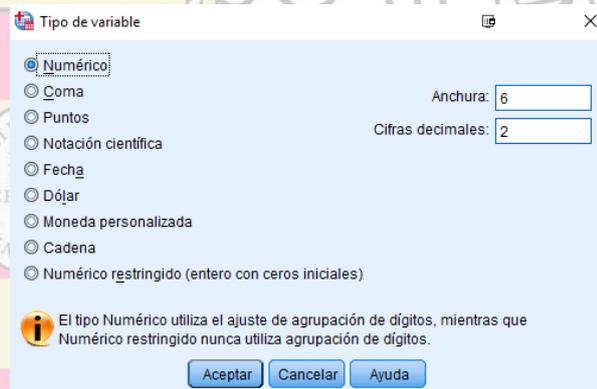
Los siguientes atributos de una variable pueden copiarse de unas a otras, sin necesidad de tener que reintroducirlas de nuevo manualmente para cada variable. Para ello, hay que seleccionar los atributos que queremos copiar, pinchar el menú Edición, y elegir Copiar (o pulsar Control + C), y después situar el puntero en la variable de destino, pinchar Edición, y elegir Pegar (Control + V). Si las variables son contiguas, también se pueden “arrastrar” las características con el puntero.

2.2.1. NOMBRE:

Introduciremos el nombre de la variable. Si no damos nombre a la variable, SPSS le dará uno por defecto (VAR00001, VAR00002, ...). El nombre debe comenzar necesariamente por una letra, el resto de los caracteres pueden ser letras, dígitos, puntos o los caracteres @, #, _ ó \$, y no puede terminar en un punto. No pueden usarse espacios en blanco ni caracteres especiales (ñ, º, ´ ó *). El nombre “corto” de la variable puede tener un máximo de ocho caracteres, y no distingue mayúsculas de minúsculas.

2.2.2. TIPO DE VARIABLE

Al pulsar, se abre el subcuadro de diálogo en el que especificaremos los tipos de datos para cada variable. Por defecto, SPSS asume que todas las variables nuevas son numéricas. Los principales tipos de variables son:



- ✓ **Numérico:** sólo contiene datos de carácter numérico (tiene sentido sumarlos, restarlos, es decir operar con ellos; o compararlos), también con decimales (variables numéricas continuas y discretas).
- ✓ **Coma:** son variables numéricas, usadas habitualmente para datos de contabilidad, en las que los miles son separados con comas y los decimales con un punto.
- ✓ **Punto:** como la anterior, pero separa los miles con puntos y los decimales con una coma.
- ✓ **Notación científica:** es una variable numérica que usa algunos caracteres especiales, como por ejemplo el número *e* elevado a *n* (exponente en base 10).
- ✓ **Fecha:** la fecha puede expresarse de diferentes formas, que aparecen automáticamente al elegir este tipo de variable. Cuidado con estas variables cuando importamos tablas de datos desde Excell. Según estén definidas en ese programa las importará bien o mal.
- ✓ **Dólar, Moneda personalizada:** para expresar cantidades en dólares u otras unidades monetarias.
- ✓ **Cadena:** variables alfanuméricas. Hay que tener en cuenta que los espacios en blanco también cuentan como otros caracteres.

Como norma general todas las variables las definiremos como numéricas (Cuantitativas continuas y discretas; y las cualitativas categóricas) salvo las que sean de tipo alfanuméricas (o texto) que las definiremos como Cadena. Solo definiremos como Cadena las

que no se puedan definir como Numérica por contener caracteres no numéricos. También definiremos como Cadenas aquellas que al estar representadas por número carecen de contenido cuantitativo como es el caso de las variables número de DNI, número de la SS, o número de teléfono, es decir aquellas cadenas de números que no tiene sentido operar con ellas o comparar las cantidades.

2.2.3. ANCHURA:

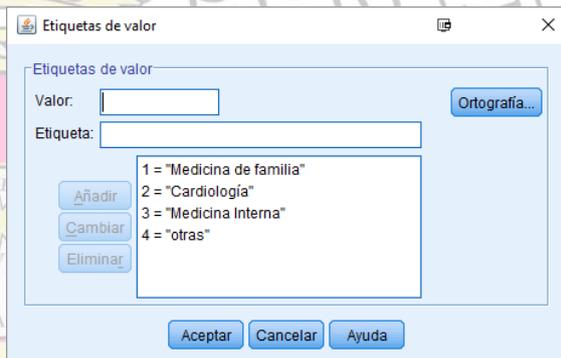
Es la anchura o número de caracteres que se pueden introducir y se presentarán en la pantalla en el hueco generado para las variables. La notación de cimal ocupa un espacio y redondeará la cantidad para que aparezca con el formato seleccionado.

2.2.4. ETIQUETA

Permite crear una etiqueta (Un segundo nombre largo y completo) que describe la variable. Las etiquetas de variable pueden tener hasta 256 caracteres (a diferencia del nombre de la variable, que sólo puede tener 8). Son útiles para describir de forma más detallada la variable; al definir la etiqueta, ésta aparecerá en los resultados, y sustituirá al nombre de la variable.

2.2.5. ETIQUETAS DE VALOR:

Las etiquetas de valor sirven para asignar etiquetas descriptivas a cada valor de la variable. Esto es muy interesante para representar variables categóricas como numéricas, asignando un valor numérico a cada categoría. Para asignar las etiquetas, se escribe el valor deseado en el campo **Valor**, el nombre de la etiqueta en el campo **Etiqueta de valor**, y se pulsa el botón **Añadir**.



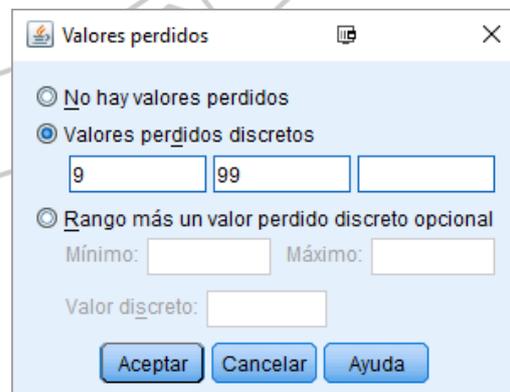
También puede hacerse pulsando directamente el botón de la barra de herramientas. Cuando definimos etiquetas de valor, éstas pueden visualizarse directamente en la **Vista de datos del Editor**, en lugar de los valores numéricos. Para ello hay que ir al menú **Ver** y luego a **Etiquetas de valor**.

2.2.6. VALORES PERDIDOS:

Esta opción permite definir qué valores de la variable deben ser considerados por el programa como valores perdidos (desconocidos o *missing data*; p.e. porque no tengamos el resultado en esa variable para ese caso). Estos valores que darán excluidos de la mayoría de los cálculos estadísticos (es decir no se cuenta como caso, solo cuando intervenga esa variable que tiene casos perdidos).

Pueden definirse hasta 3 valores perdidos

discretos, o bien intervalos entre los que se considera los valores como perdidos para variables



numéricas.

2.2.7.FORMATO DE COLUMNA

Con esta opción se puede seleccionar el ancho de la columna que visualizaremos en la Vista de datos; esta anchura es distinta de la anterior y sólo se refiere a lo que veremos en pantalla. En el caso de variables numéricas con decimales hay que tener en cuenta que el punto decimal se considera como otro carácter adicional, es decir un Ácido urico de 5,7 requiere tres huecos).

2.2.8.ALINEACIÓN

Define como queremos que se alineen los datos en las columnas cada casilla.

2.2.9.MEDIDA:

Hace referencia a la clase de variable de que se trata. Si es de tipo Escala, Ordinal o Nominal. Los atributos de una variable pueden copiarse a otras, sin necesidad de tener que introducirlos manualmente. Para ello, hay que seleccionar los atributos que queremos copiar, pinchar el menú Edición, y elegir Copiar (o pulsar Control + C), y después situar el puntero en la variable de destino, pinchar Edición, y elegir Pegar (Control + V). Si las variables son contiguas, también se pueden “arrastrar” las características con el puntero.

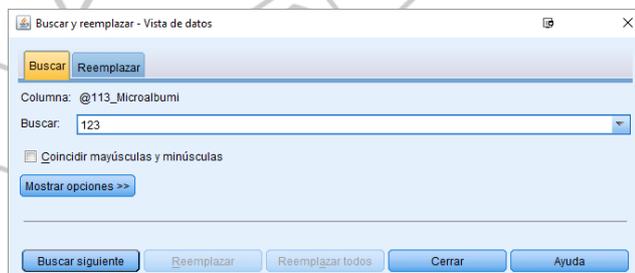
2.3. INTRODUCCIÓN DE DATOS

Para introducir datos en una casilla sólo hay que posicionar el cursor sobre la misma y pulsar con el ratón; la casilla se resaltará. En la esquina superior izquierda del Editor de datos aparecerán el nombre de la variable y el número de fila (número de caso) sobre el que estamos trabajando. El valor introducido queda registrado una vez que se pulsa Intro o que nos situamos en otra casilla.

Es posible cortar, copiar y pegar valores de casillas individuales o de grupos de valores mediante los correspondientes comandos en el menú **Edición**. También pueden insertarse nuevos casos o variables entre los ya existentes mediante los menús **Insertar caso** e **Insertar variable** del menú **Datos**.

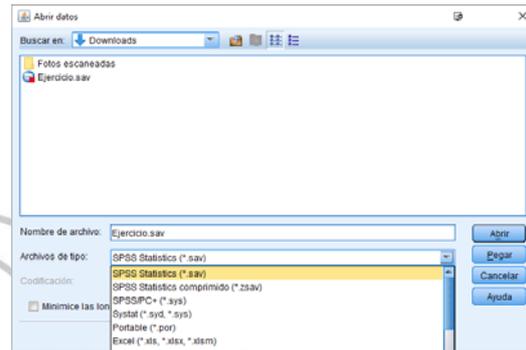
Para ir a un caso o fila concreta puede hacerse mediante búsqueda directa con el ratón o bien con la opción **Ir a caso** del menú **Datos**. Al introducir el número de caso al que queremos ir y pulsar **Aceptar**, el programa nos desplazará a dicho caso.

También pueden buscarse datos concretos en una variable de forma automática. Primero hay que situarse en una casilla de la variable en la que deseamos buscar, y después seleccionar el menú **Edición y Buscar**. Hay que introducir el valor que se quiere hallar y después pulsar **Buscar siguiente**.



2.4. IMPORTACIÓN Y LECTURA DE ARCHIVOS

Los archivos de datos en SPSS tienen siempre la extensión **.sav**, por lo que para abrir un archivo de datos existente hay que ir al menú **Archivo** y elegir **Abrir y Datos**. Entonces se abre un cuadro de diálogo donde podemos buscar la ruta donde se encuentra el fichero que queremos abrir, lo seleccionamos y pulsamos **Abrir**.



También pueden abrirse e importarse en SPSS archivos de otros programas. Para ello, en el mismo cuadro de diálogo anterior, hay que pinchar en la flecha de la casilla Tipo de archivos, desplegándose un menú contextual donde aparecen los tipos de archivos que puede abrir SPSS.

Una vez que abrimos un archivo ya existente, podemos estudiar qué variables tiene y en qué formato mediante la opción Información del archivo del menú Utilidades. Se abrirá una ventana de resultados con la caracterización de todas las variables.



3. FUSIÓN DE ARCHIVOS

En determinadas ocasiones necesitamos trabajar con más de un fichero de datos, y la información se encuentra repartida en varias tablas, por lo que es necesario juntar o fusionar dos o mas tablas en una nueva que contenga el resultado de las tablas anteriores. Podemos fusionar variables o bien fusionar casos.

3.1. AÑADIR VARIABLES

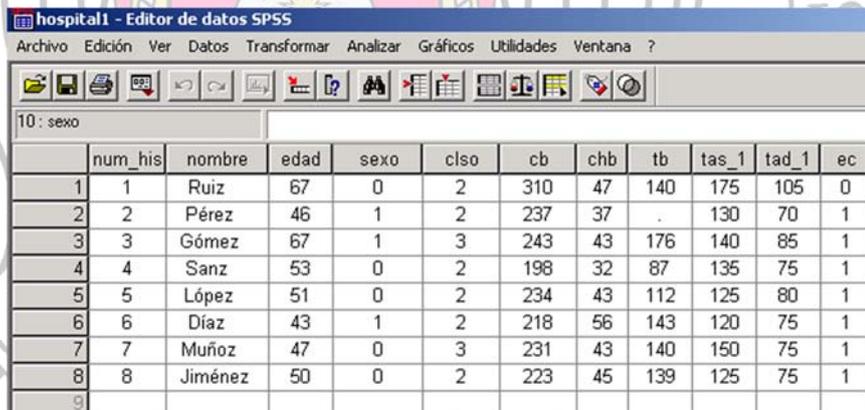
En este caso se trata de añadir nuevas características (nuevas variables, o columnas) a un archivo que ya contiene otras variables de los mismos individuos. Para ello, los archivos que vamos a fusionar tienen que cumplir unos requisitos:

- ✓ Los dos archivos deben contener una variable común a ambos ficheros que identifique cada caso en ambos ficheros (p.e. número de caso), con un valor único para cada individuo, o sea, debe estar presente en ambos ficheros, y su valor debe ser igual en ambos ficheros para cada caso.
- ✓ Los dos archivos deben estar ordenados de menor a mayor por la variable que identifica cada caso.

Puede suceder que los dos archivos tengan exactamente los mismos casos, o bien que sólo coincidan algunos. En este último supuesto, podemos hacer que los casos del archivo resultante sean la suma de los casos de los dos archivos, o que uno de los dos archivos “domine” al otro y sean sólo sus casos los que integren el archivo resultante.

Veámoslo paso a paso con un ejemplo.

Supongamos el archivo “hospital1”, con datos sobre pacientes, en los que se incluyen, entre otras, dos variables con sendas tomas de presión arterial sistólica (tas) y diastólica (tad), con



	num_his	nombre	edad	sexo	clso	cb	chb	tb	tas_1	tad_1	ec
1	1	Ruiz	67	0	2	310	47	140	175	105	0
2	2	Pérez	46	1	2	237	37	130	70	70	1
3	3	Gómez	67	1	3	243	43	176	140	85	1
4	4	Sanz	53	0	2	198	32	87	135	75	1
5	5	López	51	0	2	234	43	112	125	80	1
6	6	Díaz	43	1	2	218	56	143	120	75	1
7	7	Muñoz	47	0	3	231	43	140	150	75	1
8	8	Jiménez	50	0	2	223	45	139	125	75	1
9											

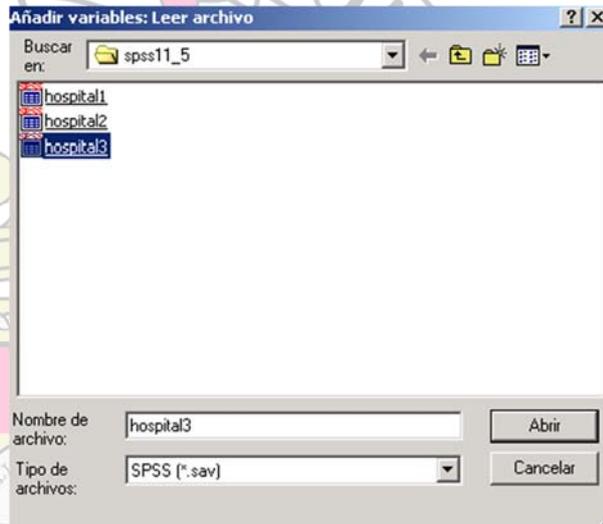
la idea de repetir esas tomas en el tiempo y observar la evolución de la presión arterial. En este archivo, además de estos datos hay una variable que incluye un código de identificación o número de historia clínica, distinto para cada paciente y que lo identifica unívocamente.

Además, tenemos un segundo archivo “hospital2”, en el que se ha recogido una nueva toma de tas y tad, y queremos añadir estos datos al archivo “hospital1”.

Primero, hemos de tener abierto el archivo sobre el que vamos a añadir las variables, es decir, “hospital1”. A continuación, vamos al menú Datos y elegimos Fundir Archivos y Añadir variables. Aparecerá un cuadro donde hemos de seleccionar el archivo que queremos añadir.

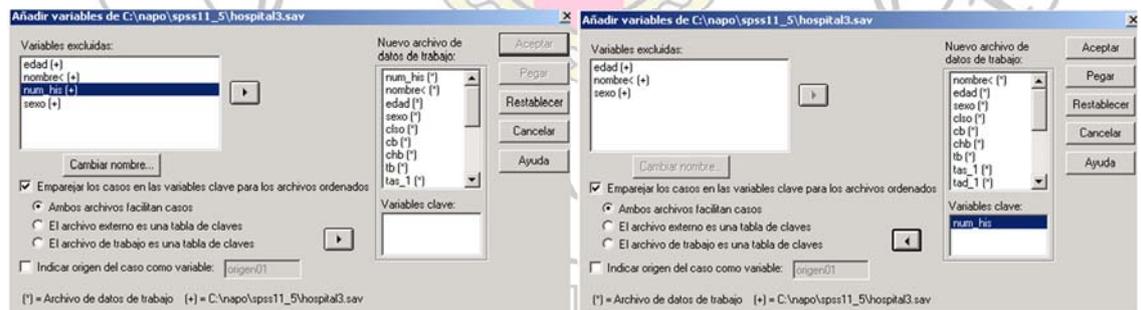
	num_his	nombre	edad	sexo	tas_2	tad_2	var
1	1	Ruiz	67	0	175	110	
2	2	Pérez	46	1	135	80	
3	3	Gómez	67	1	130	80	
4	4	Sanz	53	0	135	80	
5	6	Díaz	43	1	110	70	
6	7	Muñoz	47	0	160	100	
7	8	Jiménez	50	0	130	80	
8	9	Galíndez	58	1	145	95	
9	10	Sanz	51	0	120	80	
10							

Una vez elegido el archivo, se abre el cuadro de diálogo para realizar la fusión.



En el cuadro **Nuevo archivo de datos de trabajo**, SPSS elige por defecto todas las variables del llamado “archivo de datos de trabajo” (en nuestro caso, “hospital1”), más las variables del archivo externo (“hospital2”) que no están repetidas. En el cuadro **Variables excluidas** quedan las variables del archivo externo cuyo nombre es idéntico a otras variables del archivo de datos de trabajo. Si no hiciésemos nada más y seccionásemos **Aceptar**, los datos del

archivo externo se sobrescribirían en el archivo de datos de trabajo sin tener en cuenta si el individuo es el mismo. Para que esto suceda, hemos de seleccionar la opción **Emparejar los casos en las variables clave para los archivos ordenados**, seleccionar la variable de identificación de caso en el cuadro **Variables excluidas** y transportarla al cuadro **Variables clave**.



Aún queda por definir qué casos formarán parte del archivo resultante. Si elegimos la opción **Ambos archivos facilitan casos**, los casos serán la suma de los de ambos archivos.

	num_his	nombre	edad	sexo	clso	cb	chb	tb	tas_1	tad_1	ec	fuma	sede	peso	talla	nes	acf	tas_2	tad_2
1	1	Ruiz	67	0	2	310	47	140	175	105	0	0	0	76	1.77	2	1	175	110
2	2	Pérez	46	1	2	237	37	.	130	70	1	1	0	56	1.60	3	1	135	80
3	3	Gómez	67	1	3	243	43	176	140	85	1	0	1	65	1.65	1	0	130	80
4	4	Sanz	53	0	2	198	32	87	135	75	1	0	0	76	1.68	2	1	135	80
5	5	López	51	0	2	234	43	112	125	80	1	0	1	76	1.68	3	1	.	.
6	6	Díaz	43	1	2	218	56	143	120	75	1	0	1	54	1.65	3	1	110	70
7	7	Muñoz	47	0	3	231	43	140	150	75	1	0	0	67	1.60	1	0	160	100
8	8	Jiménez	50	0	2	223	45	139	125	75	1	1	1	68	1.72	.	1	130	80
9	9	Galíndez	58	1	145	95
10	10	Sanz	51	0	120	80

Con esta opción, además puede crearse una variable que nos dice de qué archivo procede cada caso, si del externo o del de trabajo; para ello hay que marcar la casilla donde pone **Indicar origen del caso como variable**.

Si elegimos la opción *El archivo externo es una tabla de claves*, el archivo origen “dominará” al archivo externo, y sólo se incluirán los casos existentes en el archivo origen, eso sí, añadiendo las variables nuevas del archivo externo.

	num_his	nombre	edad	sexo	clso	cb	chb	tb	tas_1	tad_1	ec	fuma	sede	peso	talla	nes	acf	tas_2	tad_2
1	1	Ruiz	67	0	2	310	47	140	175	105	0	0	0	76	1.77	2	1	175	110
2	2	Pérez	46	1	2	237	37	.	130	70	1	1	0	56	1.60	3	1	135	80
3	3	Gómez	67	1	3	243	43	176	140	85	1	0	1	65	1.65	1	0	130	80
4	4	Sanz	53	0	2	198	32	87	135	75	1	0	0	76	1.68	2	1	135	80
5	5	López	51	0	2	234	43	112	125	80	1	0	1	76	1.68	3	1	.	.
6	6	Díaz	43	1	2	218	56	143	120	75	1	0	1	54	1.65	3	1	110	70
7	7	Muñoz	47	0	3	231	43	140	150	75	1	0	0	67	1.60	1	0	160	100
8	8	Jiménez	50	0	2	223	45	139	125	75	1	1	1	68	1.72	.	1	130	80
9																			
10																			

Por el contrario, si escogemos la opción *El archivo de trabajo es una tabla de claves*, será el archivo externo el que domine al archivo de trabajo, y en el archivo resultante con todas las variables sólo estarán los casos que aporte el archivo externo.

Puede suceder que alguna de las variables excluidas se llame igual que alguna de las variables del archivo de trabajo, aunque su contenido sea diferente. En esa situación, puede cambiarse al nombre de la variable pulsando el botón **Cambiar nombre**, y automáticamente la variable pasará al cuadro de **Nuevo archivo de datos de trabajo**.

	num_his	nombre	eda	sexo	clso	cb	chb	tb	tas_1	tad_1	ec	fuma	sede	peso	talla	nes	acf	tas_2	tad_2
1	1	Ruiz	67	0	2	310	47	140	175	105	0	0	0	76	1.77	2	1	175	110
2	2	Pérez	46	1	2	237	37	.	130	70	1	1	0	56	1.60	3	1	135	80
3	3	Gómez	67	1	3	243	43	176	140	85	1	0	1	65	1.65	1	0	130	80
4	4	Sanz	53	0	2	198	32	87	135	75	1	0	0	76	1.68	2	1	135	80
5	6	Díaz	43	1	2	218	56	143	120	75	1	0	1	54	1.65	3	1	110	70
6	7	Muñoz	47	0	3	231	43	140	150	75	1	0	0	67	1.60	1	0	160	100
7	8	Jiménez	50	0	2	223	45	139	125	75	1	1	1	68	1.72	.	1	130	80
8	9	Galindez	58	1	145	95
9	10	Sanz	51	0	120	80
10

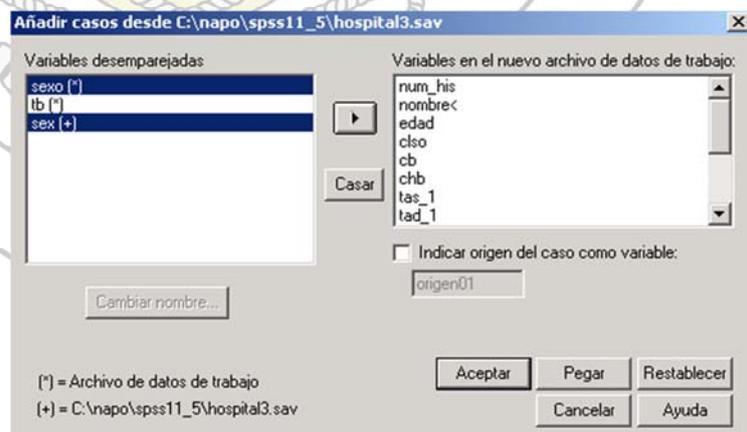
3.2. AÑADIR CASOS

Para añadir casos deberíamos tener los dos archivos con las mismas variables, con los mismos nombres, y con diferentes individuos. La operación se realiza pinchando en el menú **Datos** y eligiendo **Fundir archivos** y **Añadir casos**. Como en el apartado anterior, se abre el cuadro en el que hay que seleccionar el archivo del que queremos añadir los casos.



Entonces aparece el cuadro de diálogo para efectuar la fusión.

En el cuadro **Variables en el nuevo archivo de datos de trabajo** aparecen las variables que se encuentran con idéntico nombre en los dos archivos, que SPSS reconoce como comunes a ambos. En el cuadro **Variables desparejadas** están aquellas que sólo aparecen en uno de los dos archivos (un asterisco o una cruz indican en cuál).



Si queremos que una variable desparejada forme parte del archivo resultante, puede seleccionarse y pasarse al cuadro de la derecha, aunque lógicamente sólo contendrá datos de los casos del archivo del que procede.

Conviene examinar bien las variables desemparejadas, porque puede ocurrir que una variable tenga nombres distintos en los dos archivos que queremos fusionar, pero en realidad contenga la misma información. Si eso pasa, puede identificarse una variable con la otra, seleccionando las dos, y pulsando el botón **Casar**; en ese momento, desaparecerán del cuadro de **Variables desemparejadas** y se incluirán en el cuadro **Variables en el nuevo archivo de datos de trabajo**. Otra opción es cambiar el nombre de una de las variables con el botón **Cambiar nombre**, y ponerlo igual al de la otra, para que el programa pueda identificarlas como la misma.

Al igual que en **Añadir variables**, puede crearse una nueva variable que nos indique de qué archivo procede cada caso, seleccionando el botón **Indicar origen del caso como variable**.

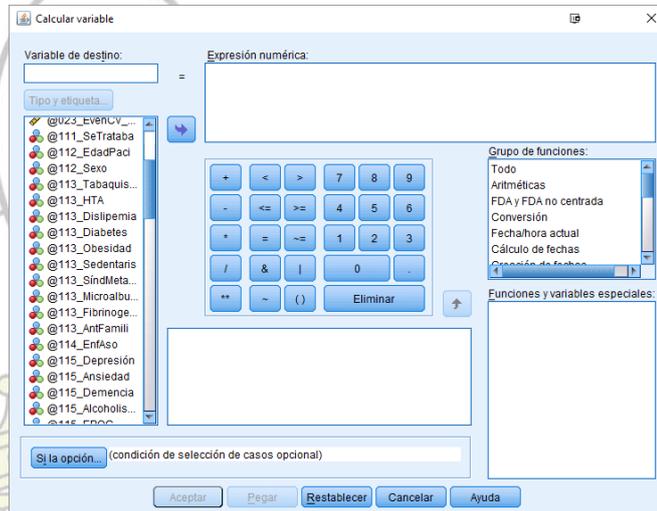


4. TRANSFORMACIÓN DE DATOS

En ocasiones es necesario crear nuevas variables a partir de otras ya existentes o modificarlas para poder explotarla de forma más adecuada. Esto se realiza mediante las opciones de transformación de SPSS.

4.1. CALCULAR VARIABLE

Esta función crea una nueva variable basándose en transformaciones a partir de otras variables. Para acceder a esta utilidad hay que seleccionar los menús **Transformar** y **Calcular**. Entonces aparece el cuadro de diálogo **Calcular variable**.



En primer lugar, hay que escribir el nombre de la variable de destino; ésta puede ser una variable ya existente o bien una variable que se vaya a añadir al archivo. Una vez escrito también pueden definirse sus atributos pulsando el botón **Tipo y etiqueta**. Después hay que construir la expresión que nos dará los valores de las variables. Pueden incluirse constantes, variables definidas previamente, operadores de la calculadora o funciones.

OPERADORES DE LA CALCULADORA: suma (+), resta (-), producto (*), cociente (/), potencia (**), igualdad (=), desigualdades (<, >, <=, >=, ~ =), paréntesis, y operadores lógicos AND (&), OR (|) y NOR (~).

FUNCIONES: existe una amplia variedad de funciones útiles para realizar transformaciones sobre variables, y algunas de las más interesantes son las siguientes:

FUNCIONES ARITMÉTICAS:

- ✓ **ABS:** calcula el valor absoluto de un número.
- ✓ **RND:** redondea una cifra por aproximación al número entero más cercano.
- ✓ **TRUNC:** elimina la parte decimal de un número.

OTRAS FUNCIONES MATEMÁTICAS: SIN, COS, EXP, LG10, LN, SQRT, etc.

FUNCIONES ESTADÍSTICAS: SUM, MEAN, SD, VAR, MIN, MAX.

FUNCIONES LÓGICAS:

- ✓ **ANY:** la función requiere 2 argumentos: una variable (prueba), y un valor o lista de valores. Crea una variable nueva que asigna el valor 1 (verdadero) si el valor de la variable prueba coincide con alguno de los valores de una lista, y el valor 0 (falso) si no coincide.
- ✓ **RANGE:** es como la anterior, pero en vez de un valor o lista de valores, hay que introducir un rango (valor mínimo y valor máximo) o rangos. Asigna el valor 1 (verdadero) si el valor de la variable prueba está incluido en el rango introducido, y el valor 0 (falso) si no está

incluido.

FUNCIONES DE GENERACIÓN DE VARIABLES ALEATORIAS:

Para generar números aleatorios (pseudo-aleatorios) puede establecerse la semilla u origen empleado por el programa, de forma que se pueda reproducir una secuencia de números pseudo-aleatorios. Para ello debe seleccionarse el menú **Transformar** y después **Semilla de aleatorización**.

- ✓ **RV.NORMAL:** crea una variable con valores aleatorios de una distribución normal estandarizada (con media y desviación típica la que indiquemos).
- ✓ **RV.UNIFORM:** genera en la variable números aleatorios en el rango que se indique.

FUNCIONES PARA EL TRATAMIENTO DE FECHAS:

- ✓ **CTIME.DAYS:** sirve para calcular tiempos transcurridos en días. Toma como referencia una fecha estable, la de instauración del calendario gregoriano (15 de octubre de 1582).
- ✓ **DATE.DMY:** reconoce una fecha y la expresa en segundos transcurridos desde la fecha de referencia.

FUNCIONES CON VARIABLES DE CADENA:

- ✓ **CONCAT:** permite añadir los caracteres de varias variables de cadena. Antes de crearla hay que definir en el cuadro de diálogo los atributos de la nueva variable.
- ✓ **LENGTH:** devuelve la longitud (número de caracteres) de una variable de cadena, incluyendo todos los espacios en blanco.
- ✓ **LPAD:** permite completar por la izquierda una variable de cadena con espacios en blanco o con los caracteres que se le indique.
- ✓ **LTRIM:** devuelve una variable de cadena eliminando los espacios en blanco que preceden a la misma.
- ✓ **RTRIM:** devuelve una variable de cadena eliminando los espacios en blanco posteriores a la misma.

El orden en que se procesa una expresión es el siguiente:

- ✓ Expresiones entre paréntesis.
- ✓ Funciones.
- ✓ Operadores aritméticos, con la siguiente prioridad:
 - Potencia.
 - Producto y cociente.
 - Suma y resta.

Además, a la hora de crear la variable puede hacerse sólo en un grupo de casos de la base de datos, mediante el botón **Sí**. Al pulsarlo se abre el cuadro de diálogo **Calcular variable: Si los casos**.

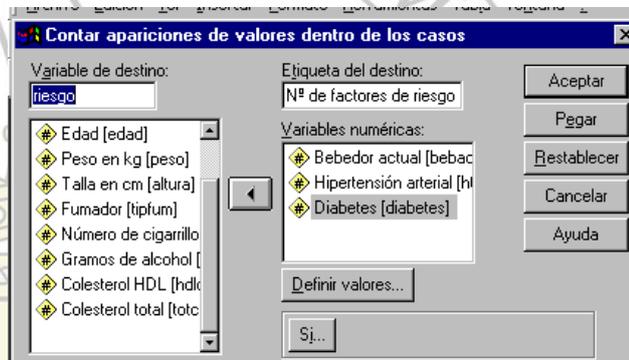


En este cuadro puede seleccionarse un conjunto de casos incluyendo una o varias condiciones sobre las variables utilizadas para la transformación. Si no se establece ninguna condición, por defecto la transformación incluirá todos los casos del archivo.

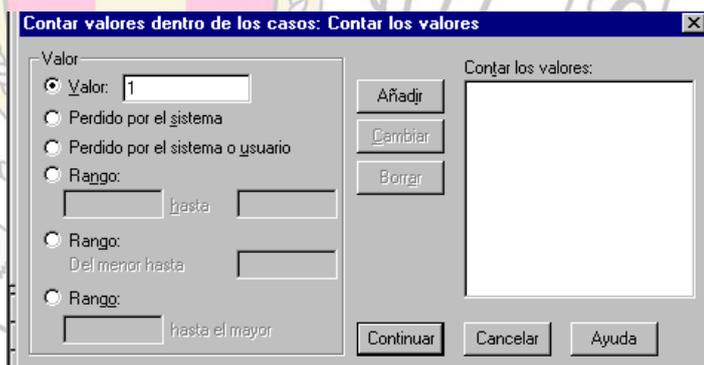
4.2. CONTAR

APARICIONES DENTRO DE LOS CASOS

Con esta función se crea una variable que permite contar el número de veces que aparece un valor en una lista de variables. Para ello es conveniente que las variables de esa lista estén expresadas o codificadas en los mismos valores, por ejemplo, 0 (No) y 1 (Sí). Para realizar esta operación hay que seleccionar el menú **Transformar** y después **Contar apariciones**. Se abrirá el cuadro de diálogo **Contar apariciones de valores dentro de los casos**.



Hay que dar nombre a la nueva variable que se va a crear, y seleccionar las variables en las que se quiere contar el número de apariciones. A continuación, se pulsa el botón **Definir valores**. En el cuadro de diálogo **Contar valores dentro de los casos: Contar los valores** se especifican los valores que se desean contar, ya sean valores discretos o rangos de valores, pulsando el botón **Añadir** cada vez, y al finalizar pulsando **Continuar**. Además, al igual que cuando calculábamos una nueva variable, pueden imponerse condiciones para realizar esta transformación sólo en un subgrupo de casos, pulsando el botón **Si**.



4.3. RECODIFICAR VALORES

La recodificación de valores se utiliza para modificar los valores de una variable ya existente. Esto es bastante útil para agrupar o combinar categorías en las variables. La recodificación puede realizarse sobre la misma variable o creando una nueva. Es aconsejable esta segunda modalidad, ya que si recodificamos sobre la misma variable se perderán los datos originales,

siendo imposible luego su recuperación. Las variables recodificadas pueden ser numéricas o de cadena. Para acceder a esta función hay que seleccionar el menú **Transformar**, y luego situar el ratón sobre **Recodificar**; automáticamente se despliega un menú con las opciones **En las mismas variables**, y **En distintas variables**.

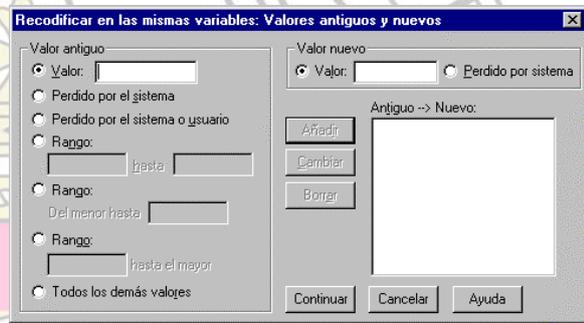
4.3.1. RECODIFICAR EN LAS MISMAS VARIABLES

Al elegir los menús citados se abre el cuadro de diálogo **Recodificar en las mismas variables**.



En primer lugar, se selecciona la variable que se quiere recodificar, y después hay que pulsar el botón **Valores antiguos y nuevos**, abriéndose el correspondiente cuadro de diálogo.

El valor antiguo es el valor o valores que se van a recodificar, y pueden ser valores discretos o rangos de valores, así como valores perdidos.



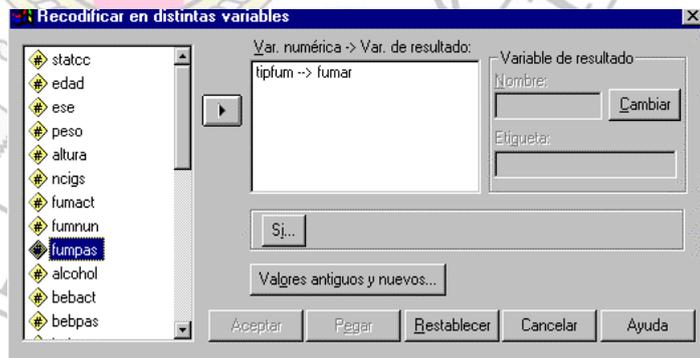
El valor nuevo es el valor individual en el que se recodifica cada valor o rango de valores antiguo.

En el recuadro **Antiguo -> Nuevo** se van reflejando todas las especificaciones para la recodificación cada vez que se pulsa el botón **Añadir**.

Al acabar hay que pulsar **Continuar**, y en el cuadro que queda, **Aceptar**. Automáticamente aparecerá la variable con los nuevos valores, desapareciendo los previos.

4.3.2. RECODIFICAR EN DISTINTAS VARIABLES

En este caso se realiza la misma operación, pero los nuevos valores aparecerán en una nueva variable, preservándose la variable de origen.



Aquí también hay que seleccionar primero la variable que se quiere recodificar, y además hay que dar nombre a la nueva variable que se crea con los nuevos valores.

Una vez hecho esto se pulsa el botón **Cambiar**, y la operación se verá reflejada en el correspondiente recuadro.

Después se pulsa **Valores antiguos y nuevos**, y como en el caso anterior se especifican todos los cambios para la recodificación, pulsando **Continuar** al terminar, y **Aceptar** después. En este caso aparecerá la nueva variable con los nuevos valores especificados para la recodificación.

4.4. ASIGNACIÓN DE RANGOS

Permite crear una nueva variable que contiene el valor de la posición que ocupan los datos cuando éstos son ordenados de mayor a menor o viceversa.

Hay que seleccionar el menú **Transformar**, y luego **Asignar rangos a los casos**, abriéndose el correspondiente cuadro de diálogo.

En él se seleccionan las variables para cuyos valores se quiere asignar los rangos. Existe la posibilidad de seleccionar diferentes métodos de asignación de rangos en el botón **Tipos de rango**. También puede seleccionarse el sistema para asignar los rangos a los valores empates, pulsando el botón **Empates**.



4.5. CATEGORIZACIÓN DE VARIABLES

Con esta función se convierten datos numéricos en varias categorías. Se crea una nueva variable. Los valores de la variable se categorizar según grupos de percentiles, de forma que cada grupo tiene aproximadamente el mismo número de casos. De esta forma el programa asignará en la nueva variable creada el valor 1 a cada valor del primer grupo, 2 a los valores del segundo grupo, y así sucesivamente. Para efectuar esto hay que escoger en el menú **Transformar**, y luego **Categorizar variables**, abriéndose el cuadro de diálogo. Hay que seleccionar en el cuadro la variable a cuyos datos se quiere categorizar.

4.6. RECODIFICACIÓN AUTOMÁTICA

Esta opción convierte los valores de una variable (ya sean numéricos o de cadena) en valores numéricos enteros consecutivos, creando una nueva variable con las mismas etiquetas y valores perdidos de la variable original. A esta función se llega eligiendo los menús **Transformar** y después **Recodificación automática**, apareciendo el correspondiente cuadro de diálogo. En él hay que seleccionar la variable que se desea recodificar y poner nombre a la nueva variable, pulsando después el botón **Nuevo nombre**. Además, puede elegirse si la recodificación comienza desde el valor inferior o superior. En el caso de las variables de cadena la recodificación comienza por orden alfabético, con las mayúsculas antes que las minúsculas.

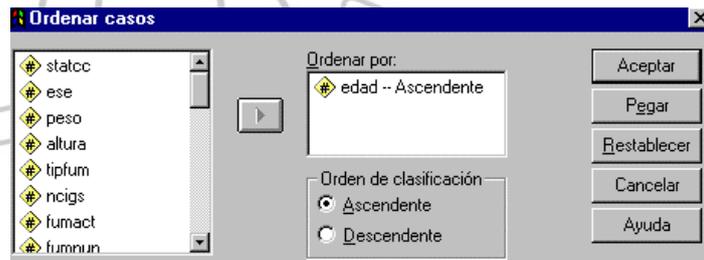


5. GESTIÓN Y TRANSFORMACIÓN DE ARCHIVOS

En muchas ocasiones los archivos de trabajo no presentan la disposición óptima para el trabajo que se pretende realizar. Es posible que tengamos un archivo o base de datos “en bruto” de forma que necesitemos efectuar determinadas transformaciones para facilitar las operaciones. Estas transformaciones incluyen cambiar la ordenación de los casos, definir subconjuntos de casos, etc. A continuación, se detallan algunas de las más interesantes.

5.1. ORDENAR CASOS

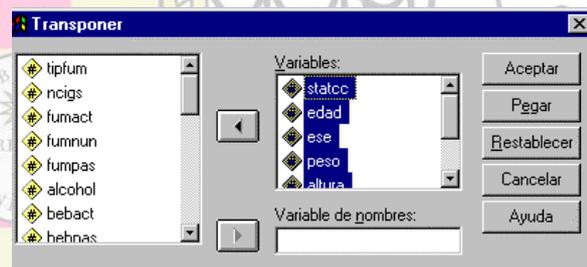
Esta función ordena todos los casos a partir de los valores de una o más variables, de forma ascendente o descendente. Hay que elegir los menús **Datos** y posteriormente **Ordenar casos**.



Si se selecciona más de una variable de ordenación, primero se ordenarán los casos en función de los valores de la primera variable, y si el valor de ésta es igual en más de un caso, quedarán ordenados en función de los valores de la segunda variable, y así sucesivamente.

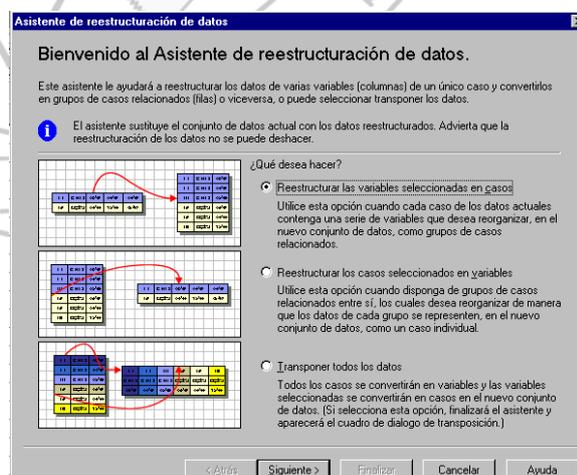
5.2. TRANSPONER ARCHIVOS

Transponer crea un nuevo archivo de datos en el que se transponen las filas y las columnas del archivo original de forma que los casos pasan a ser variables y las variables, casos, creándose automáticamente una nueva variable con los nombres de las variables originales.



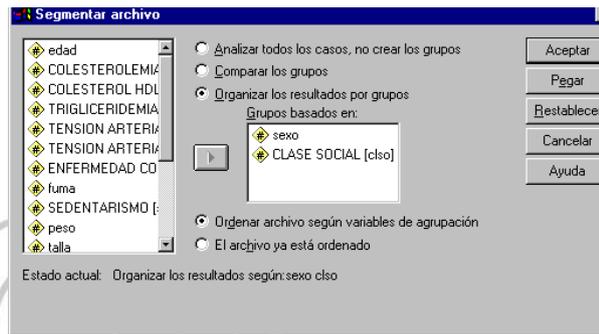
Al cuadro de diálogo se accede a través de los menús **Datos** y **Transponer**, y en él hay que seleccionar las variables que se quieren utilizar para hacer la transposición; las variables no utilizadas se perderán, por lo que será prudente grabar el nuevo archivo con distinto nombre si no se quiere perder definitivamente dicha información.

Otra posibilidad es elegir en el menú **Datos** la opción **Reestructurar**. Se abrirá un cuadro de diálogo que muestra un asistente para modificar de forma más compleja el archivo de datos en función de sus casos y sus variables. Una de las opciones de este asistente es la transponer el archivo, que es idéntica a la comentada anteriormente.

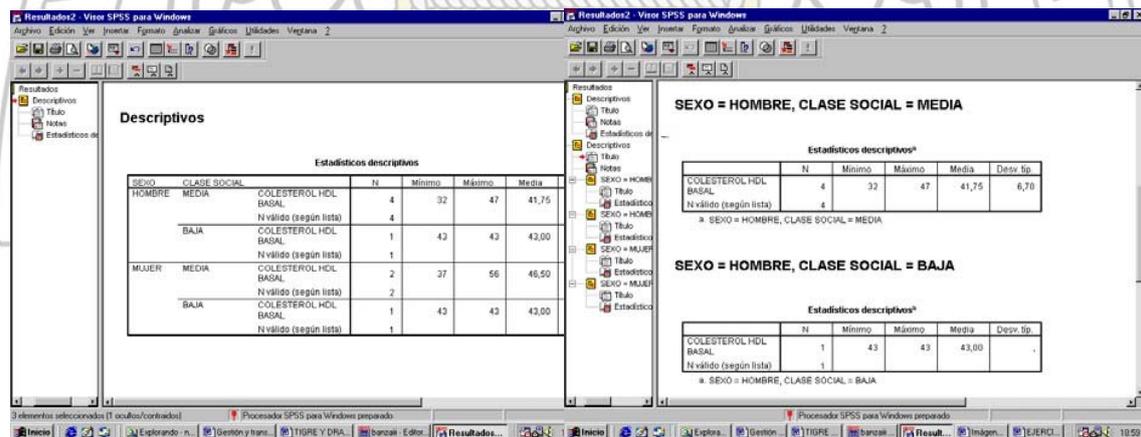


5.3. SEGMENTAR ARCHIVO

Esta función segmenta o divide el archivo de datos en distintos grupos, que serán analizados de forma aislada cuando se realice cualquier procedimiento estadístico. Pueden seleccionarse varias variables de agrupación, de forma que primero se crearán los grupos en función de la primera variable seleccionada, y luego



cada uno de los subgrupos a su vez se segmentará según la segunda variable seleccionada, y así sucesivamente. Pueden seleccionarse hasta ocho variables de segmentación. Para hacer esta operación hay que seleccionar los menús **Datos** y **Segmentar archivo**, apareciendo el cuadro de diálogo correspondiente. En la ventana de la izquierda aparecen las variables del archivo de trabajo, que podremos seleccionar para segmentarlo, pasándolos a la ventana **Grupos basados en**. El cuadro da diversas opciones: con la primera no se crea ningún grupo; con la opción **Comparar los grupos** se realiza la segmentación presentándose los datos juntos para poder compararlos, y con la tercera opción, **Organizar los resultados por grupos**, se hace la misma segmentación, pero la ventana Resultados presenta los datos organizados por separado e



identificados según las variables seleccionadas.

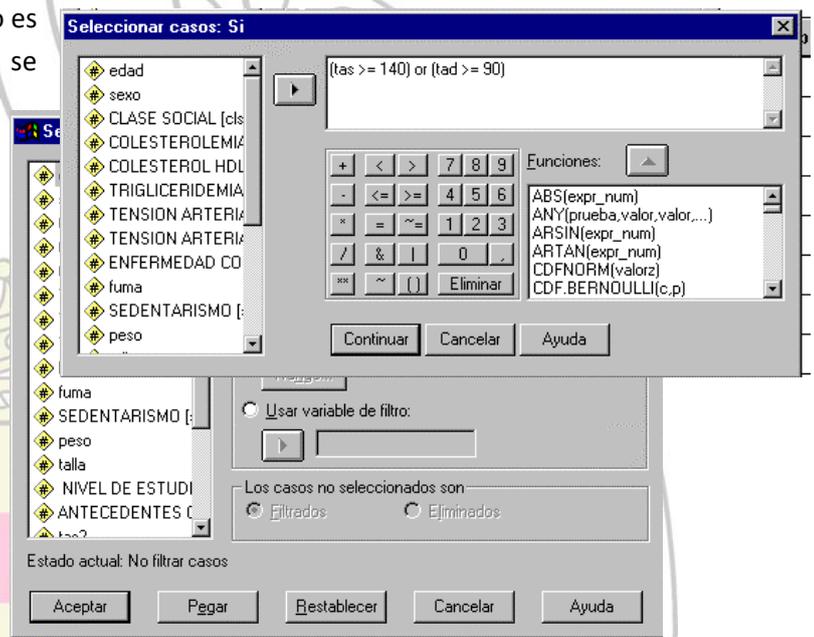
5.4. SELECCIONAR CASOS

Esta es una opción que permite seleccionar un grupo de casos del archivo de trabajo en función de diversos criterios que pueden establecerse en el cuadro de diálogo. Los casos seleccionados serán los utilizados únicamente en cualquier proceso estadístico que se realice. Los casos no seleccionados pueden ser eliminados total o temporalmente (filtrados). Es aconsejable filtrar los casos, ya que en caso contrario la información eliminada se pierde definitivamente; en caso de elegir la opción **Eliminados**, los casos se eliminarán definitivamente en el momento en que se guarden los cambios en el archivo de trabajo. Se accede a esta función seleccionando los menús **Datos y Seleccionar casos**.

La primera opción del cuadro de diálogo es **Todos los casos**; en esta ocasión no se realiza selección alguna, o se deshace una selección realizada previamente.

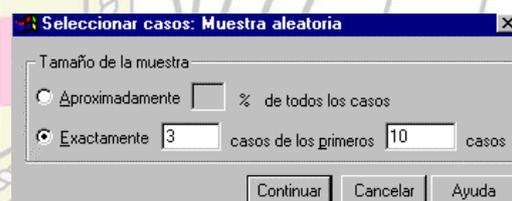
Seleccionar si se satisface la condición

Al seleccionar la opción **Si se satisface la condición** se activa el botón **Si**, que hemos de pulsar para establecer dicha condición. Aparecerá el cuadro de diálogo **Seleccionar casos: Si**. Con él podremos definir qué casos queremos seleccionar, mediante las variables del archivo de trabajo, operaciones y funciones.



5.4.1. SELECCIONAR UNA MUESTRA ALEATORIA DE CASOS

Mediante esta opción podemos seleccionar un subconjunto de casos de forma aleatoria. Elegiremos la opción **Muestra aleatoria de casos** y pulsaremos el botón **Muestra** que se activa.



Aparecerá el subcuadro de diálogo **Seleccionar casos: Muestra aleatoria**. El tamaño de la muestra seleccionada puede ser:

- ✓ **Aproximadamente**: podemos elegir un porcentaje aproximado de todos los casos del archivo de trabajo. Lógicamente, cuanto más casos tengamos en el archivo de trabajo, más se aproximará el porcentaje de casos seleccionado efectivamente al porcentaje especificado.
- ✓ **Exactamente**: nos permite elegir un número exacto de casos, especificando el número de casos a partir de los que se generará la muestra. Si queremos que sea a partir de todos los casos del archivo, en la segunda casilla el número total de casos del archivo de trabajo.

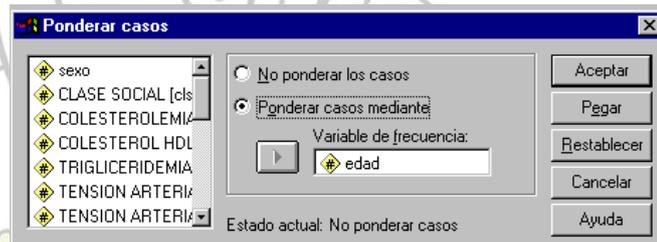
5.4.2. SELECCIONAR CASOS EN FUNCIÓN DE UN RANGO

Esta opción selecciona un subconjunto de casos comprendidos en un rango que nosotros mismos establecemos; este rango se basa en los números de caso del Editor de datos.



5.4.3. PONDERAR CASOS

Con esta aplicación podemos asignar pesos a los casos en función de una variable determinada. Al aplicarse una variable de ponderación aparecerá el mensaje Ponderado en la barra de estado, y la ponderación seguirá activada hasta que se aplique otra o se desactive la ponderación. Los casos con valores perdidos, negativos o cero para la variable de ponderación son excluidos de los análisis estadísticos. Esta opción se realiza eligiendo los menús **Datos** y **Ponderar casos**, y entonces aparece el cuadro de diálogo correspondiente, en el cual hemos de seleccionar la variable de ponderación, de manera que sus valores se usan como ponderaciones de los casos (por ejemplo, un caso con un valor de 5 en la variable de ponderación representará cinco casos al realizar un procedimiento estadístico).



Esta opción es especialmente útil para introducir datos en una tabla de contingencia de forma rápida, como se ilustra en el ejemplo siguiente:

Se realiza un estudio sobre la asociación entre la ingesta de alcohol y la presencia de enfermedad cardiaca, con 2.000 pacientes, en los que se recoge si han consumido alcohol de forma regular durante los últimos 5 años, y si han sufrido enfermedad cardiaca en el último año. Para ver esa posible asociación podríamos crear una base de datos de los 2.000 individuos, con dos variables, con valores 0 (No) y 1 (Sí). Por lo tanto, habría cuatro posibilidades de pacientes:

- ✓ Consumo de alcohol (1) sin enfermedad cardiaca (0).
- ✓ Consumo de alcohol (1) con enfermedad cardiaca (1).
- ✓ No consumo de alcohol (0) sin enfermedad cardiaca (0).
- ✓ No consumo de alcohol (0) con enfermedad cardiaca (1).

En cada una de estas posibilidades existiría un número de pacientes (frecuencia), de modo que la suma de las cuatro frecuencias es la suma de todos los pacientes del estudio (2000). Así podríamos crear una base de datos reducida con todos estos datos:

Si nosotros ponderamos este archivo por la variable "frec", estamos dando un peso a cada posibilidad, y este peso es el valor de dicha variable.

	alcohol	enf_car	frec	var
1	1	0	650	
2	1	1	370	
3	0	0	400	
4	0	1	480	
5				
6				

De este modo, si hacemos una tabla de contingencia para evaluar la asociación, no estaremos evaluando cuatro individuos, sino 2000:

Resultados2 - Visor SPSS

Archivo Edición Ver Insertar Formato Analizar Gráficos Utilidades Ventana ?

Resultados

- Tablas de contingencia
 - Título
 - Notas
 - Resumen del procedimiento
 - Tabla de contingencia
 - Pruebas de chi-cuadrado
 - Estimación de parámetros

Tabla de contingencia Enfermedad cardiaca * Consumo de alcohol

Recuento

		Consumo de alcohol		Total
		No	Sí	
Enfermedad cardiaca	No	420	650	1070
	Sí	560	370	930
Total		980	1020	2000

Pruebas de chi-cuadrado

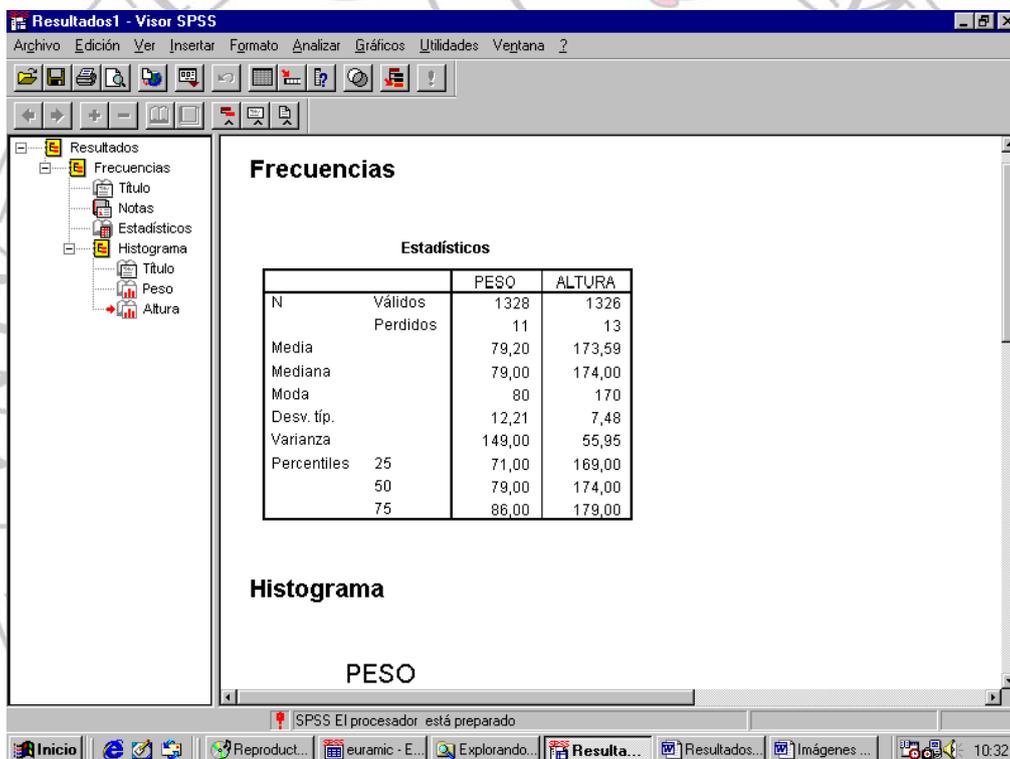
	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	87,491 ^b	1	,000		
Corrección por continuidad	86,855	1	,000		
Razón de verosimilitud	88,119	1	,000		
Estadístico exacto de					



6. VISOR DE RESULTADOS

En esta ventana se muestran todos los resultados obtenidos al ejecutar un comando o procedimiento de análisis estadístico o gráfico. Los resultados se ofrecen con un formato por defecto (normalmente tablas para los estadísticos e imágenes para los gráficos) que no siempre nos permite su uso directo por defecto, ya que ese formato puede no ser el que más nos interesa, por lo que el formato habrá que modificarlo o adaptarlo de forma que se adapte a lo que nuestro modelo de resultados requiere.

La pantalla del visor de resultados se divide en dos ventanas. La ventana de la derecha contiene los resultados propiamente dichos, mientras que la de la izquierda muestra un índice de todos los resultados de los procedimientos que hemos ido realizando y se van agregando sucesivamente en esta pantalla.



Para ir viendo todos los resultados obtenidos se puede recorrer el panel derecho con la barra de desplazamiento o bien pulsar en cada elemento del índice de titulares en el panel izquierdo. Las características del panel de resultados (tamaño y fuente de los resultados) pueden modificarse con el menú **Ver**.

6.1. MOSTRAR Y OCULTAR RESULTADOS

Cuando tenemos una gran cantidad de resultados en la ventana puede ser útil ocultar algunos de ellos sin eliminarlos para centrarnos en aquellos que nos en ese momento nos interesa. Esto podemos hacerlo de dos maneras:

- ✓ Pulsar dos veces con el ratón sobre el titular del resultado correspondiente.
- ✓ Seleccionarlo y después elegir los menús Ver y Ocultar, o bien pulsar el icono

En ese momento el icono del elemento, que hasta entonces tenía la forma de un libro abierto, cambia de aspecto, pasando a libro cerrado, y además aparecerá un mensaje en la barra de estado indicándonos que ese elemento está oculto.

Para volverlo a mostrar hay que hacer la misma operación, pero a la inversa.

Los resultados pueden desplazarse en la ventana para ordenarlos como más nos interese:

- ✓ Pulsando y arrastrando con el ratón.
- ✓ Con las opciones **Cortar** y **Pegar** del menú **Edición**.

También puede eliminarse cualquier elemento (tabla, gráfico, título) seleccionándolo y pulsando la tecla **Supro** con los menús **Edición** y **Eliminar**.

6.2. INSERTAR

También pueden insertarse objetos en la página de resultados mediante el menú **Insertar**: pueden insertarse títulos, cuadros de texto para realizar anotaciones que nos interesen, otros gráficos de otros archivos, archivos externos, o en general cualquier objeto.

Estas operaciones pueden hacerse directamente mediante el conjunto

de iconos que aparece en la barra de herramientas:



También pueden pegarse objetos procedentes de otras ventanas de resultados mediante las opciones **Cortar**, **Copiar** o **Pegar** del menú **Edición**.



6.3. FORMATO

Con este menú podemos alinear todos los resultados como más nos convenga para su presentación. Hay que seleccionar en el menú **Formato** la opción correspondiente según queramos centrar o alinear a la izquierda o derecha. Podemos ver cómo quedará la presentación mediante la presentación preliminar, en el menú **Archivo**, o con el icono

6.4. CAMBIAR EL NIVEL DE LOS TITULARES

El panel de los titulares nos muestra todos los elementos de la ventana de resultados como un árbol del que van saliendo ramas. Nosotros podemos cambiar los niveles de cada uno de los elementos de la ventana. Por ejemplo, si queremos que unos gráficos estén al mismo nivel que una tabla, o que una tabla dependa de un gráfico podemos “ascenderlo” o “degradarlo”. Estas operaciones pueden realizarse de dos modos:

- ✓ Seleccionar el elemento y pulsar el botón derecho del ratón. Elegir la opción **Ascender**

o **Degradar** del menú contextual que aparece.

- ✓ Seleccionar el elemento y pulsar uno de los iconos de la barra de herramientas:



6.5. GUARDAR LOS RESULTADOS

Como en cualquier otro trabajo, es necesario guardar los resultados que vayamos obteniendo con SPSS. Esta operación se realiza como en cualquier aplicación de Windows.

Además, en la ventana de resultados existe la opción de guardarlos con una contraseña. Para ello hay que seleccionar el menú **Archivo** y después **Guardar con contraseña**, apareciendo un cuadro donde se nos pide introducir y confirmar dicha contraseña.

Guardar con contraseña

Contraseña:

Confirmar contraseña:

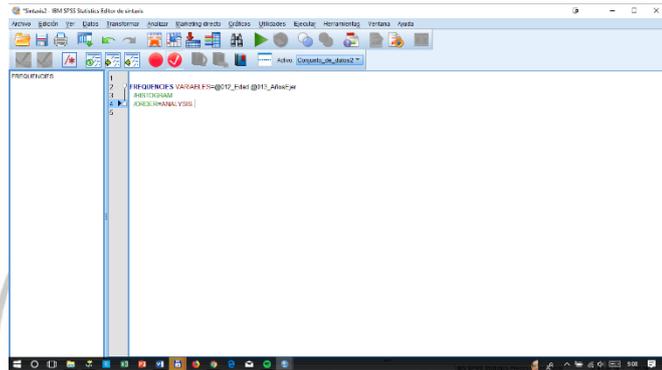
Código OEM:

Atención: Si pierde u olvida la contraseña, no podrá recuperar los datos. Es recomendable guardar las contraseñas y sus documentos correspondientes en un lugar seguro. (Recuerde que las contraseñas distinguen entre mayúsculas y minúsculas.)

Aceptar
Cancelar
Ayuda

7. SINTAXIS

Las instrucciones y procedimientos que SPSS ejecuta, pueden guardarse en una ventana aparte en forma de comandos de sintaxis, de manera que pueden guardarse en un fichero en el ordenador (fichero con extensión **.sps**) y utilizarse posteriormente (abriendo el fichero de sintaxis y ejecutándolos) para reproducir



nuevamente estos procedimientos sin tener que volver a seleccionarlos con una sola orden. El programa repetirá de nuevo todas las instrucciones que hayamos almacenado previamente en ese fichero. Estos comandos se guardan en unos archivos especiales de sintaxis, que no son más que archivos de texto que contienen comandos.

Los comandos también pueden escribirse directamente sobre una ventana de sintaxis si se conoce el lenguaje, pero existen otros procedimientos más sencillos para que el programa traslade los comandos a la ventana de sintaxis, y sin errores de escritura.

Cuando se va a ejecutar cualquier análisis (procedimiento estadístico, gráfico), en cada cuadro de diálogo aparece el botón **Pegar** que, al pulsarlo, crea automáticamente una ventana de sintaxis con los comandos para ejecutar dicho procedimiento.

Además, en los titulares de la ventana de resultados aparece un apartado de **Notas** es los que aparecen estos comandos igualmente.

Notas		
Resultados creados		05-MAY-1995 19:38:16
Comentarios		
Entrada	Datos	C:\Medicina Preventiva\prácticas\estadística spps\euramic.sav
	Filtro	<ninguna>
	Peso	<ninguna>
	Segmentar archivo	<ninguna>
	Núm. de filas del archivo de trabajo	1339
Manipulación de los valores perdidos	Definición de los perdidos	Los valores perdidos definidos por el usuario serán tratados como perdidos.
	Casos utilizados	Los estadísticos se basan en todos los casos con datos válidos.
Sintaxis		FRECUENCIAS VARIABLES=peso edad /FORMAT=NOTABLE /NTILES= 4 /STATISTICS=STDDEV VARIANCE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM /ORDER= ANALYSIS .

En la ayuda de cada cuadro de diálogo, puede hallarse información sobre los comandos disponibles para ese procedimiento, ya que en la lista de Temas relacionados hay una opción que nos la da: **Command Syntax**.

7.1. Archivo diario

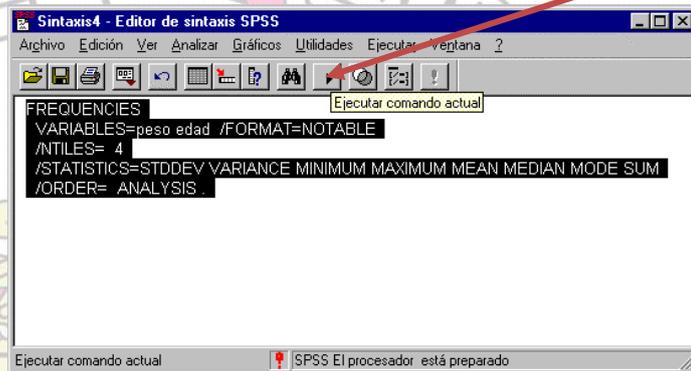
No obstante, todos los procedimientos que vayamos ejecutando van almacenándose en un archivo diario que se llama **spss.jnl**. En este archivo se van almacenando automáticamente y por defecto todos los procedimientos efectuados en cada sesión, y puede abrirse como un archivo de sintaxis (de texto). Para localizarlo tendremos que buscar y elegir del archivo **spss.jnl**

en **Opciones** del menú **Edición** del editor de datos, en la pestaña **General**.



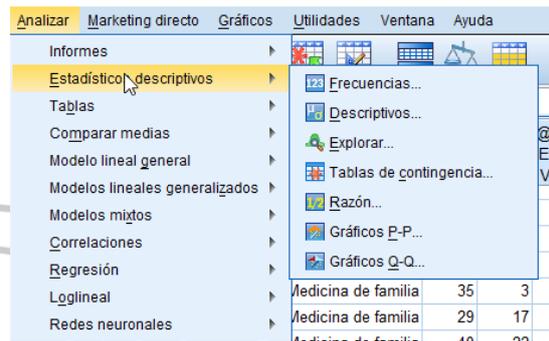
7.2. Ejecución de comandos desde la sintaxis

Para ejecutar comandos de sintaxis primero hemos de seleccionarlos en la ventana en la los tengamos, y después hay que pulsar el botón de ejecución en la barra de herramientas, o bien elegir el menú **Ejecutar** y después la opción que queramos del menú desplegable que aparece.



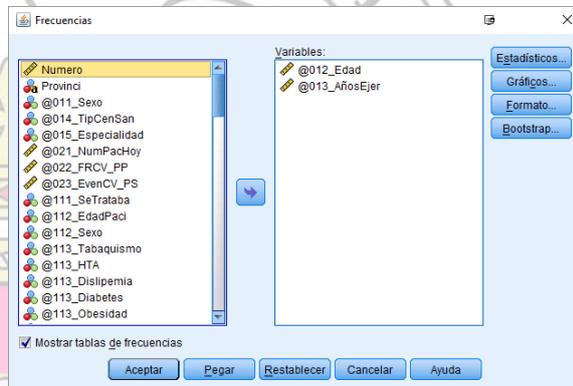
8. ESTADÍSTICOS DESCRIPTIVOS

El submenú **Estadísticos descriptivos** está en el menú **Analizar**, y ofrece una serie de opciones para describir datos de una forma sencilla. Este es, siempre, el primer paso que debemos hacer en un estudio estadístico, para resumir y representar de forma fácil gran cantidad de datos; y además nos permite controlar la presencia de errores en la carga de los datos. Así mismo nos permite saber qué tipo de distribución siguen nuestras variables.



8.1. FRECUENCIAS

El procedimiento **Frecuencias (Análisis, Descriptivo, Frecuencias)** permite seleccionar proporción de estadísticos descriptivos y gráficos que queremos calcular para describir y representar las variables de interés. Es el mejor procedimiento para una primera aproximación a los datos, inspección inicial de los datos.



El cuadro de diálogo **Frecuencias** muestra el cuadro de variables disponibles en el lado izquierdo, y el cuadro de variables que vamos a usar para el análisis en el lado derecho. Además, aparecen unos botones para seleccionar los datos que queremos solicitar, además de los habituales botones de comando de todos los cuadros de diálogo. También hay una opción de selección **Mostrar tablas de frecuencias**, útil sobre todo en variables categóricas, que nos mostrará la frecuencia y porcentaje de aparición de cada valor.

8.1.1. ESTADÍSTICOS

Al pulsar el botón **Estadísticos** se abre el subcuadro de diálogo correspondiente, en el que se muestran los valores estadísticos que podemos solicitar en este procedimiento.

8.1.1.1. VALORES PERCENTILES

Se trata de estadísticos que informan sobre la posición relativa (posición porcentual) de cada dato en relación a todos los demás, una vez ordenados de menor a mayor. Es decir, el percentil 12 nos da el valor que ocupa la posición 12% respecto del total de datos de nuestra muestra cuando están ordenados de menor a mayor. Podemos solicitar los **Cuartiles** (valores que dividen la distribución en cuatro partes iguales (Primer cuartil es el valor que ocupa la posición 25%, Segundo cuartil es el valor que ocupa la posición 50% y el Tercer cuartil es el valor que ocupa la posición 75%), o los valores que ocupan unas determinadas posiciones en tramos



iguales de la distribución de datos que nos interese, mediante la opción **Puntos de corte para n grupos iguales**. También podemos pedir un determinado percentil individual, seleccionando la opción **Percentiles** y especificando cuál o cuáles queremos ver, pulsando el botón **Añadir** para cada uno de ellos.

8.1.1.2. MEDIDAS DE TENDENCIA CENTRAL

Las medidas de la tendencia central calculan estadísticos que informa de alrededor de qué valores se distribuye nuestra muestra, de cual es el valor “más representativo” o “más probable” de la muestra, o alrededor de qué valor es el que más repite. En este procedimiento se incluyen, además de los cuantiles, la media aritmética, mediana, moda, y la suma de todos los valores.

8.1.1.3. MEDIDAS DE DISPERSIÓN

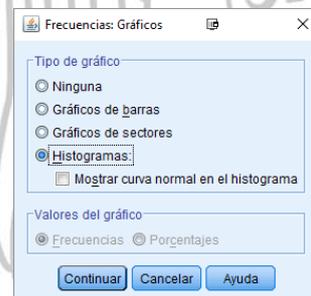
Estas medidas nos informan del grado de variabilidad de los datos de la muestra, de cómo los valores de la variable se sitúan alrededor de la media, es decir, más agrupados o más dispersos mas cercano o más lejanos. Se incluyen La desviación típica (este es el término en español de desviación estándar –*standard deviation*- en inglés), varianza, rango, mínimo, máximo y el error típico de la media (este ultimo es el término en español de error estándar –*standard error*- en inglés).

8.1.1.4. FORMA Y SIMETRÍA DE LA DISTRIBUCIÓN

Presenta dos índices: asimetría y curtosis. De muy escaso uso en Ciencias de la Salud.

8.1.2. GRÁFICOS

A partir de aquí podemos seleccionar un tipo de gráfico en función del tipo de variables que hayamos elegido para el análisis estadístico. La ventana de resultados nos mostrará un gráfico para cada una de las variables escogidas. El cuadro de diálogo **Frecuencias: Gráficos** nos ofrece varias opciones, comenzando por no solicitar ningún gráfico.



8.1.2.1. GRÁFICOS DE BARRAS Y SECTORES

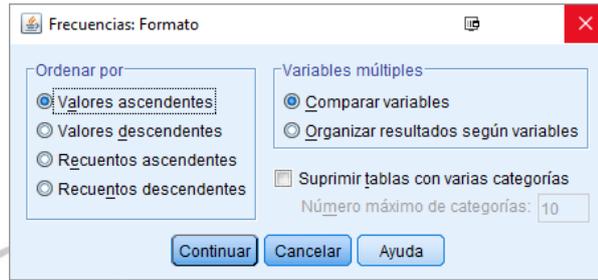
Se emplean para la caracterización de variables categóricas o cuantitativas discretas. Cuando escogemos uno de estos gráficos se activan las opciones de **Valores del gráfico**, pudiendo seleccionar ver las frecuencias absolutas para cada valor o bien los porcentajes.

8.1.2.2. HISTOGRAMAS

Son utilizados para variables cuantitativas continuas. Al seleccionar este gráfico se activa la opción **Con curva normal**. Si activamos esta última opción, el gráfico sobrepondrá la curva normal teórica que correspondería a esa muestra.

8.1.3.FORMATO

Al pulsar el botón **Formato** se muestra el correspondiente cuadro de diálogo, con algunas opciones. La primera de ellas es la posibilidad de suprimir las tablas con un número de valores muy elevado (que nosotros hemos de especificar).



8.1.3.1. ORDENAR POR

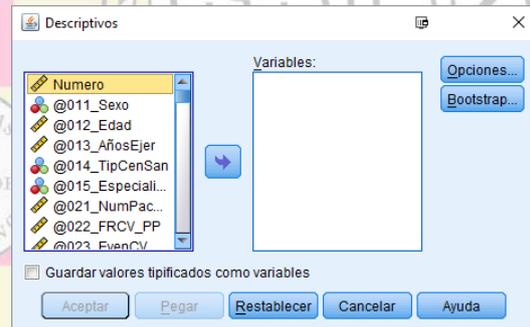
Cuando hemos solicitado una tabla de frecuencias, esta opción está activada, y nos permite especificar cómo queremos que esa tabla esté ordenada, en función de los valores o de las frecuencias de aparición, y de forma ascendente o descendente. Si se solicita un histograma o valores percentiles, SPSS asumirá que la variable es cuantitativa y mostrará los valores de forma ascendente.

8.1.3.2. MÚLTIPLES VARIABLES

Si vamos a analizar varias variables los resultados pueden salir en una sola tabla, o bien en tablas individuales para cada variable.

8.2. DESCRIPTIVOS

Mediante el procedimiento **Descriptivos** pueden obtenerse un grupo de estadísticos de resumen para varias variables en una única tabla, y además pueden calcularse los valores normalizados para los valores de las variables. Es útil fundamentalmente para trabajar con variables cuantitativas continuas. En el cuadro de diálogo **Descriptivos** se muestran las variables disponibles a la izquierda, y en el lado derecho las variables que se quiere analizar. También se halla la opción **Guardar valores tipificados como variables**, que calcula los valores Z de la variable tipificada y los guarda en una nueva variable.

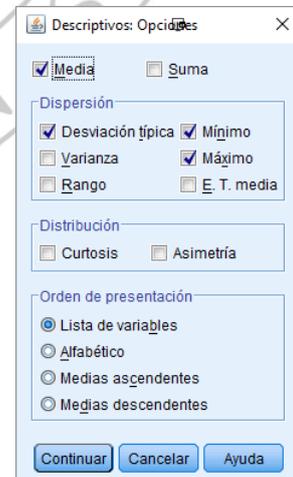


Además de los habituales botones de comandos, está el botón **Opciones**, queda paso al cuadro de diálogo **Descriptivos: Opciones**, en el que podremos escoger los estadísticos que deseemos.

Pueden elegirse la media y la suma total de los valores de la variable como medidas de la tendencia central. Como medidas de dispersión, las mismas que en la opción **Frecuencias**, así como

los índices de asimetría y curtosis.

Los resultados pueden visualizarse en diferente orden: según la lista de variables, por orden alfabético, o bien por orden ascendente o descendente de las medias aritméticas.

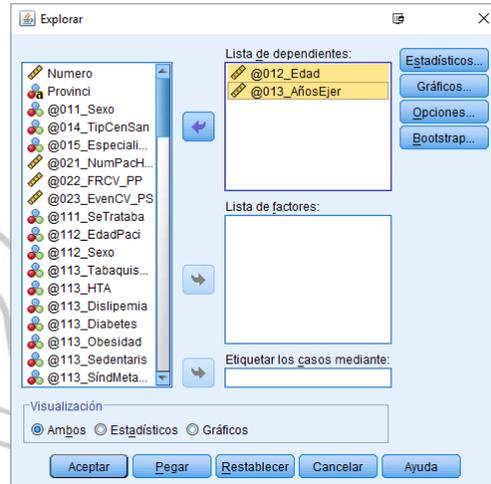


8.3. EXPLORAR

Es un procedimiento que genera estadísticos y gráficos de forma un poco más compleja que los anteriores, para todos los casos de las variables elegidas o para grupos de casos seleccionados. Nos permite estratificar los resultados sin necesidad de segmentar el archivo, por unas variables que llama *Factores*. Al elegir la opción **Explorar** se abre su cuadro de diálogo, con las distintas opciones.

En el lado derecho del cuadro hay tres campos. En el superior (**Dependientes**) seleccionaremos las variables que deseamos analizar. Si queremos incluir todos los casos no hay que seleccionar nada más, pero podemos dividir los casos en función de otra u otras variables, pasándolas al campo **Lista de Factores**. Además, cuando solicitemos (como se ve más adelante) datos para casos individuales, podemos etiquetar éstos en función de los valores de la variable que elijamos en el campo **Etiquetar los casos mediante**.

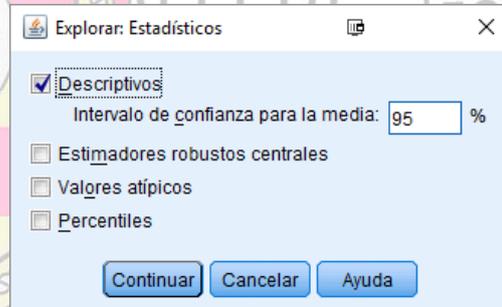
Una vez realizada esta selección hemos de escoger la información que queremos obtener. Ésta puede ser en forma de estadísticos, gráficos o ambos, y para ello marcaremos la correspondiente opción en el cuadro **Mostrar**.



8.3.1. ESTADÍSTICOS

Al pulsar este botón se abre el subcuadro de diálogo **Explorar: Estadísticos**, en el que se presentan cuatro posibilidades no excluyentes.

1. **Descriptivos**. Muestra medidas de la tendencia central y de dispersión. Además, se muestra el intervalo de confianza al 95 % para la media, aunque puede cambiarse el nivel de confianza.
2. **Estimadores robustos centrales**. Se trata de estimadores especiales alternativos a la media y la mediana.
3. **Valores atípicos**. Muestra los cinco valores de la variable mayores y menores. Si hemos seleccionado una variable para etiquetar, al lado de cada caso aparece la etiqueta correspondiente.
4. **Percentiles**. Muestra los valores de los percentiles 5, 10, 25, 50, 75, 90 y 95.

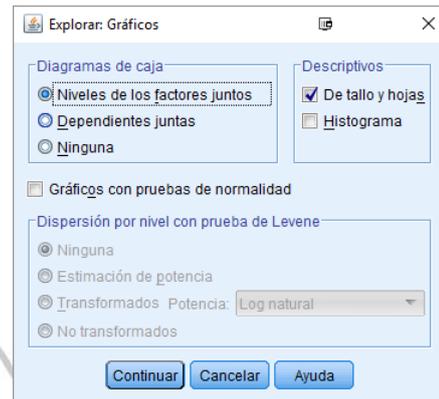


8.3.2. GRÁFICOS

5. **Diagramas de Caja**. Pueden solicitarse los diagramas para cada variable independiente seleccionada mediante la opción Niveles de los factores juntos. Si se desea ver los diagramas para todas las variables seleccionadas hay que escoger la opción Dependientes juntas; ésta es bastante útil cuando las variables representan una misma

característica medida en momentos distintos.

6. **Descriptivos.** Pueden solicitarse el diagrama de tallo y hojas y el histograma.
7. **Gráficos con pruebas de normalidad.** Realiza pruebas de normalidad (Kolmogorov- Smirnov y Shapiro- Wilk) a las variables seleccionadas, y muestra dos tipos de gráficos para explorar el ajuste a la normal.
8. **Dispersión por nivel con prueba de Levene.**



Efectúa una prueba de homogeneidad de la varianza (Levene) y permite seleccionar diversas transformaciones para los datos.

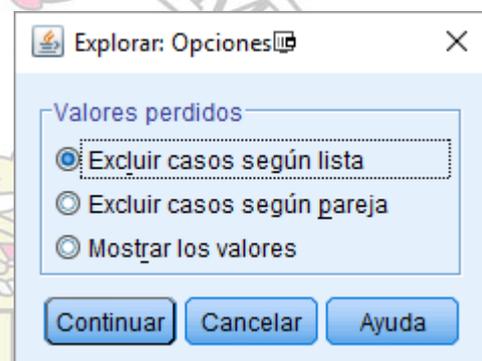
8.3.3. OPCIONES

Tras pulsar el botón **Opciones** se abre el correspondiente cuadro de diálogo, donde podemos tratar de distinto modo en el análisis los valores perdidos.

Excluir casos según lista. En este caso quedarán excluidos del análisis todos los casos que tengan valor perdido para cualquiera de las variables seleccionadas.

Excluir casos según pareja. Sólo excluye del análisis a los valores que tengan valor perdido en cada una de las variables individualmente.

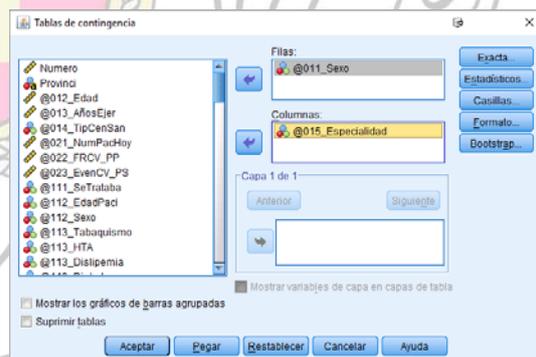
Mostrar los valores. Genera una categoría diferente en el análisis para los valores perdidos de las variables de factor.



8.4. TABLAS DE CONTINGENCIA

Este procedimiento crea tablas para analizar la asociación entre variables categóricas. En el cuadro de diálogo **Tablas de contingencia** aparecen las diferentes opciones.

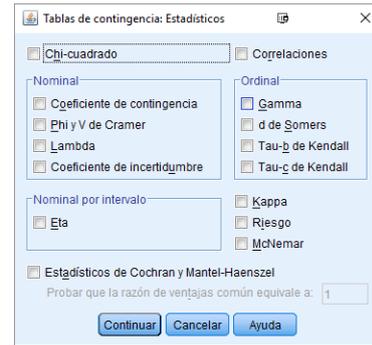
En los campos **Filas y Columnas** incluiremos las variables que queremos analizar. También se nos ofrece la opción de ver los gráficos de barras correspondientes a las variables elegidas, así como no ver las tablas, sino solamente los estadísticos.



Además, podemos realizar estratificación por diferentes variables si las incluimos en el campo **Capa**, obteniendo todos los resultados para el análisis estratificado.

8.4.1. ESTADÍSTICOS

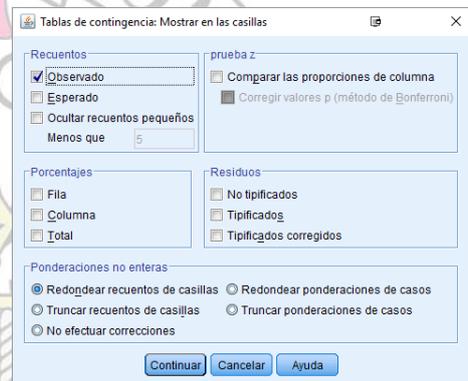
- ✓ **Chi-cuadrado.** Calcula los valores de chi-cuadrado de Pearson, de la razón de verosimilitud, de Fisher (equivalente no paramétrico del Pearson) y el corregido de Yates.
- ✓ **Correlaciones** da los coeficientes de correlación como medida de asociación lineal entre las variables seleccionadas.
- ✓ **Riesgo.** En tablas 2x2 estima el riesgo relativo y el odds ratio.
- ✓ **Mc Nemar.** Prueba no paramétrica para dos variables dicotómicas dependientes.



8.4.2. CASILLAS

Ofrece la posibilidad de mostrar en la tabla de contingencia las frecuencias esperadas y observadas para cada variable, así como los porcentajes correspondientes a las filas y las columnas.

Además pueden incluirse los residuos (diferencia entre los valores observados y esperados).

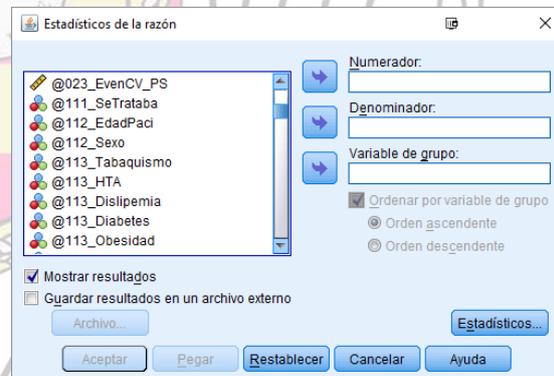


8.4.3. FORMATO

Puede ordenar las filas en orden ascendente o descendente de los valores de la variable de fila.

8.5. RAZÓN

Este procedimiento permite describir el comportamiento de una razón entre dos variables cuantitativas, mediante una serie de estadísticos descriptivos. Para acceder a él hay que elegir la opción **Razón** de los menús **Análisis** y **Estadísticos Descriptivos**, abriéndose un cuadro de diálogo. En él hay que elegir las variables **Numerador** y **Denominador** de la razón que se describe, y además se puede escoger una **Variable de agrupación**, o factor. En el botón **Estadísticos** pueden escogerse los parámetros que desea utilizar para describir la razón.



9. REPRESENTACIONES GRÁFICAS

SPSS incluye una amplia gama de posibilidades para realizar representaciones gráficas de variables o de procedimientos estadísticos. Antes de detallarlos, repasaremos brevemente qué tipos de variables podemos encontrar en un análisis estadístico.

Las variables pueden ser de dos tipos:

Cuantitativas: son cuantificables, toman valores numéricos. Tiene sentido operar con ellas (se puede calcular una diferencia de edades o de pesos, por ejemplo) y se pueden comparar (por ejemplo, ver si un grupo de pacientes tienen una concentración de ácido úrico en suero mayor que otros) Pueden ser:

- ✓ Continuas: la variable puede adoptar cualquier valor numérico dentro de sus valores posibles (Pueden tener tantos decimales como seamos capaces de medir). EJEMPLO: cifras de colesterol en sangre, talla.
- ✓ Discretas: sus valores son siempre números enteros (aunque en ocasiones se usa, conceptualmente no podemos aplicarles decimales). EJEMPLO: número de hijos.

Cualitativas: expresan cualidades, y no adoptan valores numéricos. EJEMPLO: sexo (masculino, femenino). Dentro de las variables cualitativas se distinguen a su vez tres clases:

- ✓ Cadena: En ocasiones también se etiquetan como variables tipo **Texto**. Son cadenas de tipo alfanuméricas, pudiendo tomar cualquier valor. Son las que aportan menos potencia en el análisis estadístico. EJEMPLO: Nombre, Dirección, DNI, Número de historia. En ocasiones, aunque estén representadas solo por números no tiene un sentido cuantitativo, y no es coherente operar con ellas o compararlas. Solo se utilizan para los datos identificativos. En ocasiones registramos respuestas de los pacientes como texto libre. En estos casos su explotación estadística exige una recodificación previa en otra variable de tipo categórica.
- ✓ Nominales categóricas: los datos se agrupan por categorías en las cuales se está o no incluido. EJEMPLO: grupo sanguíneo ABO (1=A, 2=B, 3=AB, 4=0) o Sexo (1=Hombre y 2=Mujer). Son exhaustivas y mutuamente excluyentes. En SPSS se introducirán como numéricas y posteriormente se etiquetarán con su nombre correspondiente. A estas variables a pesar de estar introducidas como un número, no les serán de aplicación pruebas estadísticas de variables cuantitativas, como la media o la desviación típica.
- ✓ Ordinales: existe una relación de orden dentro de las categorías. EJEMPLO: gravedad de una enfermedad (leve, moderada, grave).

La tabulación y la representación gráfica de los datos pretende favorecer una fácil y más intuitiva interpretación y comprensión de los resultados, así como su análisis. Debe buscarse la claridad a través de la simplificación y la sencillez. Los gráficos tienen la ventaja de que pueden comunicar mucha información de una forma casi instantánea. A continuación, se presentan los más utilizados.

9.1. VARIABLES CUANTITATIVAS CONTINUAS:

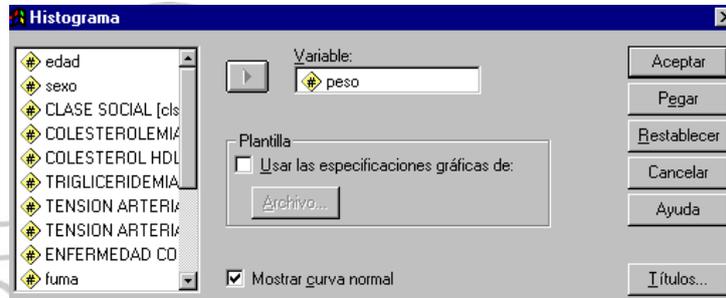
9.1.1. HISTOGRAMA:

En el eje de abscisas se seleccionan los intervalos en los que se dividen los valores de la variable, y en el eje de ordenadas la frecuencia correspondiente de cada intervalo. Así se obtienen unos rectángulos cuyas áreas son proporcionales a las frecuencias de cada intervalo.

Para realizarlo elegiremos los menús **Gráficos** e **Histograma**, apareciendo el correspondiente cuadro de diálogo:

Elegiremos una variable continua y pulsaremos **Aceptar**.

Puede elegirse que sobre el histograma se superponga una curva normal, y además pueden insertarse títulos y subtítulos a través del botón **Títulos**.

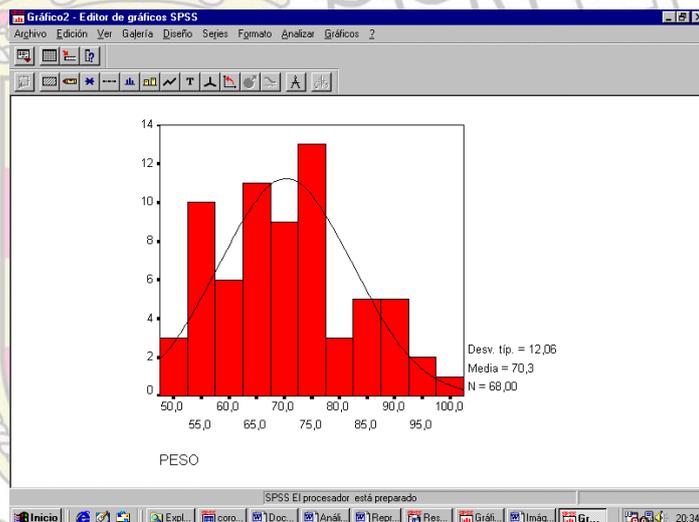


9.1.2. EDITOR DE GRÁFICOS

Para cambiar las características del histograma iremos al Editor de Gráficos de SPSS pulsando dos veces con el ratón sobre el gráfico.

9.1.2.1. MENÚ DISEÑO

- **Opciones:** permite incluir o eliminar la curva normal y la leyenda. También se puede acceder pulsando el botón
- **Ejes:** con esta opción pueden modificarse las características de los ejes de coordenadas. Para ello antes hay que seleccionar uno de los ejes. Pueden modificarse el título del eje, el rango de valores visualizados, el número de intervalos en el eje X, pueden introducirse etiquetas y variar su formato, o variar las marcas señalizadoras.
- **Título, Nota a pie de página, Leyenda, Anotación:** introduce, elimina o modifica las características de estos elementos. También puede hacerse haciendo doble "click" sobre cada uno de ellos.
- **Línea de referencia:** permite insertar una o varias líneas de referencia en ambos ejes.
- **Marco interior, Marco exterior:** muestra u oculta estos elementos.

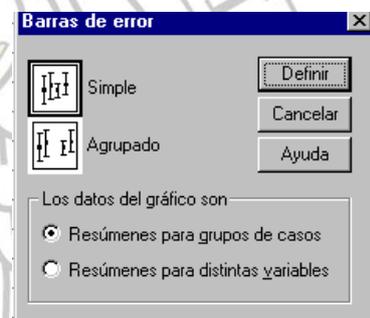


9.1.2.2. MENÚ FORMATO

- **Trama de relleno:** aplica al interior de los rectángulos la trama seleccionada. Previamente hay que seleccionar los rectángulos. También puede aplicarse la trama de relleno al área de todo el gráfico. Acceso directo con el botón .
- **Color:** puede modificar el color de cualquier elemento del gráfico (rectángulos, curva, texto, etc) seleccionándolo previamente. También mediante el botón .
- **Estilos de línea:** modifica cualquier línea seleccionada. Botón .
- **Texto:** modifica la fuente y el tamaño del texto. También con el botón .
- **Intercambiar ejes:** transpone los ejes de abscisas y ordenadas. Puede realizarse directamente con el botón .

9.1.3. BARRAS DE ERROR

Se trata de un gráfico muy usado tradicionalmente que presenta un dato que describe una variable mediante un parámetro (la media aritmética) y un intervalo de confianza. Puede elegirse la barra de error para una única variable o para comparar grupos en función, por ejemplo, de otra variable. Esta opción se encuentra en cuadro de diálogo para las barras de error.



En el Editor de Gráficos pueden modificarse las características del gráfico de forma parecida al histograma.

9.1.4. DIAGRAMA DE CAJA O BOXPLOT

La información que aportan las barras de error es a veces bastante pobre, e incluso difícil de interpretar, a la hora de describir una variable. Existe un tipo de gráfico desarrollado en los años 70 por Tukey que, siendo parecido a las barras de error, presenta múltiples ventajas, ya que ofrece una representación sintética de una distribución y permite ver de forma clara sus principales características, como son la tendencia central, dispersión, asimetría y valores anómalos. Una de sus principales particularidades es que se elabora a partir de índices estadísticos basados en ordenaciones (mediana, cuartiles), y no en momentos (media).

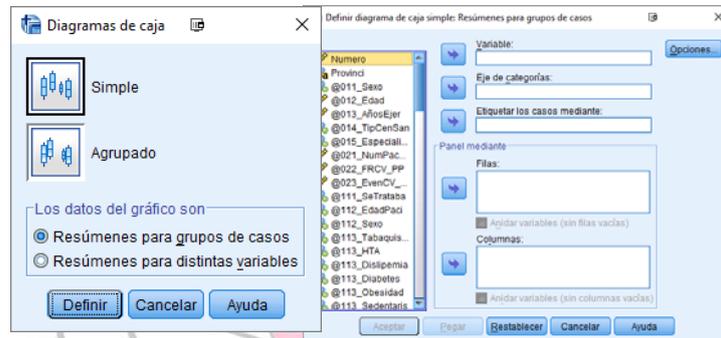
9.1.5. CONSTRUCCIÓN DEL DIAGRAMA DE CAJA

Su parte central está formada por un rectángulo de anchura arbitraria, y sus dos bases representan el primer y tercer cuartiles de la distribución. Dentro del rectángulo hay una línea de trazo grueso que representa la mediana. Además, hay otros dos trazos horizontales (patillas) que representan los valores máximo y mínimo que no se consideran anómalos (valores adyacentes). Por último, por encima y por debajo de las patillas están representados de forma individual todos los valores anómalos de la distribución. Para hallar los valores adyacentes se multiplica la amplitud intercuartil ($H_3 - H_1$) por 1,5 y el resultado se suma a H_3 y se resta a H_1 .

Con este tipo de representación gráfica se puede ver de forma muy intuitiva la tendencia central

de la distribución, la distribución del 50% central, la posible asimetría de las colas así como los valores más extremos. Es muy útil para comparar datos cuantitativos en diferentes grupos o para estudiar la evolución de un carácter diferentes momentos temporales.

El diagrama de caja puede realizarse eligiendo los menús *Gráficos* y *Diagramas de caja*. En el cuadro de diálogo puede seleccionarse si se quiere hacer el boxplot para una variable o bien comparar varios grupos. Tras pulsar el botón *Definir*



aparecerá el subcuadro de diálogo en el que hay que seleccionar las variables sobre las que se construirán los diagramas de caja.

Una vez que se pulse *Aceptar* aparecerán los resultados. En el editor de gráficos pueden modificarse algunas de sus características.

9.2. VARIABLES CUALITATIVAS Y CUANTITATIVAS DISCRETAS

9.2.1. DIAGRAMA DE BARRAS

El diagrama de barras (bar chart) es una de las opciones más utilizadas para la representación de variables categóricas o cuantitativas discretas. Se confecciona colocando en el eje de abscisas las categorías de la distribución que se quiere representar, y sobre cada categoría se construye un rectángulo de altura igual a la frecuencia absoluta o relativa de cada categoría. La base tiene anchura arbitraria pero constante. Para representar un diagrama de barras hay que elegir el menú *Barras* en *Gráficos*. Se abrirá el primer cuadro de diálogo.

Este cuadro nos permite escoger tres tipos de representación gráfica, el diagrama simple, agrupado o apilado.

Además hay otro grupo de opciones de análisis, representar grupos de casos de una misma distribución, comparar distintas variables o representar valores individuales.



9.2.2. DIAGRAMA SIMPLE

Es el caso más sencillo. En él hemos de elegir la variable que queremos representar en el cuadro Eje de categorías, y posteriormente escoger una de las opciones para el eje de ordenadas. También podemos poner títulos y subtítulos en botón *Títulos*, e incluir o no en el diagrama el grupo de los valores perdidos, mediante el botón *Opciones*.

9.2.3. DIAGRAMA AGRUPADO

En este caso representaremos una variable que elegiremos en el campo *Eje de categorías*,

pero comparándola en función de los grupos de otra variable de agrupación, que colocaremos en el campo *Definir grupos por*.

9.2.4. DIAGRAMAS APILADOS

Es como el caso anterior pero en lugar de salir las barras adosadas saldrán una encima de la otra.

9.2.5. RESÚMENES PARA DISTINTAS VARIABLES

El diagrama de barras para una variable nos mostraba las frecuencias de cada una de sus categorías. Esta opción nos dará un resumen de varias variables categóricas en función de un parámetro estadístico que nosotros seleccionemos en el botón Cambiar resumen.

9.2.6. VALORES INDIVIDUALES DE LOS CASOS

Esta opción representa en el eje de abscisas cada uno de los casos, y en el eje de ordenadas la categoría para cada caso. Es poco ilustrativo cuando hay muchos casos y la variable es dicotómica.

9.3. DIAGRAMA DE PARETO

Es un tipo especial de gráfico de barras que se usa sólo con variables categóricas puras. Tiene las barras ordenadas de forma decreciente según la frecuencia de cada categoría. Además, sobre las barras se dibuja una línea que representa las frecuencias acumuladas de las categorías. En cada barra también figura una etiqueta con la frecuencia de cada categoría. En el eje de ordenadas se representan las frecuencias tanto absolutas como relativas, una en cada lado del diagrama. Este diagrama puede ser útil cuando tenemos distribuciones con muchas categorías y se quiere destacar de forma visual su importancia relativa. Para realizar este diagrama se selecciona el menú *Pareto* en el menú Gráficos. También ofrece las opciones parecidas a las del diagrama de barras.

9.4. GRÁFICOS DE LÍNEAS

Se trata de gráficos con utilidad semejante a los diagramas de barras. También ofrecen opciones similares, con diagrama simple, múltiple y de líneas verticales, y para representar grupos de casos, distintas variables y casos individuales. Se accede a esta opción en el menú ***Líneas*** de ***Gráficos***.

9.5. DIAGRAMA DE SECTORES

En este tipo de diagrama cada clase se corresponde con un sector de superficie proporcional a su frecuencia. Las opciones para los datos y tipos de gráfico son parecidas a las anteriores.

10. CORRELACIÓN

Las técnicas de correlación permiten evaluar la existencia o no de una relación lineal entre variables cuantitativas, es decir, si a un aumento o disminución en una de las variables corresponde de forma fija un aumento o disminución en otra. Podemos analizar la correlación de forma gráfica o mediante parámetros estadísticos.

10.1. DIAGRAMA DE DISPERSIÓN

Mediante este diagrama puede observarse la distribución, caso a caso, de las variables que analicemos. Lo más habitual es hacer un gráfico entre dos variables, de forma que cada caso le corresponde un valor para cada variable, obteniéndose así una nube de puntos.

Para crear el diagrama de dispersión hay que elegir los menús *Gráficos* y *Dispersión*.

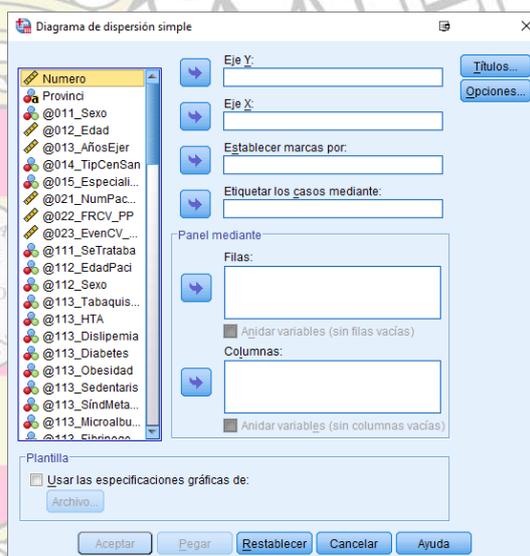


10.1.1. DIAGRAMA SIMPLE

En este caso se seleccionan dos variables, una para cada eje, y se obtiene una nube de puntos.

Puede obtenerse el diagrama en función de las categorías de otra variable cualitativa seleccionando ésta en el campo *Establecer marcas por*. Además puede etiquetarse cada caso en función de otra variable con la opción *Etiquetar los casos mediante*.

Las opciones para modificar el aspecto del gráfico se hallan en el Editor de gráficos, y son parecidas a las del resto de gráficos. Sin embargo, hay que destacar algunas.



Si seleccionamos en el Editor de gráficos los menús *Diseño* y *Opciones*, aparece un cuadro de diálogo:

En él podemos incluir en el gráfico la línea que ajusta de forma lineal los puntos del diagrama; podemos elegir la línea de ajuste para el modelo total o para cada uno de los grupos. También pueden verse todas las etiquetas de los casos, aunque esta opción sólo es útil cuando la muestra no es muy numerosa.

Pueden, no obstante etiquetarse los casos individualmente pulsando el  botón y después pulsando cada punto de la nube. Además, pueden mostrarse las líneas de referencia de la media en el eje Y. Cuando hay muchos puntos en el diagrama, algunos pueden quedar superpuestos, de forma que podemos agruparlos en los llamados “girasoles”, en los que cada “pétalo” representa un número determinado de casos.

10.1.2. DIAGRAMA SUPERPUESTO

En esta opción podemos crear varios diagramas de puntos para varios pares de variables, y así poder compararlos. Se usa fundamentalmente para comparar relaciones de variables iguales en grupos diferentes de casos.

Para crearlos hay que seleccionar al menos dos pares de variables y llevarlos al campo Pares Y-X. También pueden etiquetarse los casos en función de las categorías de otra variable.

10.1.3. DIAGRAMA MATRICIAL

El diagrama matricial muestra varias nubes de puntos para la correlación lineal de varias variables intercambiando los ejes X e Y.

Pueden elegirse más de dos variables para la matriz, visualizarse según otra variable categórica, etiquetar los casos.

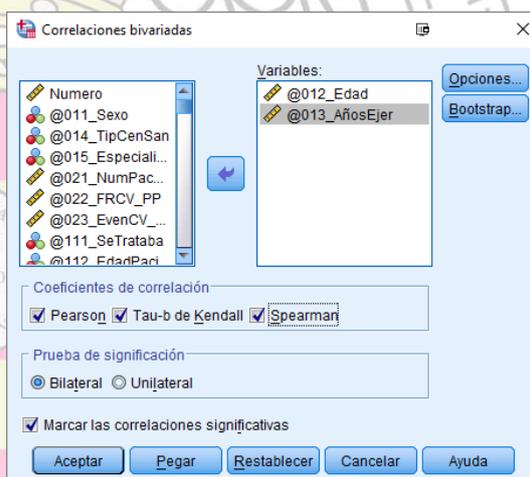
10.1.4. GRÁFICO 3-D

En este caso pueden representarse tres variables relacionadas entre sí, apareciendo la nube de puntos. En el editor de gráficos pueden utilizarse las opciones de giro en 3 dimensiones.

10.2. CORRELACIONES BIVARIADAS

Mediante estos procedimientos se pueden calcular parámetros estadísticos que nos informen de la existencia o no de correlación lineal entre variables.

Para acceder a esta opción hay que elegir los menús *Analizar*, *Correlaciones* y *Bivariadas*, apareciendo el correspondiente cuadro de diálogo. En él podemos seleccionar las variables que queremos analizar. Si elegimos más de dos, SPSS realizará todos los análisis parciales de correlación. Como parámetros de correlación podemos elegir:



10.2.1. COEFICIENTE DE CORRELACIÓN LINEAL DE PEARSON (R):

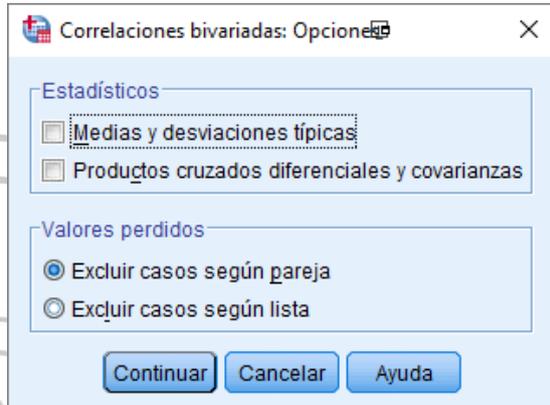
El coeficiente de correlación de Pearson varía entre -1 y $+1$, y es recíproco, es decir la correlación entre X e Y es idéntica a la correlación entre Y e X. Además, es independiente de las unidades de medida de X e Y, y por tanto permite evaluar la correlación entre variables con escalas diferentes. Una correlación positiva ($r > 0$) indica que a aumentos en X corresponden aumentos en Y, y viceversa, mientras que una correlación negativa ($r < 0$) indica que a aumentos en X corresponden disminuciones en Y, y viceversa. Si $r = +1$ ó $r = -1$ la correlación es exacta, y en lugar de una nube de puntos aparecerá una línea recta. Si $r = 0$ quiere decir que no hay asociación lineal entre X e Y. Es importante tener en cuenta que r no mide la magnitud de la asociación, esto habrá de determinarse mediante modelos de regresión lineal. El coeficiente de correlación tampoco nos dice si el modelo de correlación lineal es el más adecuado para

relacionar dos variables, pese a que entre ellas sí exista correlación.

10.2.2. RHO DE SPEARMAN Y TAU-B DE KENDALL.

Se trata de pruebas no paramétricas utilizadas principalmente con variables con categorías ordenadas. Sus valores van igualmente de -1 a $+1$ y su interpretación es equivalente a la del coeficiente de correlación de Pearson.

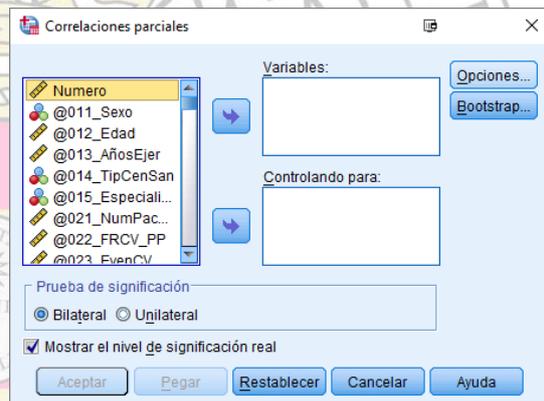
El cuadro dispone de una posibilidad de marcar las correlaciones que resulten estadísticamente significativas. En el botón Opciones también podemos pedir que aparezcan las medias y las desviaciones típicas de las variables, así como los productos cruzados y las covarianzas.



10.3. CORRELACIONES PARCIALES

Este procedimiento calcula los coeficientes de correlación de variables mientras se controlan los efectos de una o más variables adicionales. Se accede eligiendo los menús *Analizar* y *Correlaciones parciales*.

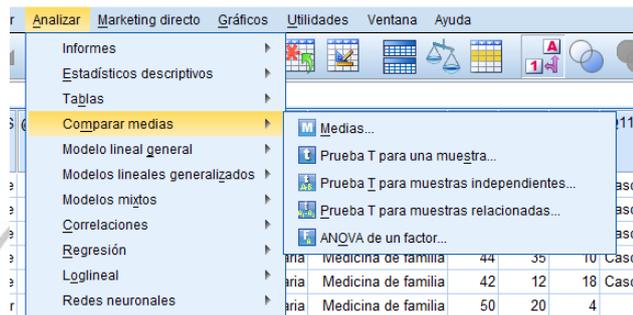
Hay que seleccionar las variables que se quiere analizar y también las variables que se quiere controlar. La ventana de resultados nos dará, en primer lugar, la correlación entre todas las variables utilizadas (si elegimos esta opción en el botón *Opciones*), y después la correlación de las variables elegidas controlando por las variables de control.



El nivel de significación estadística de la correlación sin controlar hay que seleccionarlo directamente si quiere verse en la salida.

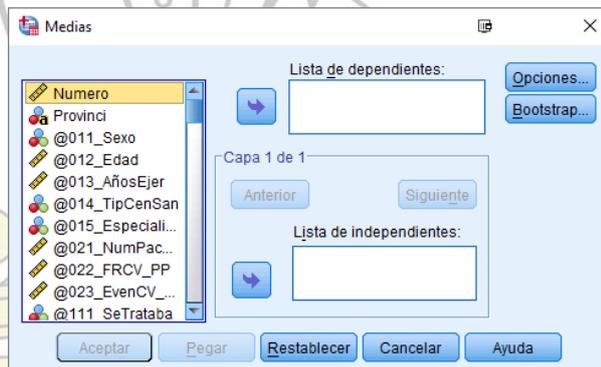
11.COMPARACIÓN DE MEDIAS

Los procedimientos incluidos en el menú *Comparar medias* permiten el cálculo de medias y otros estadísticos, así como la comparación de medias para diferentes tipos de variables, mediante las pruebas *t de Student* y ANOVA.



11.1. MEDIAS

Mediante el procedimiento *Medias* pueden calcularse medias y otros estadísticos para variables dependientes dentro de las categorías de una o varias variables independientes. Además puede realizarse el análisis de la varianza para variables cualitativas de varias categorías. Por lo tanto las variables dependientes o variables resultado serán cuantitativas y las variables independientes, categóricas.

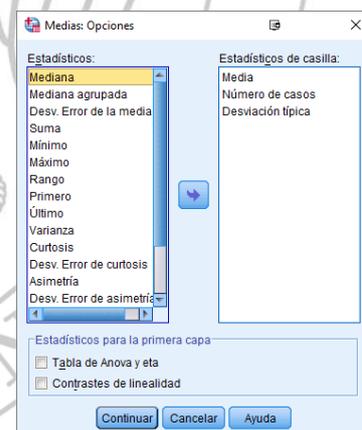


Pueden seleccionarse varias variables dependientes, y también varias independientes de forma que los resultados irán mostrando los valores de las variables resultado para cada variable independiente por separado.

11.1.1. OPCIONES

El subcuadro de diálogo *Medias: Opciones* nos muestra la información adicional que podemos pedir en este procedimiento; se trata de una serie de estadísticos cuyo orden de aparición será el mismo en el que quedan en el cuadro *Estadísticos de casilla*.

Además, puede obtenerse una tabla de análisis de la varianza si la variable independiente tiene más de dos categorías, y parámetros de asociación (η y η^2). Otra opción es la de presentar también parámetros de linealidad como son R y R^2 .



11.2. PRUEBAS *t* DE STUDENT

La prueba *t de Student* es utilizada para la estimación de medias y proporciones en variables cuantitativas y para la comparación de medias y proporciones en distintas poblaciones. El procedimiento *Comparar medias* ofrece varias opciones en función de la procedencia de los datos que queremos comparar.

11.2.1. PRUEBA *T* PARA MUESTRAS INDEPENDIENTES

La *t de Student* para datos independientes se usa para comparar medias de variables en dos

grupos de casos que son independientes entre sí, de forma que los sujetos de cada grupo deberían haber sido elegidos de forma aleatoria.

En caso contrario elegiríamos la opción de *t de Student* para datos apareados o dependientes.

Al hacer la comparación de medias hay que tener en cuenta otro factor, que es la igualdad o no de las varianzas en los dos grupos, pues los resultados no serán iguales en ambos casos. Si asumimos igualdad de varianzas podremos calcular un estimador único y más estable de la varianza poblacional a partir de las varianzas muestrales.



Si las varianzas no son iguales no podremos hacer esto. El procedimiento *Comparar medias* realiza automáticamente un test de hipótesis para la igualdad de las varianzas (test de Levene).

En el cuadro de diálogo podemos seleccionar las variables que queremos analizar, y en función de qué vamos a comparar los grupos, mediante la variable de agrupación, que puede ser cualitativa o cuantitativa. Una vez elegida esta variable hemos de definir los grupos de comparación pulsando el botón *Definir grupos*.

Aquí tenemos dos opciones. Pueden elegirse dos valores especificados para comparar las medias (en general, dos categorías en variables cualitativas), pero también puede establecerse un punto de corte para variables cuantitativas que divida esa variable en dos grupos, por encima y por debajo de ese valor.

Los resultados que SPSS muestra vienen en dos tablas: la primera, *Estadísticos de grupo*, incluye el tamaño muestral, la media, la desviación típica y el error estándar de la media. En la segunda se efectúa la prueba t.

Estadísticos de grupo				
Grupo	N	Media	Desviación tp.	Error tp. de la media
Edad - Hombre	527	51,82	9,031	,393
Edad - Mujer	225	48,62	9,514	,314

Prueba de muestras independientes											
		Prueba de Levene para la igualdad de varianzas				Prueba T para la igualdad de medias					
		F	Sig.	t	df	Sig. (bilateral)	Diferencia de medias	Error tp. de la diferencia	95% intervalo de confianza para la diferencia		
										Inferior	Superior
Edad	Se han asumido varianzas iguales	4,528	,034	7,387	750	,000	5,399	,731	3,965	6,834	
	No se han asumido varianzas iguales			7,234	404,042	,000	5,399	,748	3,932	6,867	

En primer lugar, se muestra el resultado para la prueba de Levene de homogeneidad de varianzas. Si éste es estadísticamente significativo, asumiremos que las varianzas no son homogéneas y entonces tomaremos los datos de la comparación de medias de la fila inferior (*No se han asumido varianzas iguales*). Si el test de Levene no es significativo, asumiremos homogeneidad de varianzas y nos quedaremos con los datos de la fila superior. Los datos que el procedimiento nos muestra son:

- *t*: estadístico utilizado para el contraste de hipótesis.

- *gl*: número de grados de libertad del estadístico t. Cuando se asumen varianzas iguales, *gl* es igual a $(n_1 + n_2 - 2)$, pero cuando las varianzas no son iguales hay que calcularlo de otras formas. La más usada es la aproximación de Satterthwaite, que SPSS calcula. Este valor será menor que el anterior porque se pierde precisión con la desigualdad de las varianzas.
- *Sig. (bilateral)*: valor p de significación estadística obtenido para el contraste de hipótesis.
- *Diferencia de medias*: estimador puntual de la diferencia de medias.
- *Error típ. de la diferencia*: valor del error estándar de la diferencia de medias (obtenido a partir de los errores estándar de la tabla superior).
- *Intervalo de confianza para la media*: límites inferior y superior del intervalo de confianza que valora la precisión de la estimación que estamos realizando para la diferencia de medias.

11.2.2. PRUEBA T PARA MUESTRAS RELACIONADAS

En este caso las muestras no son independientes, sino que están relacionadas. Lo más habitual es que se trate de variables mediadas en los mismos casos en dos momentos temporales diferentes, o que sean casos apareados en función de otras variables.

En el cuadro de diálogo hay que seleccionar el par de variables cuyas medias vamos a comparar, y pasarlo al campo de la derecha *Variables relacionadas*.

Igual que en el caso anterior primero aparece una tabla con los estadísticos de las muestras

y después la prueba *t de Student*. En esta ocasión se tratan los datos como parejas de datos, y se halla la media de las diferencias entre cada pareja, así como su desviación típica para calcular el error estándar de la media. El resto de los valores son equivalentes a los del caso anterior.

The screenshot shows the SPSS 'Prueba T para muestras relacionadas' dialog box. In the 'Variables emparejadas:' table, 'Par 1' is assigned 'alfaA' and 'Variable 2' is assigned 'alfaB'. Below the dialog box, the SPSS output window displays the following data:

Estadísticos de muestras relacionadas				
Par	Media	N	Desviación tp.	Error tp. de la media
Par 1 alfaA	6,837	30	2,0178	,3684
Par 1 alfaB	16,11	30	2,660	,486

Correlaciones de muestras relacionadas			
Par	N	Correlación	Sig.
Par 1 alfaA y alfaB	30	-.195	,301

Prueba de muestras relacionadas						
		Diferencias relacionadas			95% Intervalo de confianza para la diferencia	
Par	Media	Desviación Sp.	Error tp. de la media	Inferior	Superior	t
Par 1 alfaA- alfaB	-9,2733	2,6392	,664	-10,6322	-7,9144	-13,957
						20
						20
						0,000

11.2.3. PRUEBA T PARA UNA MUESTRA

La prueba t para una muestra efectúa un contraste de hipótesis para comprobar si la media de una variable difiere de forma significativa de un valor que nosotros mismos seleccionamos.

Prueba T para una muestra

Variables para contrastar: alfaA

Valor de prueba: 8

Estadísticos para una muestra

	N	Media	Desviación tip.	Error tip. de la media
alfaA	30	6,937	2,0178	,3684

Prueba para una muestra

Valor de prueba = 8

	t	gl	Sig. (bilateral)	95% intervalo de confianza para la diferencia	
				Inferior	Superior
alfaA	-3,158	29	,004	-1,1633	-1,917

En este caso hay que seleccionar la variable elegida y el valor que queremos contrastar. Al pulsar Aceptar se efectuará el contraste de hipótesis.

11.3. ANOVA DE UN FACTOR

Este procedimiento es el equivalente a la *t de Student* para más de dos muestras, es decir, realiza una comparación de medias en variables cuantitativas para más de dos grupos de casos. Se usa para contrastar la hipótesis de que varias medias son iguales. En el correspondiente cuadro de diálogo seleccionaremos la variable cuantitativa que vamos a analizar así como la variable cualitativa de más de dos categorías (*Factor*). Además, tenemos varios botones que nos ofrecen distintas posibilidades.

OPCIONES

Podemos pedir los estadísticos descriptivos para cada grupo de casos, así como la prueba de Levene para la homogeneidad de las varianzas. También puede presentarse un gráfico con las medias para cada grupo.

ANOVA de un factor

Lista de dependientes: VARIABLE

Factor: FX2

ANOVA de un factor

Descriptivos

VARIABLE	N	Media	Desviación típica	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
1	6	63,17	10,572	4,316	52,07	74,26	47	75
2	6	63,00	16,125	6,583	46,08	79,92	40	81
3	6	69,50	18,447	7,531	50,14	88,86	48	90
4	6	57,50	15,834	6,464	40,98	74,12	35	75
Total	24	63,29	15,104	3,883	58,91	69,67	35	90

Prueba de homogeneidad de varianzas

Estadístico de Levene	gl1	gl2	Sig.
1,758	3	20	,188

ANOVA de un factor

VARIABLE	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	433,125	3	144,375	,600	,633
Intra-grupos	4813,833	20	240,692		
Total	5246,958	23			

12. REGRESIÓN LINEAL

Mediante la regresión lineal se busca hallar la línea recta que mejor explica la relación entre unas variables independientes o variables de exposición y una variable dependiente o variable respuesta. Se trata de cuantificar cuánto varía la variable respuesta con cada cambio en las variables independientes. Cuando sólo incluimos en el modelo una variable independiente hablamos de regresión lineal simple, mientras que si hay más de una nos referiremos a regresión lineal múltiple. En los modelos de regresión lineal la variables deben ser cuantitativas y ajustarse a la normal. La estructura del modelo de regresión lineal (ecuación de la recta), por lo tanto, será la siguiente:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + E$$

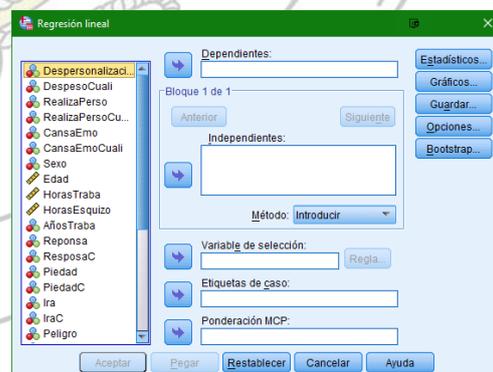
donde Y es la variable dependiente (variable efecto) y $X_1, X_2 \dots X_n$ son las variables independientes y E es una variable "error" que representa el componente de Y que no puede explicarse mediante las variables X. $\beta_0, \beta_1 \dots \beta_n$ son los coeficientes de regresión correspondientes a cada una de las variables independientes.

Para hacer un modelo de regresión lineal hay que establecer unos supuestos que deben cumplirse:

- Independencia: los valores de Y son independientes entre sí.
- Linealidad: la relación entre la variable dependiente y cada variable independiente debe ser lineal.
- Homogeneidad de las varianzas: la varianza de Y es constantes para todos los valores de X.
- Normalidad: para cada valor de la variable independiente, Y sigue una distribución normal.
- Las variables explicativas deben ser linealmente independientes, ninguna puede ser combinación lineal de las otras.

Para crear un modelo de regresión lineal hay que elegir los menús *Analizar, Regresión y Lineal*, apareciendo el cuadro de diálogo correspondiente. Hay que seleccionar una variable dependiente o resultado, y una o más variables independientes o efecto.

Pueden agruparse varias variables independientes en bloques y hacer varios análisis en cada procedimiento.



En el campo *Variable de selección* puede elegirse una variable para hacer el análisis sólo a un subconjunto de casos definido por nosotros mismos, pulsando el botón *Regla*.

Además, puede seleccionarse una variable para etiquetar los casos en función de sus categorías si se realiza algún gráfico.

Con esto, la información que la ventana de resultados ofrece por defecto es la siguiente:

12.1. INTERPRETACIÓN DE LOS COEFICIENTES

Un coeficiente de regresión debe interpretarse como la variación de la variable resultado producida por cada unidad de variación en la variable independiente. Si utilizamos varias variables independientes, entenderemos que esa variación se produce manteniendo constantes el resto de variables, es decir, ajustando por cada una de ellas.

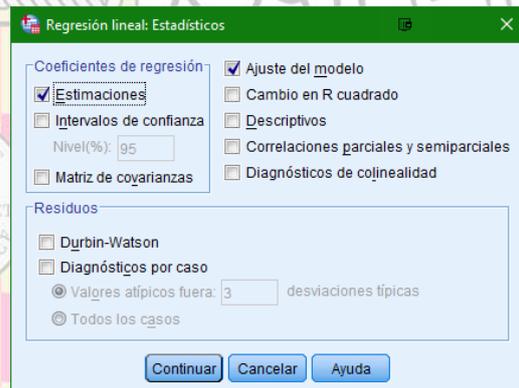
Para cada uno de los coeficientes obtendremos un intervalo de confianza del coeficiente y una significación estadística (probabilidad de que el coeficiente sea igual a 0).

12.2. GRÁFICOS

Pueden crearse gráficos de dispersión a través del botón *Gráficos*. Además, pueden realizarse gráficos de dispersión mediante el procedimiento *Dispersión* del menú *Gráficos*. En él se puede construir el gráfico con la nube de puntos, la recta de ajuste y las bandas de confianza, que serían como los intervalos de confianza de la recta de regresión.

12.3. ESTADÍSTICOS

- Coeficientes de regresión: la opción *Estimaciones* muestra los estadísticos que incluye por defecto. *Intervalos de confianza* muestra intervalos de confianza para coeficiente de regresión. *Matriz de covarianza* muestra una tabla con las correlaciones y covarianzas de las variables exposición.
- Ajuste del modelo: muestra las tablas con el coeficiente de determinación y el análisis de la varianza.
- Descriptivos: muestra el tamaño muestral, media y desviación típica y una matriz de correlaciones.



12.4. OPCIONES

- Criterios del método por pasos: establece las condiciones para que una variable sea incluida o eliminada del modelo cuando la selección de variables se realiza hacia adelante, hacia atrás o por pasos.
- Incluir la constante: puede no incluir la constante o coeficiente β_0 en el análisis.

