

Tutorial de *SPSS Statistics*

Josep M. Domènech y Roberto Sesma

Laboratori d'Estadística Aplicada i de Modelització
Universitat Autònoma de Barcelona

Tabla de contenidos

Presentación.....	3
Preparación de los ordenadores	3
1. Estudio sobre hábitos de salud.....	5
1.1 Preparación de los datos	5
1.2 Cómo se introducen los datos de un estudio.....	6
1.3 Valores desconocidos (<i>Missing values</i>) y Valores no aplicables.....	6
1.4 Estructura de los datos de un estudio: matriz de datos y clasificación de las variables según la escala de medida	7
2. El Sistema <i>SPSS Statistics</i>	8
2.1 Menú ayuda	8
2.2 Estudios de casos.....	8
2.3 Ventanas de <i>SPSS Statistics</i>	9
2.4 Cuadros de diálogo de <i>SPSS Statistics</i>	10
2.5 Modos de trabajo con <i>SPSS Statistics</i>	10
2.6 Propiedades de las variables.....	11
Tabla con las propiedades de las variables del estudio.....	12
2.7 Esquema de un proceso de datos con <i>SPSS Statistics</i>	13
2.8 Configuración de <i>SPSS Statistics</i>	14
3. Lectura de los datos	17
3.1 Introducción de los datos almacenados en un fichero	17
3.2 Definición de las propiedades de las variables	18
3.3 Final de una sesión de trabajo: guardar las ventanas de datos y sintaxis	23
4. Creación de nuevas variables.....	25
4.1 Nueva sesión de trabajo: abrir las ventanas de datos y de sintaxis	25
4.2 Calcular (COMPUTE).....	26
Operaciones aritméticas: crear IMC	26
Operaciones aritméticas: crear CProv.....	27
Cálculo del tiempo transcurrido a partir de dos fechas	28

4.3 Calcular (COMPUTE) con expresiones lógicas	30
Expresiones lógicas	30
Creación de una variable binaria: procedimiento de elección.....	30
Creación de la variable Hipertensión: uso del operador NOT para definir una expresión lógica complementaria.....	32
4.4 Calcular Si (IF).....	33
4.5 Recodificar (RECODE).....	35
Recodificación de una variable cadena en una numérica.....	35
Recodificación de una variable continua con valores discretos	37
Recodificación de una variable continua con valores continuos.....	39
4.6 Contar (COUNT).....	40
4.7 Construcción del indicador de salud.....	42
4.8 Final de la etapa de creación de variables: guardar las ventanas de datos y sintaxis	47
4.9 Comprobación de la sintaxis creada	48
5. Ordenación y Selección de registros	49
5.1 Ordenación de registros.....	49
5.2 Selección de registros	51
Selección temporal	51
Selección permanente	53
6. Descripción de datos	55
6.1 Clasificación de los procedimientos descriptivos de <i>SPSS Statistics</i>	55
6.2 Procedimiento FREQUENCIES	55
Descripción de una variable categórica ordinal	55
Descripción de una variable cuantitativa	56
Descripción de variables que contengan tiempos cumplidos	57
6.3 Procedimiento DESCRIPTIVES	58
6.4 Procedimiento MEANS.....	59
6.5 Procedimiento EXAMINE	60
6.6 Procedimiento CROSSTABS	62
6.7 Copiar las tablas de la ventana de resultados en un documento Word	63
6.8 Final de la sesión: guardar las ventanas de sintaxis y resultados	64
7. Edición de Resultados	65
7.1 Manipulación de tablas pivote y objetos gráficos.....	65
7.2 Edición de una tabla pivote	66
7.3 Edición de gráficos.....	68

Presentación

El objetivo de este Tutorial es aprender a utilizar el programa *SPSS Statistics*, que es un programa de uso universal para el análisis estadístico de datos, a través de un trabajo dirigido que consiste en leer los datos de un hipotético estudio sobre **Hábitos de Salud**, guardados en el archivo **Salud.xls**, definir las propiedades de las variables, crear nuevas variables para contestar las hipótesis del estudio y finalmente realizar un breve estudio descriptivo que comporta.

Para trabajar con este Tutorial es **imprescindible imprimirlo** y, con el ordenador al lado, debe ir realizando las tareas que se proponen; el PDF está preparado para imprimirlo a doble cara. Tenga en cuenta que las imágenes pueden variar ligeramente en función de la versión de *SPSS Statistics* y de los programas y dispositivos instalados en el ordenador.

Los capítulos 1 y 2 son teóricos; contienen conceptos y terminología básica de proceso de datos y una breve presentación del *SPSS Statistics*. Las **tareas** que se proponen empiezan en el apartado 2.8 Configuración de *SPSS Statistics* (p. 14).

El Tutorial se ha preparado con la versión 18 del programa, pero puede realizarse con versiones más recientes ya que la forma de realizar los ejercicios es idéntica. La única y **curiosa diferencia** es que, en lugar de *SPSS Statistics*, en las ventanas aparece *PASW Statistics*. Esto es debido a que en la versión 18 la compañía SPSS decidió cambiar el nombre SPSS por PASW (*Predictive Analytics Software Statistics*), pero unos meses más tarde IBM compró la empresa y el programa volvió a recuperar el nombre de *SPSS*.

Para poder realizar este tutorial es necesario tener **conocimientos básicos de Windows**.

Preparación de los ordenadores

El Tutorial está formado por un documento y el archivo de datos **Salud.xls**. En primer lugar se debe **crear una nueva carpeta** con el nombre **Tutorial SPSS** y dentro de ella copiaremos el archivo de datos. Los archivos que se crearán a lo largo del Tutorial se irán guardando en esta carpeta.

Puesto que este Tutorial trabaja con diferentes tipos de archivos (Excel, Datos, Sintaxis y Resultados SPSS) y algunos iconos que los identifican son bastante similares, se recomienda que Windows enseñe la extensión del archivo. Para ello basta abrir cualquier carpeta, por ejemplo, la carpeta **Tutorial SPSS** del escritorio, y elegir el menú de la carpeta **Herramientas**, luego **Opciones de carpeta...** y finalmente la ficha **Ver**:



En todo el Tutorial para indicar una secuencia se escriben las diferentes opciones del menú seguidas del signo |.

Así, por ejemplo, para expresar la secuencia anterior escribiremos: **Herramientas | Opciones de carpeta... | Ver**

En la ficha **Ver** se **desactiva la casilla**: ☐ Ocultar extensiones para los tipos de archivos conocidos y se sale con el botón **Aceptar**.

Después de haber realizado esta acción veremos los nombres de los archivos con su extensión, de manera que en el archivo Excel de datos en lugar de ver **Salud** veremos **Salud.xls**.

1. Estudio sobre hábitos de salud

Se ha realizado un estudio sobre hábitos de salud administrando el siguiente cuestionario a una muestra de 52 sujetos. Para preparar el cuestionario se han seguido las siguientes **reglas de uso universal**:

- **En un estudio sólo se registran variables originales, nunca variables calculadas.** Así, si se estudia la obesidad introduciremos el peso y la talla, **nunca** el índice de masa corporal ni la variable binaria obeso / no obeso, porque estas variables las puede crear *SPSS Statistics* a partir del peso y talla de cada sujeto.
- **En un estudio conviene registrar fechas, nunca edades ni tiempos transcurridos.** Así, es mejor registrar la fecha de nacimiento de los sujetos y las fechas en que se producen los acontecimientos que nos interesan, porque *SPSS Statistics* permite calcular de forma exacta los tiempos transcurridos entre las diferentes fechas. En este cuestionario se recoge sólo la fecha de nacimiento porque fue contestado el día **14 de abril de 2006**.

Cuestionario de salud				Número: _____
Sexo: () Masculino () Femenino Peso: ___,_ kg Talla: __ cm Fecha nacimiento: __/__/____				
Código postal: _____ Presión arterial sistólica: _____ mmHg Presión arterial diastólica: _____ mmHg				
¿Fuma actualmente o ha fumado? () Nunca () Ex fumador () Fumador				
Si fuma o ha fumado ¿a qué edad empezó? ____ años () No recuerdo				
Si fuma ¿cuántos cigarrillos fuma en promedio cada día? ____ c/d () NS/NC				
Hábitos positivos	Nunca	Ocasional	Habitual	
H1: Práctica deportiva	()	()	()	
H2: Dieta equilibrada	()	()	()	
H3: Descanso regular	()	()	()	
Hábitos negativos	Nulo / Bajo	Medio	Alto	
H4: Consumo de alcohol	()	()	()	
H5: Consumo de caféina	()	()	()	

1.1 Preparación de los datos

Una vez recogidos todos los datos de este estudio se han introducido en el archivo **Salud.xls** teniendo en cuenta las siguientes reglas:

- **Las variables categóricas se entran codificadas, nunca el texto de las categorías.** Así, si el Sexo se registra como “Femenino/Masculino” en lugar de estos textos deben introducirse códigos más sencillos, por ejemplo 0/1 ó F/M, aunque es preferible **utilizar siempre códigos numéricos** porque facilita el posterior tratamiento de las variables. Las respuestas a los ítems sobre hábitos positivos se han codificado como: Nunca=0, Ocasional=1, Habitual=2, y sobre los hábitos negativos se han codificado como Bajo=1, Medio=2, Alto=3.
- Los códigos adecuados para las **variables binarias** (No/Sí, Femenino/Masculino...) son los valores 0/1.
- En un estudio es frecuente encontrar sujetos con valores desconocidos en alguna variable. En este caso la regla es **dejar el campo vacío** porque el programa de análisis estadístico lo interpreta automáticamente como un valor desconocido y le asigna el código interno de **system missing** que se visualiza con un punto (.).

La siguiente imagen enseña el archivo **Salud.xls** con los datos de este estudio entrados con la codificación expuesta. En la primera fila **se deben situar los nombres de las variables** que utilizará *SPSS Statistics* y las respuestas de cada sujeto se introducen en una de las siguientes filas.

Nombres de las variables (Véase p.11)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	Caso	Sex	Peso	Talla	FN	CP	PAS	PAD	Fuma	Edad	F	Tab	H1	H2	H3	H4	H5
2	79	F	64,7	169	07/04/1935	8430	120	70	2	16	20	2	0	2	3	3	
3	84	M	101,5	170	04/11/1933	8080	130	92				0	1	0		1	
4	91		55,0		27/07/1937	8190	140	86	0		0	2		2	1	1	
5	92	F	66,7	175	17/06/1949	8401	150		2		99						
6	87	M	75,5	178		88788	188	88	0	22	15	0	1	0	0	0	

Salud

1.2 Cómo se introducen los datos de un estudio

El procedimiento de elección es utilizar una **base de datos**, ya que permite estructurar los datos de un estudio en forma de tablas relacionadas y definir formularios para capturar datos de forma fácil y segura ya que se pueden incluir comprobaciones sobre cada dato en el mismo momento en que se ha introducido (validaciones *on-line*). La más utilizada, para estructuras de datos cuyos diseños no sean excesivamente complejos, es Microsoft Office **Access** (tiene la limitación de un máximo de 256 campos por tabla). No obstante, por su sencillez de uso, lo más habitual es introducir los datos en Microsoft Office **Excel**.

Una vez introducidos los datos, si no han sido capturados con validaciones, es imprescindible someterlos a un **proceso de depuración** que incluya comprobaciones de rango y lógicas. El mayor nivel de calidad se alcanza cuando, además, se han capturado con un **proceso de doble entrada** (un segundo operador, diferente del primero, introduce de nuevo los mismos datos en modo verificación y se corrigen las discrepancias).

Los datos del archivo **Salud.xls** han sido sometidos a un proceso de depuración y están libres de errores.

1.3 Valores desconocidos (*Missing values*) y Valores no aplicables

Los valores que faltan en la matriz de datos pueden ser **valores desconocidos** (*missing*) o **valores no aplicables**. El siguiente cuadro presenta la parte de un cuestionario en el que pueden darse estos dos tipos de valores.

¿Fuma actualmente o ha sido fumador? () Nunca () Ex fumador () Fumador		Pregunta de cribado ← Códigos de valores desconocidos (<i>user missing</i>)
Si fuma o ha fumado: ¿A qué edad comenzó a fumar? ____ años		
Si fuma actualmente: ¿Cuánto fuma en un día? ____ c/d		
No recuerda (codificar 98) No desea contestar (codificar 99)		

Valores no aplicables

Las preguntas que sólo pueden ser contestadas por una parte de los sujetos del estudio generan **valores no aplicables**. En el cuadro, si un sujeto nunca ha fumado se saltan las preguntas sobre tabaquismo, y el valor de la *edad en que empezó a fumar* y el del *consumo de tabaco* quedarán vacíos: En este caso no se trata de **valores desconocidos** (*missing*) sino de **valores no aplicables**; sin embargo *SPSS Statistics* no dispone de un código especial que permita distinguir estos dos importantes tipos de vacíos y en ambos casos los considera *system missing*.

Valores desconocidos (*missing*)

Cuando no se conoce el valor de una variable lo recomendable es **dejar el campo vacío** ya que *SPSS Statistics* lo interpreta automáticamente como un valor desconocido y le asigna el código interno de **system missing** (perdido por el sistema) que se visualiza como un punto (.) en la ventana de datos.

En unas pocas ocasiones interesa poner en el campo un código que indique los motivos de no conocer su valor; en el cuadro anterior si el entrevistador olvida formular la pregunta el valor de la edad quedará vacío, pero si el sujeto no recuerda *a qué edad comenzó a fumar* se pone el valor 98 (que en este ejemplo corresponde al código “No recuerda”) mientras que si no quiere contestar se pone el valor 99 (código de “No contesta”). Para que *SPSS Statistics* reconozca los códigos (98, 99, etc.) como diferentes tipos de valores desconocidos de una determinada variable (y los excluya de los cálculos estadísticos) deberán ser declarados **user missing**.

En la siguiente tabla de datos hay señalados algunos de estos valores.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Caso	Sex	Peso	Talla	FN	CP	PAD	PAS	Fuma	EdadF	Tab	H1	H2	H3	H4	H5
2	79	F	64,7	169	07/04/1935	8430	70	120	2	16	20	2	0	2	3	3
3	84	M	101,5	170	04/11/1933	8080	92	130				0	1	0		1
4	91		55,0		27/07/1937	8190	86	140	0		0	2		2	1	1
5	92	F	66,7	175	17/06/1949	8401		150	2		99					
6	97	M	75,5	170		29700	80	126	2	99	15	0	1	0	3	2
7	100	F	74,0	175	17/06/1949	85000	100		1	16	0	0	1	0	1	0



Variable Tab (consumo de tabaco): Lo usual es que si un sujeto manifiesta que no fuma (variable de cribado Fuma=0) se salta a la siguiente sección y todas las variables sobre tabaco quedan vacías (valor “no aplicable”). Sin embargo, en este estudio, a los sujetos que no fuman se les ha asignado el **valor deducible** Tab=0.

1.4 Estructura de los datos de un estudio: matriz de datos y clasificación de las variables según la escala de medida

Los datos de cualquier estudio se recogen, siguiendo un determinado protocolo, mediante formularios o cuestionarios, que permiten registrar la información de cada uno de los sujetos de forma estandarizada y estructurada.

La **matriz de datos** es una estructura en forma de tabla que contiene los valores de los sujetos en las diferentes variables. Las **filas** de esta tabla representan a cada sujeto y las **columnas** representan las diferentes variables.

Se entiende por variable cada uno de los caracteres o aspectos que se registran en los sujetos del estudio.

Variable Edad a la que empezó a fumar ↓															
Caso	Sex	Peso	Talla	FN	CP	PAD	PAS	Fuma	EdadF	Tab	H1	H2	H3	H4	H5
79	F	64,7	169	07/04/1935	8430	70	120	2	16	20	2	0	2	3	3
84	M	101,5	170	04/11/1933	8080	92	130				0	1	0		1
Sujeto 91 → 91		55,0		27/07/1937	8190	86	140	0		0	2		2	1	1
92	F	66,7	175	17/06/1949	8401		150	2		99					
97	M	75,5	170		29700	80	126	2	99	15	0	1	0	3	2
100	F	71,0	175	17/02/1940	85000	100		1	10	0	0	1	0	1	0

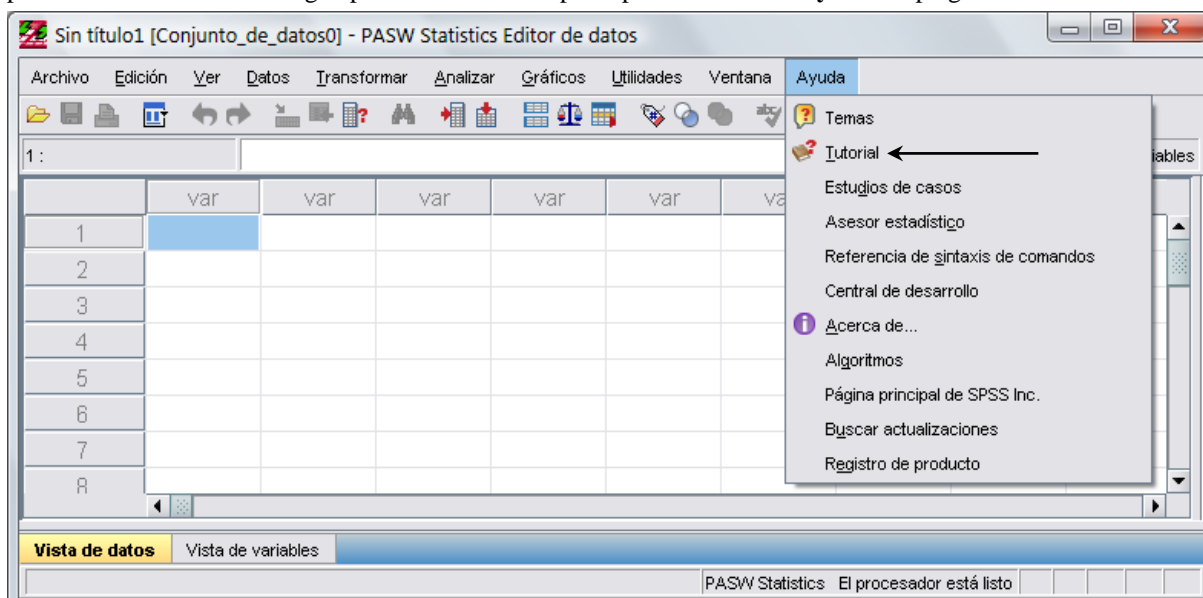
La estadística clasifica las variables en dos grandes grupos según se registren **datos no métricos** (variables categóricas) o **datos métricos** (variables cuantitativas). Dentro de las variables categóricas se distinguen dos tipos según tengan o no orden las diferentes categorías de la variable. Dentro de las variables cuantitativas se distinguen las **discretas** (que siempre son el resultado de un recuento) de las **continuas** (que son medidas que pueden tomar como valor un número real).

Esta clasificación asociada a la escala de medida de las variables, se ilustra en el cuadro siguiente y es muy importante porque las técnicas estadísticas utilizadas para analizar los datos varían según la **escala de medida** de las variables implicadas.

	Tipo de variable y escala de medida	Ejemplos
Datos no métricos (categóricos)	Categórica (escala nominal)	
	Binaria (2 categorías)	Sexo (Femenino/Masculino)
	Con más de 2 categorías	Tratamiento (A, B, C); Código postal
	Categorías ordenadas (escala ordinal)	Práctica deportiva (Nunca/Ocasional/Habitual) Nivel socioeconómico
Datos métricos (cuantitativos)	Cuantitativa (escala de intervalo o razón)	
	Discreta (recuentos)	Número de síntomas, Número de hijos ...
	Continua (medidas)	Edad, Peso, Talla ...

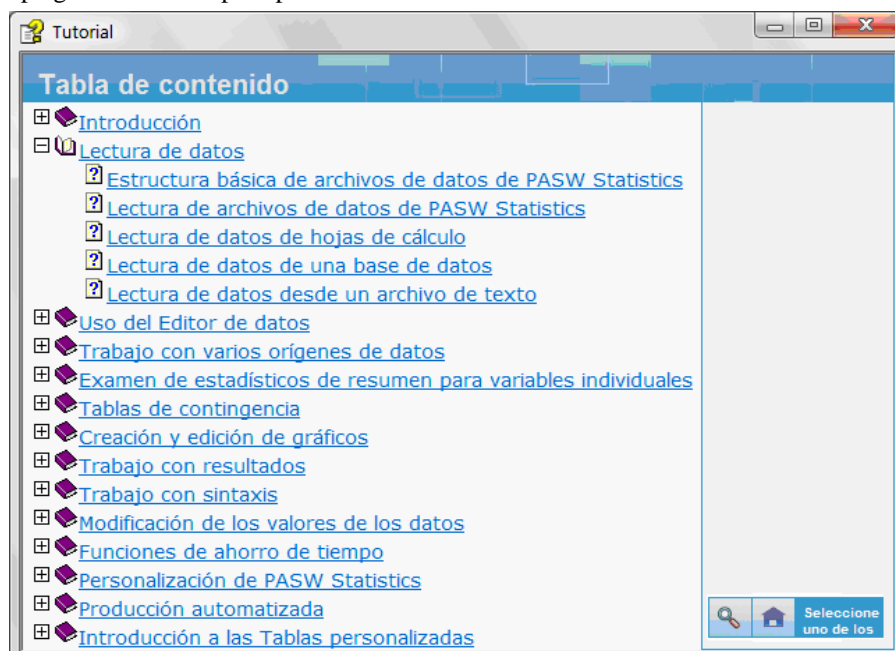
2. El Sistema *SPSS Statistics*

SPSS Statistics es la nueva denominación del clásico SPSS (*Statistical Package for Social Sciences*) que en 2009 fue adquirido por IBM. Es un programa de uso universal para el análisis estadístico de datos. Permite gestionar archivos de datos, crear y modificar variables, generar listados de sujetos, producir tablas y gráficos y realizar pruebas estadísticas. La imagen presenta la ventana principal con el menú **Ayuda** desplegado:



2.1 Menú ayuda

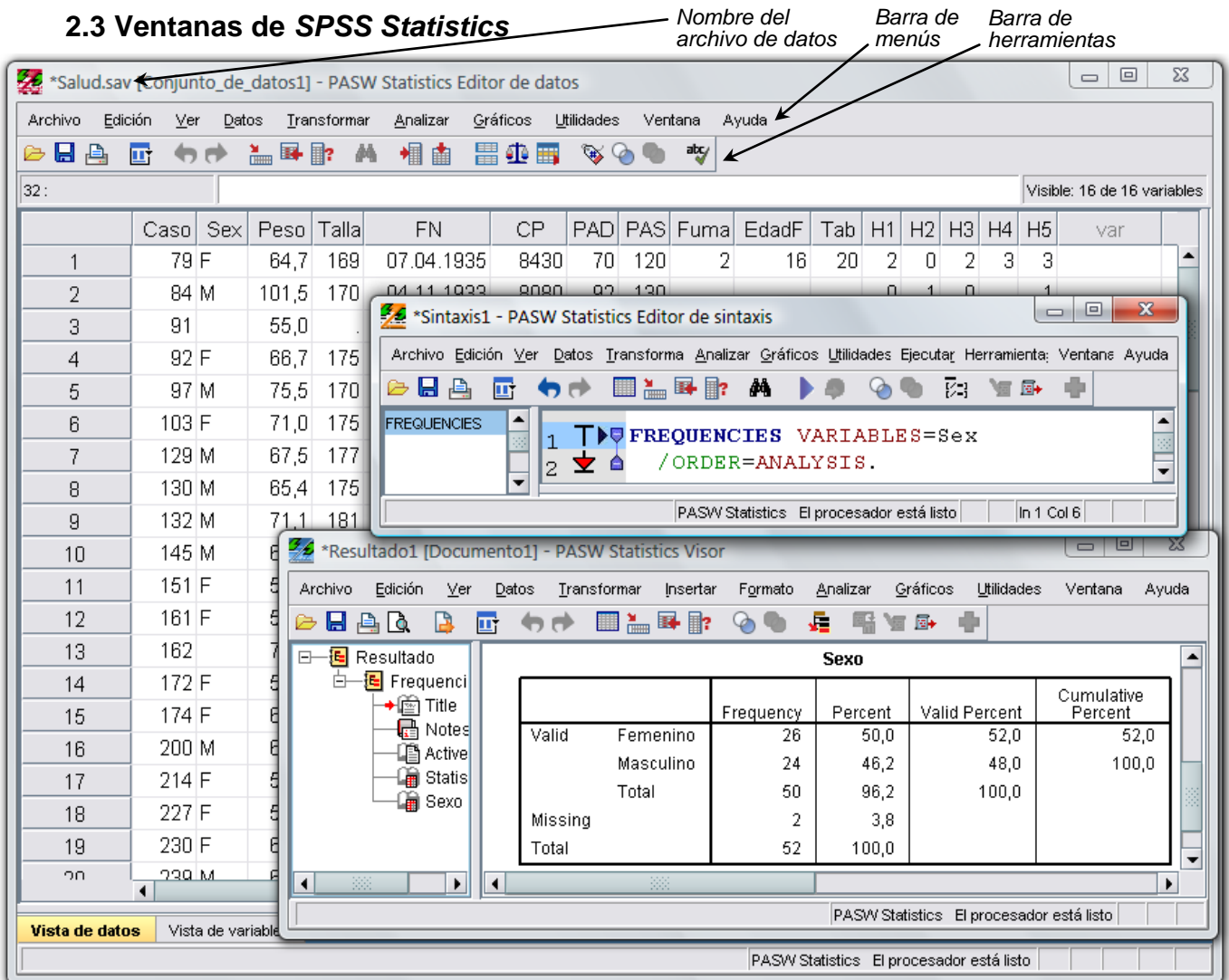
Una vez abierto el programa, el menú **Ayuda | Tutorial** permite acceder a un completo Tutorial estructurado por temas (que no es el Tutorial que está realizando ahora). Seguir los diferentes tutoriales que incorpora el programa es una manera rápida para aprender de forma visual y guiada los aspectos más importantes para trabajar con *SPSS Statistics*. La imagen tiene desplegado el tutorial por aprender a leer los datos de un estudio:



2.2 Estudios de casos

Es un conjunto de tutoriales en inglés que proporcionan ejemplos detallados de todos los procedimientos estadísticos implementados en los diferentes módulos de *SPSS Statistics*. Se enseña a realizar los análisis estadísticos y a interpretar los resultados. Estos ejemplos asumen que se está familiarizado con las operaciones básicas de *SPSS Statistics* que se explican en los temas del **Tutorial**.

2.3 Ventanas de SPSS Statistics




Si se abre **SPSS Statistics** haciendo doble clic sobre su icono aparece la ventana principal vacía con el título: “Sin Título – Editor de datos”. Si se accede haciendo doble clic sobre un archivo de datos en formato **SPSS Statistics** (p.e. Salud.sav) en el título aparece el nombre del fichero. En ella se encuentra:

La barra de menús: Así, si se escoge **Archivo**, se abrirá un menú con varias opciones, entre ellas la opción **Salir**. Así, por ejemplo, para salir del programa se selecciona **Archivo | Salir**.

La barra de herramientas: Permite un acceso rápido a las opciones más utilizadas de los menús.


Tabla de datos: Cada fila recoge los datos de un sujeto y cada columna los de una variable. Al salir de esta ventana preguntará si queremos guardar el contenido de la ventana de datos en un fichero. Los archivos de datos tienen extensión **.SAV**.

 **SPSS Statistics** permite tener abiertas simultáneamente varias ventanas de datos.

Ventanas secundarias:

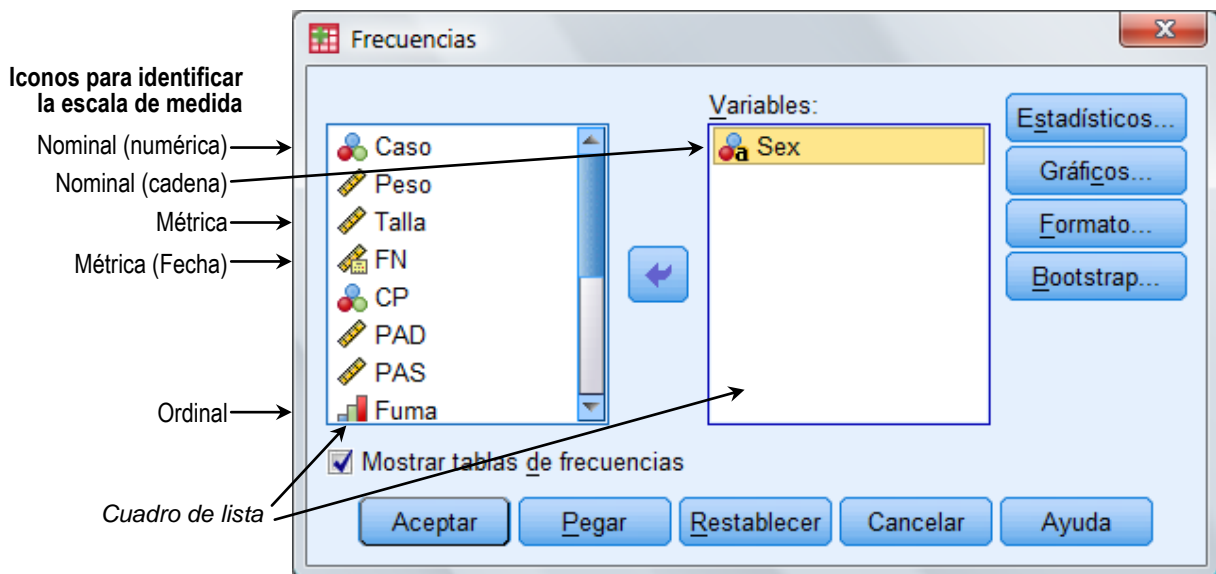
Resultados1-Visor: Esta ventana es dónde el programa enviará las salidas (resultados y mensajes de error). Los archivos de resultados tienen extensión **.SPV**.

Sintaxis1- Editor de sintaxis: Esta ventana permite guardar las instrucciones que se vayan ejecutando. Los archivos de sintaxis tienen extensión **.SPS**.

 Las diferentes ventanas se iconizan en la barra de tareas de Windows, como sucede con la mayoría de aplicaciones Windows.

Si al salir del programa no se guarda el contenido de estas ventanas la información se pierde y no se puede recuperar.

2.4 Cuadros de diálogo de SPSS Statistics




La mayor parte de los cuadros de diálogo contienen un conjunto de botones de comando y cuadros de lista. Los botones que generalmente aparecen en todos los cuadros son:

- Aceptar:** Ejecuta las opciones del cuadro de diálogo que se han seleccionado.
- Pegar:** Inserta la sintaxis del procedimiento generado por este cuadro en la ventana de sintaxis.
- Restablecer:** Elimina las especificaciones introducidas la última vez que se utilizó el cuadro de diálogo.
- Cancelar:** Cierra el cuadro de diálogo sin completar el procedimiento.
- Ayuda:** Abre una ventana con ayuda específica sobre operaciones a realizar.

Los puntos suspensivos (...) a continuación del texto de un botón de comando indican que al pulsar ese botón aparecerá un nuevo cuadro de diálogo.

Cuadros de lista

Suelen contener listas de archivos o listas de variables. Existen dos tipos de cuadros de lista de variables: los que contienen todas las variables del archivo de datos y los que contienen las variables que se han seleccionado.

Las variables se seleccionan en el primer cuadro y se copian en el segundo con el botón  o viceversa. También se mueven de un cuadro a otro haciendo doble clic sobre el nombre de la variable o arrastrando.

Los nombres de las variables van precedidos por un icono que identifica la escala de medida. El icono incorpora la letra **a** cuando tienen formato cadena.


Las opciones de los cuadros de diálogo aparecen atenuadas cuando carece de sentido seleccionarl

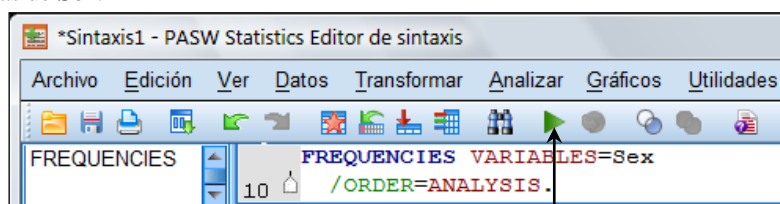
2.5 Modos de trabajo con SPSS Statistics

El programa permite trabajar con el sistema de **Menús**. Así, si en el cuadro de diálogo anterior se sale pulsando el botón **Aceptar** se genera automáticamente la distribución de la variable **Sex** en la ventana de resultados.

Sin embargo, el **uso profesional** del programa **requiere trabajar con sintaxis**. Puesto que la mayor parte (pero no todos) los procedimientos se pueden realizar a través de cuadros de dialogo, para evitar tener que escribir la sintaxis, en este Tutorial se enseña a trabajar en **modo mixto**: se trabaja con Menús y se sale del cuadro de diálogo con el botón **Pegar** que escribe automáticamente la sintaxis del procedimiento en la ventana de sintaxis.

Si se sale del anterior cuadro con el botón **Pegar**, al final de la ventana de sintaxis se añade el siguiente comando que generará la distribución de frecuencias de **Sex**.

Seleccionando esta sintaxis y pulsando el botón  se genera la distribución de frecuencias de la página siguiente.



Botón para ejecutar la sintaxis

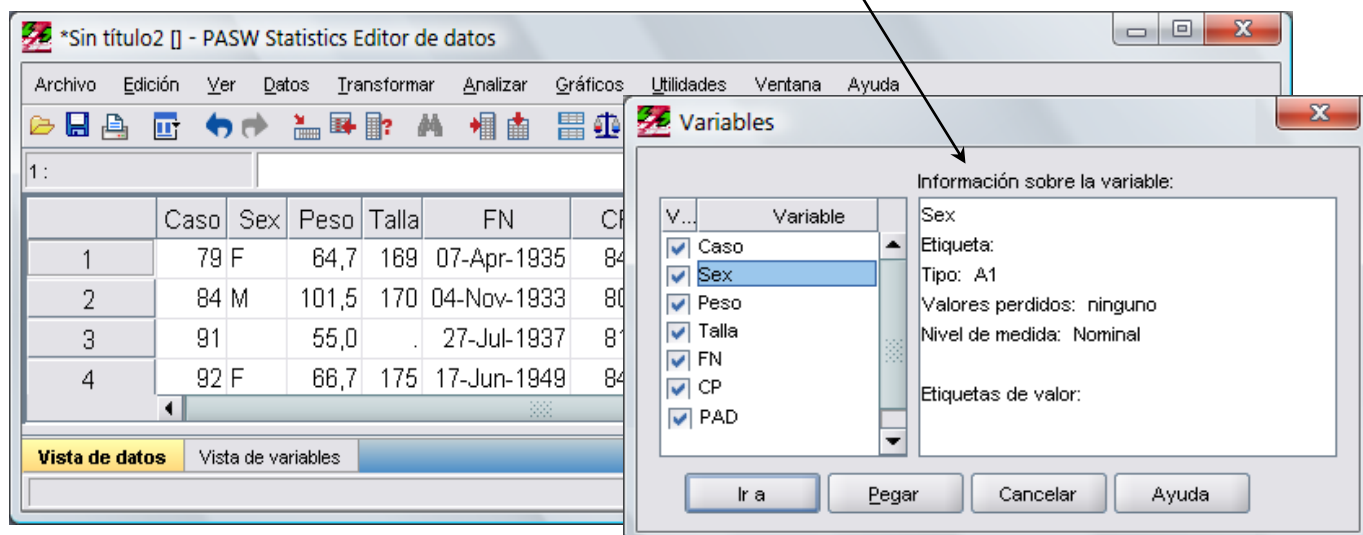
2.6 Propiedades de las variables

Los programas de análisis estadístico, además de los datos, tienen asociado a cada variable un diccionario que contiene una serie de propiedades cuyo principal cometido es presentar con un formato adecuado las tablas y gráficos resultantes de los análisis.

La ventana siguiente presenta los datos inmediatamente después de haber sido leídos desde el Excel.

Para ver el diccionario se elige la opción de menú **Utilidades | Variables**.

Si nos situamos sobre la variable **Sex** aparece el diccionario casi vacío:



Si a continuación se pide la distribución de frecuencias de la variable **Sex** se obtiene la siguiente tabla en la que los 2 casos sin valor en esta variable aparecen como válidos y sin ninguna etiqueta descriptiva del contenido de la variable ni de sus categorías:

Secuencia de menús para obtener este resultado

Descripción antes
de definir las propiedades
de la variable Sex

Analizar | Estadísticos descriptivos | Frecuencias... | Variables: ► Sex

Sex				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2	3,8	3,8	3,8
F	26	50,0	50,0	53,8
M	24	46,2	46,2	100,0
Total	52	100,0	100,0	

Para disponer de una tabla correcta se deben definir las propiedades de esta variable, entrando en el diccionario la etiqueta de nombre "Sexo", las etiquetas de valores "Masculino" y "Femenino" y definiendo el código blanco como valor perdido (*missing*). Si se pide de nuevo la distribución de frecuencias de la variable Sex se obtiene la siguiente tabla que se lee sin dificultad porque incorpora las etiquetas e identifica a los 2 sujetos sin valor como valores desconocidos.

Descripción después
de definir las propiedades
de la variable Sex

Etiqueta de variable

Etiquetas de valores

Valores desconocidos

Analizar | Estadísticos descriptivos | Frecuencias... | Variables: ► Sex

Sexo					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Femenino	26	50,0	52,0	52,0
	Masculino	24	46,2	48,0	100,0
	Total	50	96,2	100,0	
Missing		2	3,8		
Total		52	100,0		

Las principales propiedades que se pueden definir para cada variable son las siguientes:

Nombre	Debe empezar con una letra (A - Z) o @, y puede tener hasta 64 caracteres entre letras, cifras y los signos: punto (.), @, \$ y guión bajo (_).
Etiqueta del nombre	Breve descripción del contenido de la variable. Si la variable es cuantitativa esta descripción debe incluir las unidades de medida.
Etiquetas de los valores (sólo para variables categóricas)	Breve descripción de los códigos que definen cada una de las categorías. Es conveniente asignar etiquetas a todas las variables categóricas.
Escala de medida	Categórica nominal, categórica ordinal o cuantitativa (escala).
Tipo	Númérico y Cadena (<i>string</i>) también llamada Alfanumérica. El tipo Fecha es una variable numérica que tiene un formato de tiempo.
Valores desconocidos (<i>user missing</i>)	Códigos asignados a los valores desconocidos definidos por el usuario. Nota: Los valores <i>system missing</i> no hay que definirlos porque se reconocen de forma automática.

Ejemplos

<i>Nombre:</i>	Sex	<i>Nombre:</i>	Tab
<i>Etiqueta de variable:</i>	Sexo	<i>Etiqueta de variable:</i>	Consumo de tabaco (c/d)
<i>Etiqueta de valores:</i>	M= Masculino F= Femenino	<i>Formato:</i>	Númérico (0 dec.)
<i>Formato:</i>	Cadena (1 carácter)	<i>Escala de medida:</i>	Métrica
<i>Escala de medida:</i>	Nominal	<i>Valores desconocidos:</i>	99
<i>Valores desconocidos:</i>	' ' (blanco)		

Tabla con las propiedades de las variables del estudio

Nombre variable	Tipo y Formato	Missing usuario	Escala medida	Etiqueta de la variable	Etiquetas de los códigos
Caso	Num. (0d)		Nominal		
Sex	Cadena	<i>blanco</i>	Nominal	Sexo	M= Masculino F = Femenino
Peso	Num. (1d)		Métrica	Peso (kg)	
Talla	Num. (0d)		Métrica	Talla (cm)	
FN	Fecha eur.		“Métrica”	Fecha de nacimiento	
CP	Num. (0d)		Nominal	Código postal	
PAD	Num. (0d)		Métrica	Presión arterial diastólica (mmHg)	
PAS	Num. (0d)		Métrica	Presión arterial sistólica (mmHg)	
Fuma	Num. (0d)		Ordinal	¿Fuma o ha fumado?	0= No fumador 1= Ex fumador 2= Fumador
EdadF	Num. (0d)	99	Métrica	Edad a la que empezó a fumar (años)	99=No recuerda 99=No contesta
Tab	Num. (0d)	99	Métrica	Tabaco (c/d)	
H1	Num. (0d)		Ordinal	Práctica deportiva	0= Nunca; 1= Ocasional 2= Habitual
H2	Num. (0d)		Ordinal	Dieta equilibrada	
H3	Num. (0d)		Ordinal	Descanso regular	
H4	Num. (0d)		Ordinal	Consumo de alcohol	1= Bajo; 2= Medio
H5	Num. (0d)		Ordinal	Consumo de cafeína	3= Alto

2.7 Esquema de un proceso de datos con SPSS Statistics

Fase 1: Preparación de la matriz de datos

Lectura de datos →

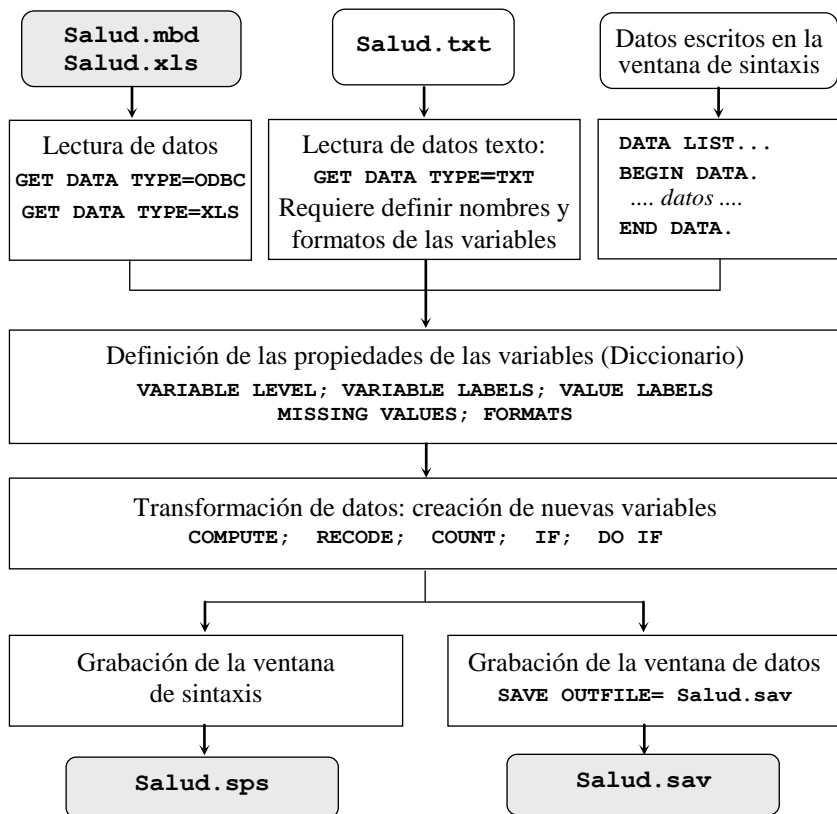
! Lo habitual es tener los datos en Access o Excel

Definición de las propiedades de las variables originales →

Creación de nuevas variables necesarias para comprobar las hipótesis del estudio →

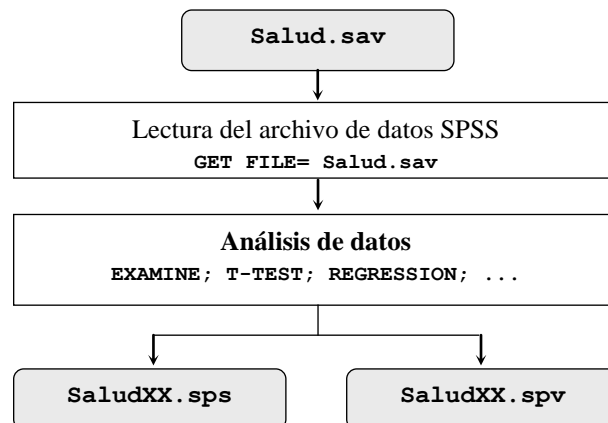
Grabación de la sintaxis y de los datos →

! El archivo de sintaxis es imprescindible para crear nuevos archivos SPSS de datos con más sujetos o con sujetos diferentes



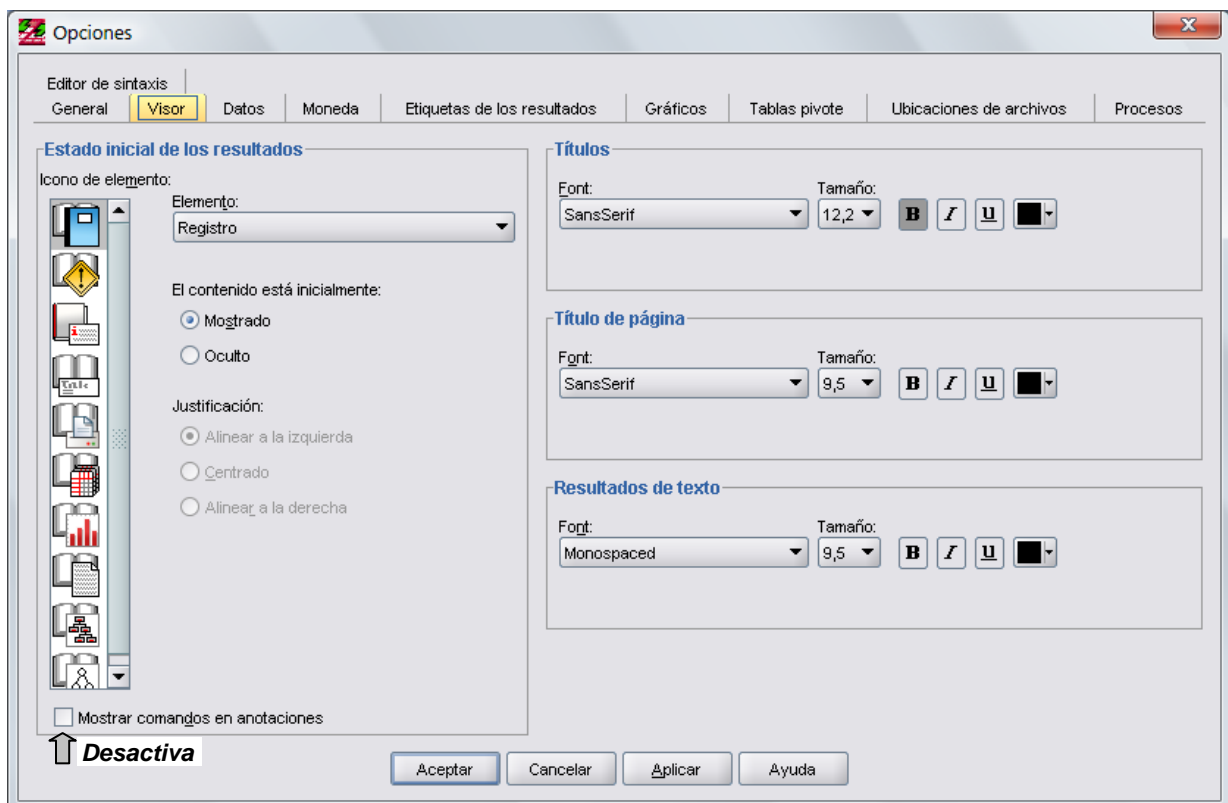
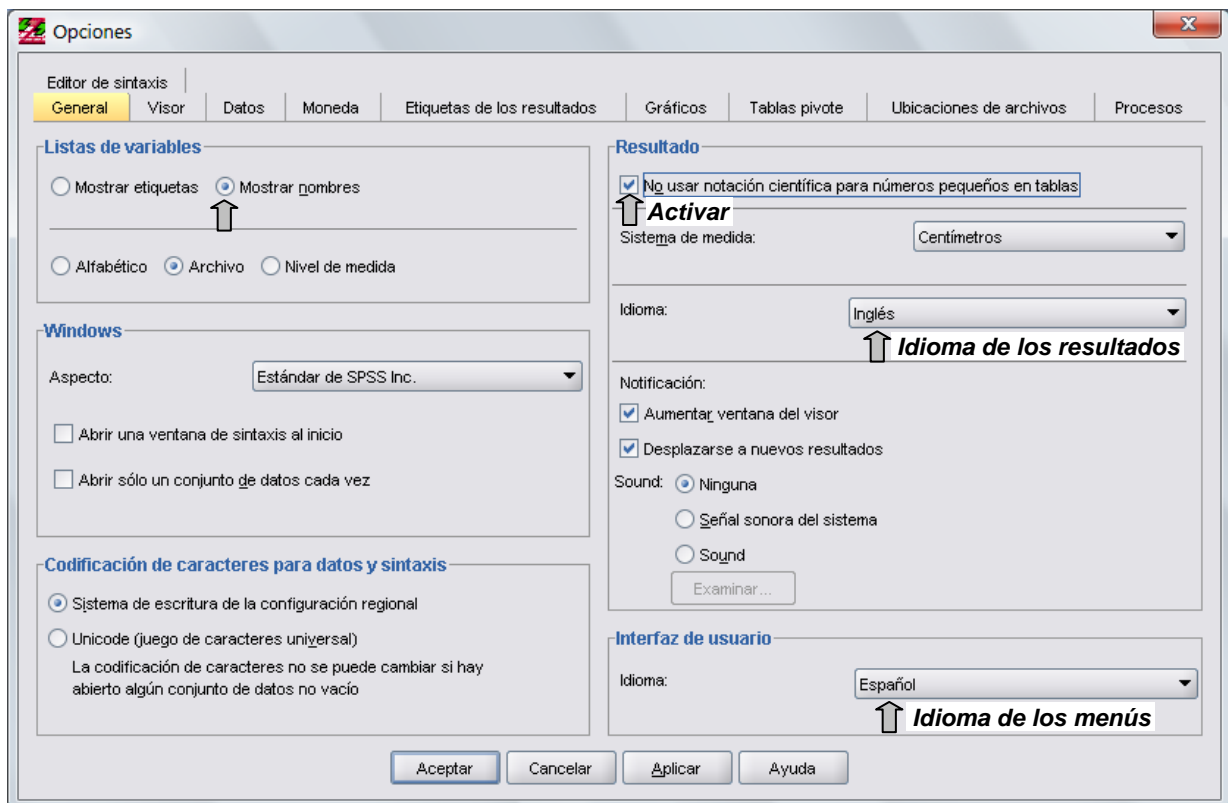
Fase 2: Explotación estadística de los datos

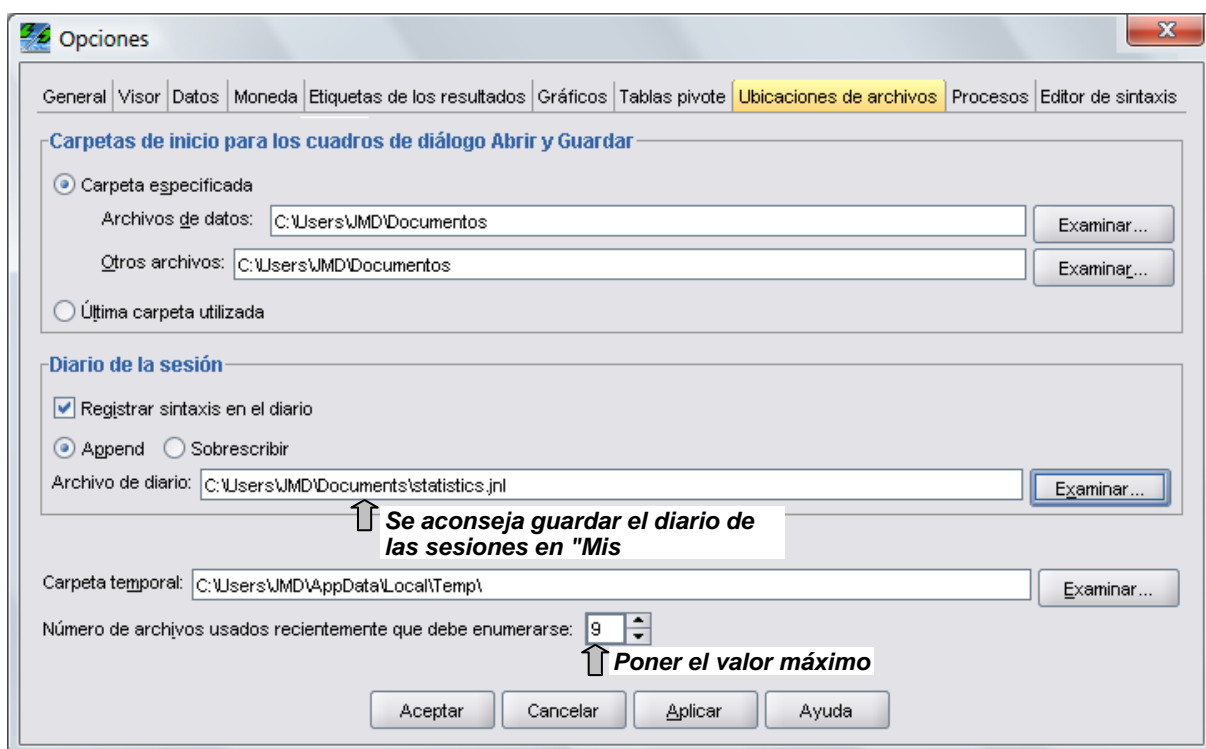
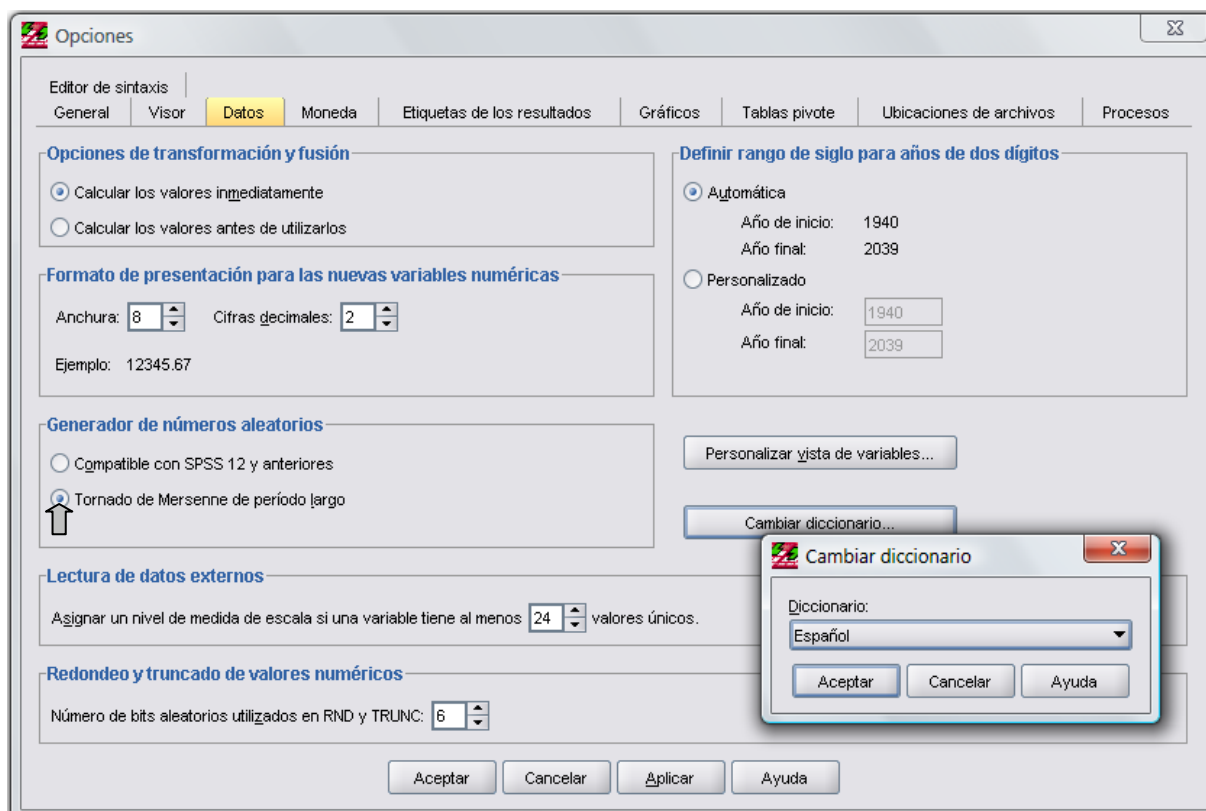
En esta fase se pueden crear uno o más archivos con la sintaxis (.sps) y uno o más archivos de resultados (.spv)



2.8 Configuración de SPSS Statistics

Se accede a ellas con el menú **Edición | Opciones...** Se señalan con ↑ las que se recomiendan activar.





Inicio de ejercicio con SPSS Statistics

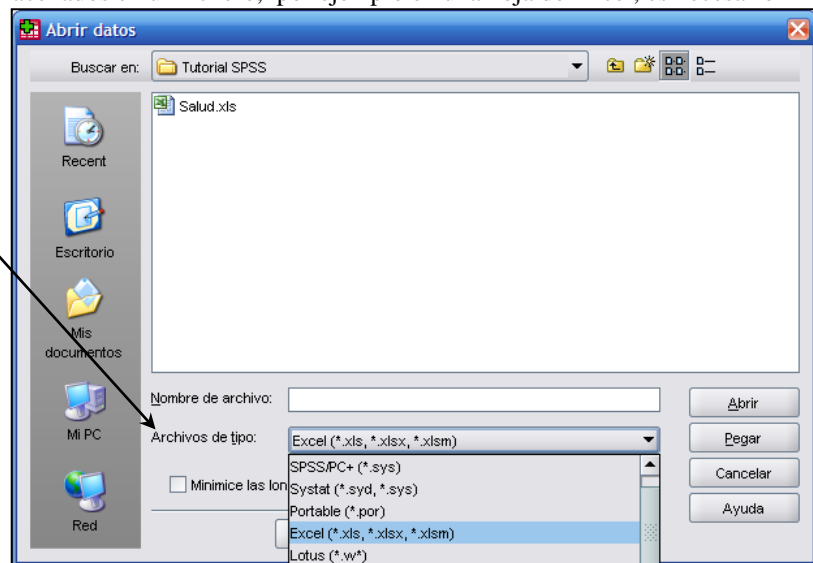
1. Abrir SPSS Statistics para configurar el programa de acuerdo con las especificaciones anteriores.
2. Abrir el menú **Edición | Opciones...** y modificar todas las opciones señaladas con ↑.
3. Salir de la ventana Opciones con el botón **Aceptar**.

3. Lectura de los datos

3.1 Introducción de los datos almacenados en un fichero

4. Si queremos trabajar con datos almacenados en un fichero, por ejemplo en una hoja de Excel, es necesario capturarlas utilizando el menú Archivo | Abrir | Datos...

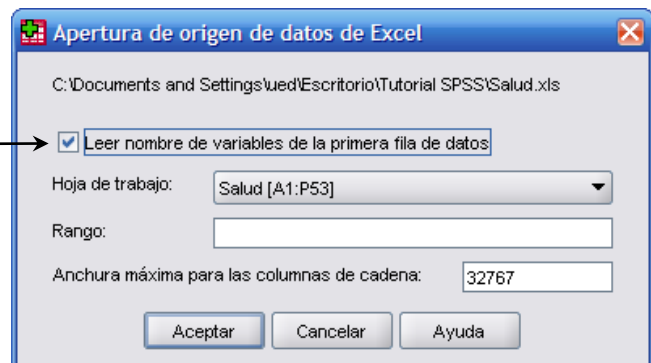
5. Una vez en la ventana *Abrir datos* nos situaremos en la carpeta **Tutorial SPSS** que contiene el archivo de datos *Salud.xls*. La lista desplegable *Archivos de tipo* permite escoger el tipo de fichero que queremos abrir. Por defecto será *SPSS Statistics (*.sav)*, pero también podemos capturar datos de Excel (*.xls). Marcamos Excel (*.xls, *.xlsx, *.xlsm) y seguidamente aparecen todos los archivos Excel que hay en la carpeta. Escogemos el archivo *Salud.xls*.



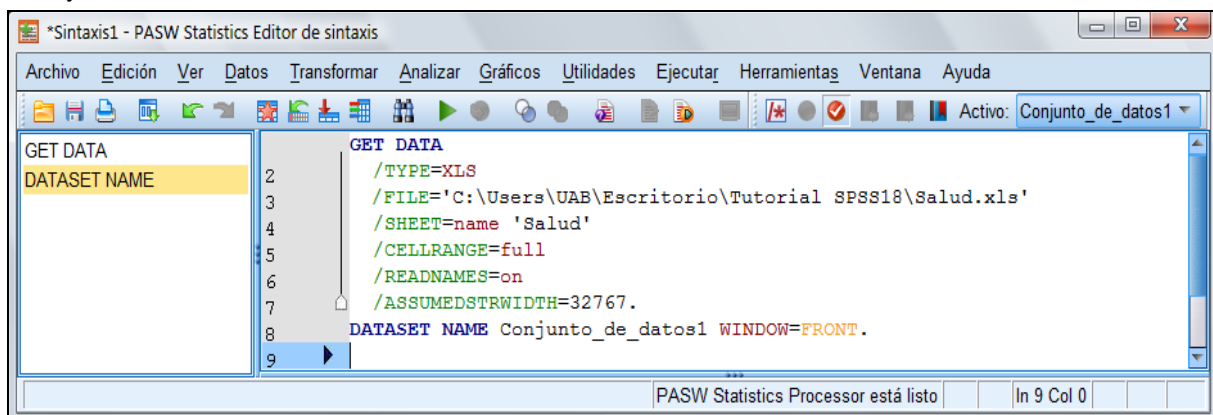
6. Hay dos posibilidades para ejecutar la acción: pulsar el botón **Abrir**, que leerá directamente los datos, o bien pulsar el botón **Pegar**, que enviará la instrucción de lectura de datos escrita en lenguaje SPSS a la ventana del **Editor de sintaxis**. Pulsamos el botón **Pegar**.

7. Se abre la ventana *Apertura de origen de datos de Excel* que permite configurar la forma en que se leerán los datos.


La casilla *Leer nombre de variables de la primera fila de datos* está activada por defecto: indica que la primera fila de la hoja de Excel contiene los nombres de las variables que corresponden a cada columna de datos. Este es nuestro caso, así que dejamos la casilla marcada y pulsamos el botón **Aceptar**.




8. Si no hay ninguna ventana de sintaxis abierta, el sistema abrirá una y añadirá la instrucción de lectura de datos. Si ya tenemos una ventana de sintaxis abierta, se añadirá la instrucción al final.

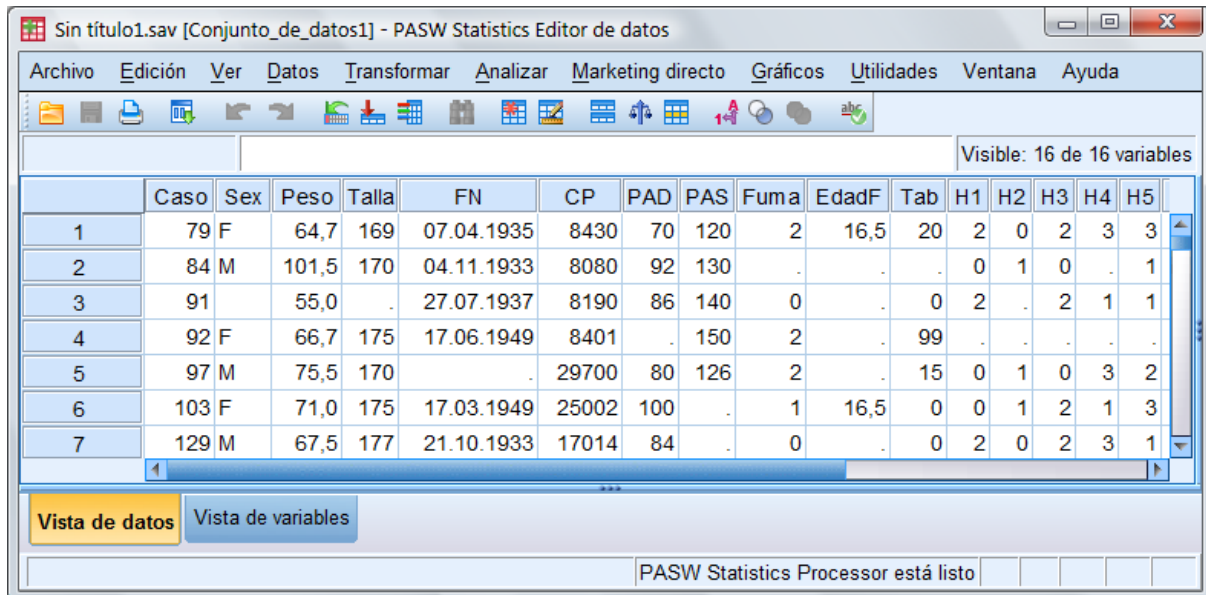


⚠ Por defecto la sintaxis que aparece en esta ventana se escribe en una fuente de espaciado proporcional. Es **importante** cambiar la fuente a **espaciado fijo** para mejorar su legibilidad. Para ello, en la ventana de sintaxis acceda al menú Ver | Fuentes..., elija **Courier New de 10pt**, y pulse **Aceptar**. A partir de este momento y en las próximas sesiones de trabajo con *SPSS Statistics* la sintaxis se escribirá con letra Courier.

9. Desde esta ventana podemos ejecutar las instrucciones de sintaxis SPSS para que tengan efecto. Marcamos las instrucciones que queremos ejecutar (normalmente todas las que ha pegado el asistente, en nuestro caso el GET DATA y el DATASET NAME) y pulsamos el botón  o utilizamos la opción de menú Ejecutar | Selección (Ctrl + R).

 El procedimiento DATASET NAME da a esta ventana de datos el nombre `Conjunto_de_datos1`. Este nombre tiene interés cuando está abierta más de una ventana de datos. En este Tutorial se trabaja con una sola ventana de datos y por lo tanto no se usará este nombre.

10. Inmediatamente se cargarán en la ventana de datos los datos contenidos en la hoja de Excel.



Sin título1.sav [Conjunto_de_datos1] - PASW Statistics Editor de datos


Archivo Edición Ver Datos Transformar Analizar Marketing directo Gráficos Utilidades Ventana Ayuda

Visible: 16 de 16 variables

	Caso	Sex	Peso	Talla	FN	CP	PAD	PAS	Fuma	EdadF	Tab	H1	H2	H3	H4	H5
1	79	F	64,7	169	07.04.1935	8430	70	120	2	16,5	20	2	0	2	3	3
2	84	M	101,5	170	04.11.1933	8080	92	130	.	.	.	0	1	0	.	1
3	91	.	55,0	.	27.07.1937	8190	86	140	0	.	0	2	.	2	1	1
4	92	F	66,7	175	17.06.1949	8401	.	150	2	.	99
5	97	M	75,5	170	.	29700	80	126	2	.	15	0	1	0	3	2
6	103	F	71,0	175	17.03.1949	25002	100	.	1	16,5	0	0	1	2	1	3
7	129	M	67,5	177	21.10.1933	17014	84	.	0	.	0	2	0	2	3	1

Vista de datos Vista de variables


PASW Statistics Processor está listo

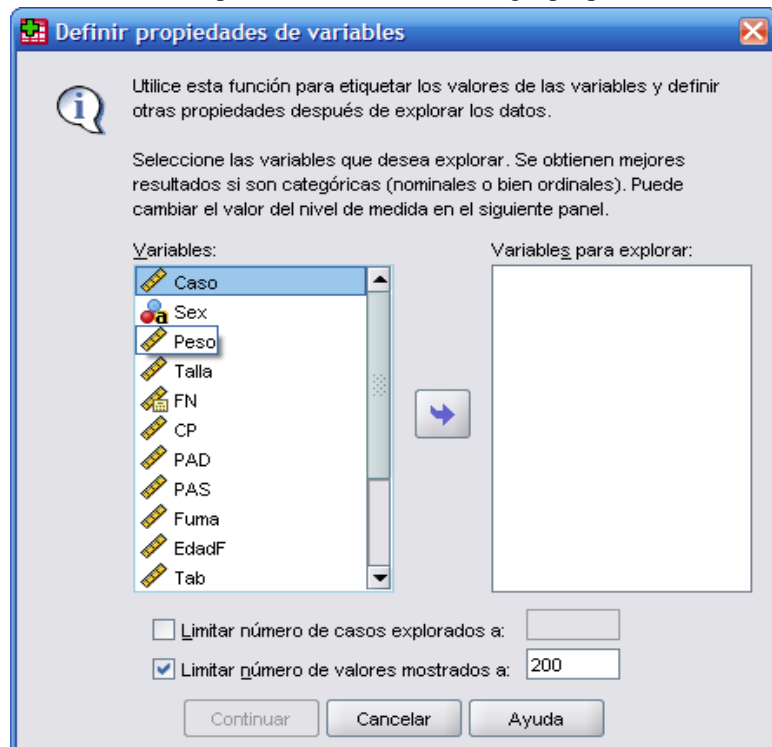
 Utilizar el editor de sintaxis y guardar las instrucciones es muy útil porque a menudo un mismo proceso debe repetirse varias veces. De esta forma es como se recomienda trabajar en este Tutorial.

3.2 Definición de las propiedades de las variables

11. El menú Datos | Definir propiedades de variables... presenta un cuadro de diálogo que permite definir las propiedades para todas las variables.

En primer lugar aparece la ventana *Definir propiedades de variables*, que permite escoger las variables de las que vamos a definir las propiedades.

Se seleccionan todas las variables y se pasan a la lista *Variables para explorar* con el botón .



Definir propiedades de variables

Utilice esta función para etiquetar los valores de las variables y definir otras propiedades después de explorar los datos.

Seleccione las variables que desea explorar. Se obtienen mejores resultados si son categóricas (nominales o bien ordinales). Puede cambiar el valor del nivel de medida en el siguiente panel.

Variables:

- Caso
- Sex
- Peso
- Talla
- FN
- CP
- PAD
- PAS
- Fuma
- EdadF
- Tab

Variables para explorar:

☐ Limitar número de casos explorados a:

☒ Limitar número de valores mostrados a:

Continuar Cancelar Ayuda

12. Al pulsar el botón **Continuar** se abre la ventana *Definir propiedades de variables*, que permite definir las propiedades para todas las variables seleccionadas.

Definir propiedades de variables

Lista de variables exploradas: Sin... Me... Variable

Variable actual: Caso Etiqueta:

Nivel de medida: Escala Sugerir Tipo: Numérico dd-mmm...

Valores sin etiqueta: 52 Atributos... Anchura: 11 Decimales: 0

Rejilla etiq. valores: Añada etiquetas a la rejilla o editelas. Puede añadir valores abajo.

	Cambiado	Perdidos	Recuento	Valor	Etiqueta
1	<input type="checkbox"/>	<input type="checkbox"/>	1	79	
2	<input type="checkbox"/>	<input type="checkbox"/>	1	84	
3	<input type="checkbox"/>	<input type="checkbox"/>	1	91	
4	<input type="checkbox"/>	<input type="checkbox"/>	1	92	
5	<input type="checkbox"/>	<input type="checkbox"/>	1	97	
6	<input type="checkbox"/>	<input type="checkbox"/>	1	103	
7	<input type="checkbox"/>	<input type="checkbox"/>	1	129	
8	<input type="checkbox"/>	<input type="checkbox"/>	1	130	
9	<input type="checkbox"/>	<input type="checkbox"/>	1	132	

Casos explorados: 52

Límite lista valores: 200

Copiar propiedades

De otra variable... A otras variables... Etiquetas automáticas

Aceptar Pegar Restablecer Cancelar Ayuda

Para cada variable tendremos que definir las siguientes propiedades:

- **Etiqueta.** Descripción de la variable que aparecerá en todos los resultados.
- **Nivel de medida.** Puede ser Nominal, Ordinal o Escala (métrica).
- **Tipo.** Los más importantes son *Numérico*, *Cadena* (letras, cifras, signos) o *Fecha*.
- **Anchura y Decimales.** Si es numérico es posible definir el número de espacios que ocupará la variable, así como el número de decimales. La *Anchura* es el ancho total que tendrá la variable, contando los decimales y el separador decimal. Por ejemplo, un peso de 101,5 kg tendrá 5 espacios de *Anchura*: 3 cifras para la parte entera, 1 decimal y el separador decimal (en este caso la coma).
- Si la variable es categórica será necesario escribir la etiqueta descriptiva junto a cada valor de cada categoría en la columna **Etiqueta**.
- Si hay valores perdidos (*user missing*) será necesario marcarlos con la casilla **Perdidos**.

! Sólo se definen los *User missing*; los *System missing* son automáticos y **no se especifican en este cuadro**.

La tabla 2.4.1 (pág. 12) presenta las propiedades de todas las variables del estudio que deberemos definir en los siguientes pasos.

13. Por ejemplo, a la variable *Caso* no es necesario definirle una etiqueta, pero sí le asignaremos el nivel de medida *Nominal*, y le asignaremos *Anchura* 3.

Definir propiedades de variables

Lista de variables exploradas: Sin... Me... Variable

Variable actual: Caso Etiqueta:

Nivel de medida: Nominal Sugerir Tipo: Numérico dd-mmm...

Valores sin etiqueta: 52 Atributos... Anchura: 3 Decimales: 0

Rejilla etiq. valores: Añada etiquetas a la rejilla o editelas. Puede añadir valores abajo.

	Cambiado	Perdidos	Recuento	Valor	Etiqueta
1	<input type="checkbox"/>	<input type="checkbox"/>	1	79	
2	<input type="checkbox"/>	<input type="checkbox"/>	1	84	
3	<input type="checkbox"/>	<input type="checkbox"/>	1	91	

Casos explorados: 52

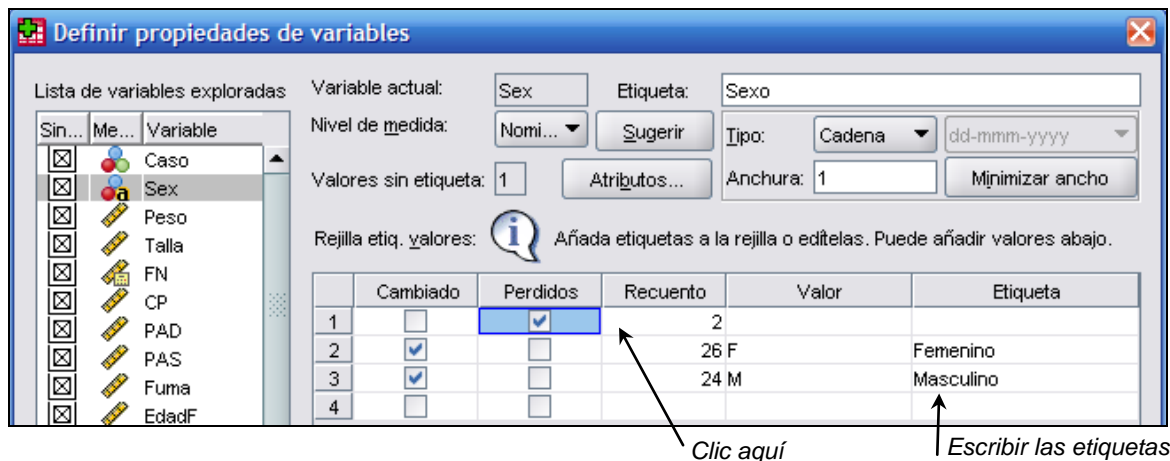
Límite lista valores: 200

Copiar propiedades

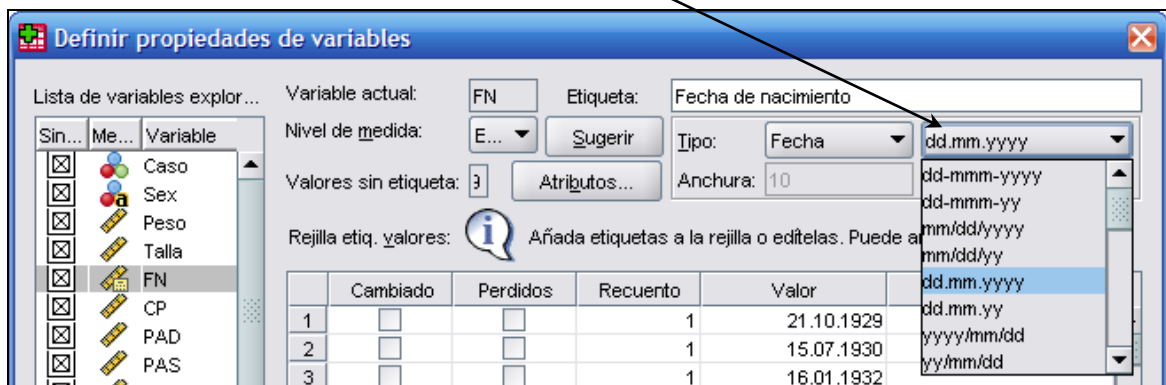
De otra variable... A otras variables... Etiquetas automáticas

Aceptar Pegar Restablecer Cancelar Ayuda

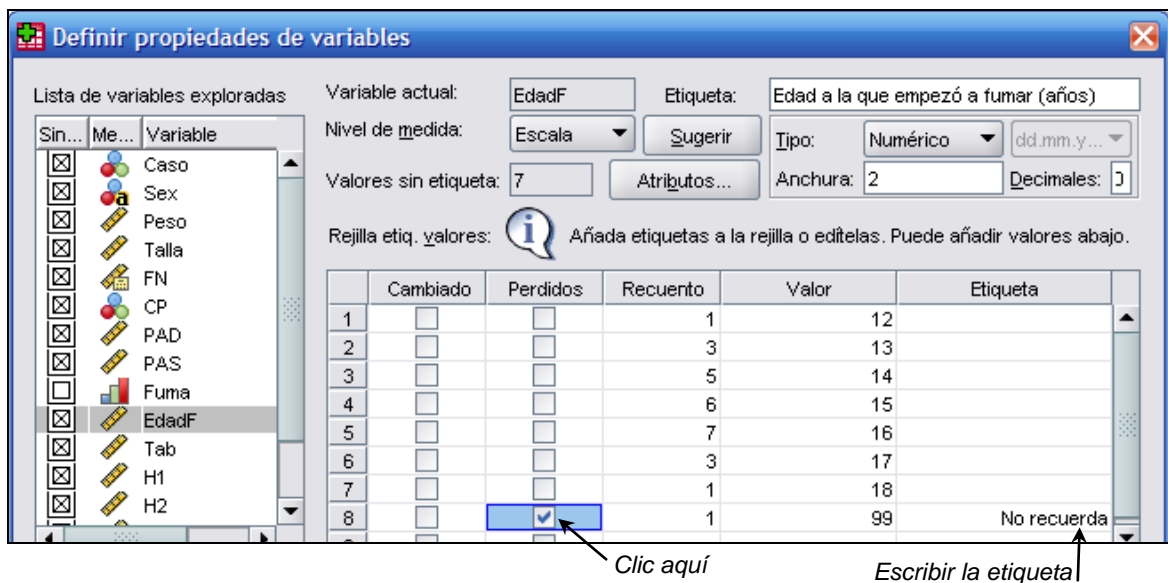
14. Para la variable *Sex* definimos la descripción *Sexo*, le asignamos las descripciones *Femenino* al código *F* y *Masculino* al código *M* y marcamos el *blanco* como un valor *Perdido*.



15. Para la variable *FN* definimos la descripción *Fecha de nacimiento*. SPSS asigna por defecto el formato *dd-mmm-yyyy*., pero preferimos ver las fechas en formato *dd.mm.yyyy*.



16. Para la variable *EdadF* definimos la etiqueta *Edad a la que empezó a fumar (años)* y anchura 2. Para esta variable el valor 99 es *Perdido*, con la etiqueta *No recuerda*.



17. Para la variable *H1* definimos la etiqueta *Práctica deportiva*, nivel de medida *Ordinal*, anchura 1 y las etiquetas asociadas a sus categorías: 0 *Nunca*, 1 *Ocasional*, 2 *Habitual*.

Definir propiedades de variables

Lista de variables exploradas: Sin... Me... Variable

Variable actual: H1 Etiqueta: Práctica deportiva

Nivel de medida: Ordinal Sugerir Tipo: Numérico dd.mm.y... Anchura: 1 Decimales: 0

Valores sin etiqueta: 0 Atributos...

Rejilla etiq. valores: Añada etiquetas a la rejilla o editelas. Puede añadir valores abajo.

	Cambiado	Perdidos	Recuento	Valor	Etiqueta
1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	24	0	Nunca
2	<input checked="" type="checkbox"/>	<input type="checkbox"/>	15	1	Ocasional
3	<input checked="" type="checkbox"/>	<input type="checkbox"/>	12	2	Habitual
4	<input type="checkbox"/>	<input type="checkbox"/>			

Casos explorados: 52 Límite lista valores: 200

Copiar propiedades: De otra variable... A otras variables... Etiquetas automáticas

Aceptar Pegar Restablecer Cancelar Ayuda

18. Las variables *H2* y *H3* tienen las mismas propiedades que *H1*, sólo cambia la etiqueta. SPSS permite copiar las propiedades de una variable a otras con los botones *De otra variable* y *A otras variables*. Si desde *H1* pulsamos el botón *A otras variables* se abre la ventana *Aplicar etiquetas y medida a*, que permite escoger a qué variables queremos copiar las propiedades. Escogemos las variables *H2* y *H3* y pulsamos el botón **Copiar**: al volver al cuadro de diálogo *Definir propiedades de variables* podemos comprobar que las variables *H2* y *H3* tienen las etiquetas de valor, el nivel de medida y la anchura de *H1*; sólo faltará definir las etiquetas del nombre de estas variables.

Aplicar etiquetas y medida a

Seleccione las variables en las que se copiarán las etiquetas de valor, el nivel de medida, la definición de valores perdidos y el formato de la variable actual. (Se funden etiquetas de valor y valores perdidos.)

Variables:

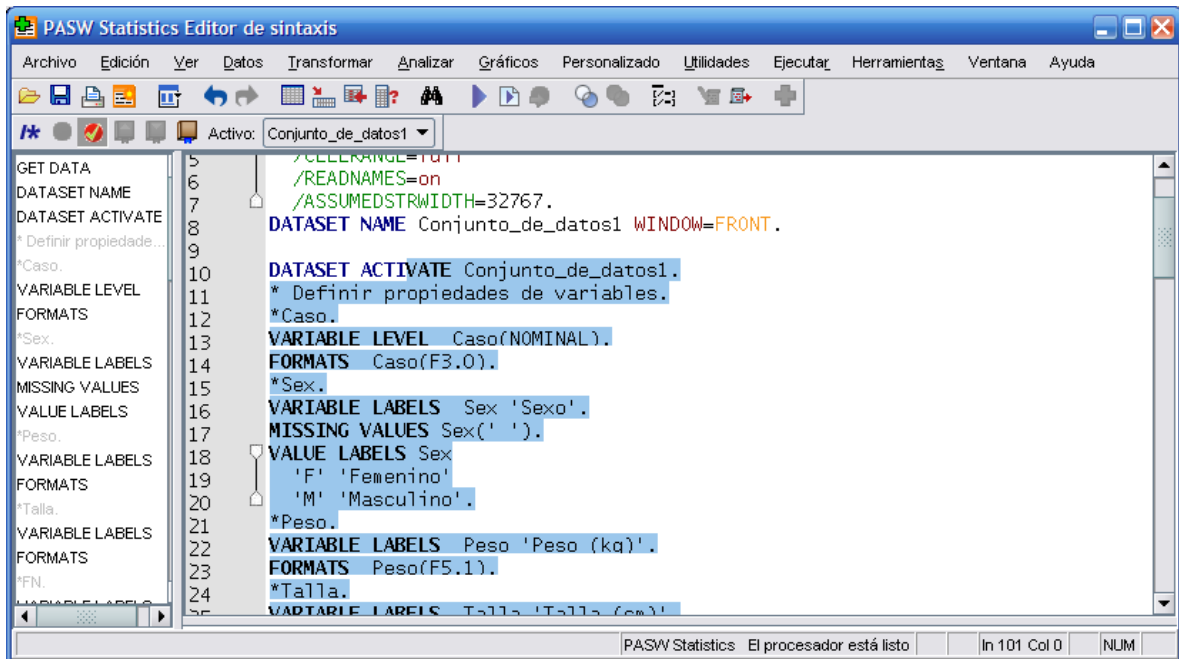
- PAD
- PAS
- Fuma
- EdadF
- Tab
- H2
- H3
- H4
- H5


Copiar Cancelar Ayuda

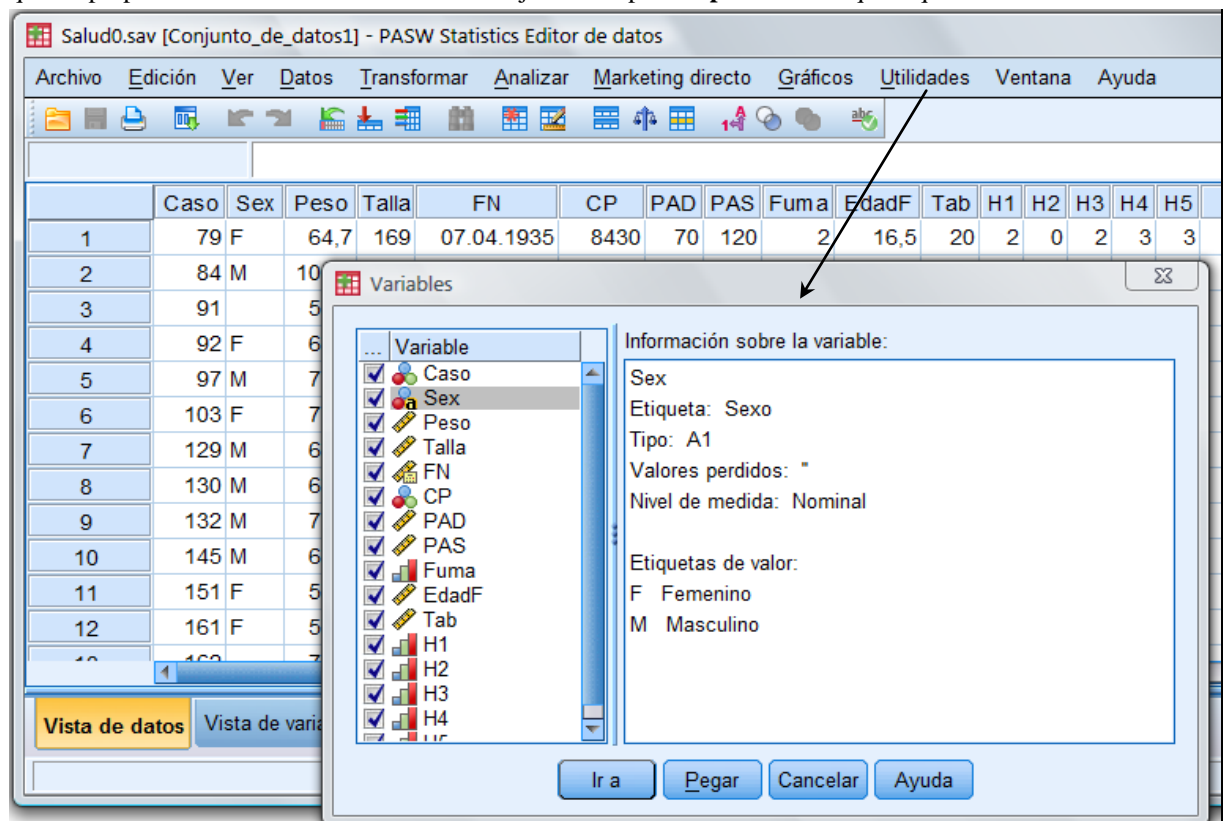
19. A continuación se define las propiedades de la variable *H4*. Y seguidamente, puesto que la variable *H5* tiene las mismas propiedades que la *H4*, usaremos el botón *A otras variables* para copiarlas a *H5*.

! El cuadro de diálogo *Definir propiedades de variables* guarda memoria de las definiciones. Por lo tanto, podemos definir las propiedades de todas las variables a la vez y al final hacer un único **Pegar**.

20. Siguiendo la tabla 2.4.1 (pág. 12) deberemos definir las propiedades de las restantes variables. Cuando acabemos de definir las propiedades de **todas** las variables, pulsamos el botón **Pegar** para que se escriban las instrucciones en la ventana de sintaxis:



21. Para ejecutar las nuevas instrucciones, se marcan con el ratón **todas las instrucciones de definición de los datos** y se pulsa el botón  para que tengan efecto.
 ! Se deben marcar **todas** las instrucciones que pega el asistente **excepto** el GET DATA que ya se ha ejecutado.
22. Finalmente deberemos comprobar que todas estas propiedades se han almacenado correctamente. Para ello se escoge la opción de menú **Utilidades | Variables...** y nos iremos situando sobre cada variable para comprobar que las propiedades son las correctas. Es un trabajo tedioso pero **imprescindible** que requiere mucha atención:



23. Además de la comprobación anterior, realizaremos una nueva comprobación para asegurarnos que los valores desconocidos de usuario (*users missing*) se han definido. Esta definición es importante porque afecta a la creación de nuevas variables y a los análisis estadísticos que se realicen con estas variables.

Para ello se abre el menú **Utilidades | Variables...** y nos iremos situando sobre las siguientes variables:

- **Sex** ¿Tiene definido el valor perdido ' '? () Sí () No
- **EdadF** ¿Tiene definido el valor perdido 99? () Sí () No
- **Tab** ¿Tiene definido el valor perdido 99? () Sí () No

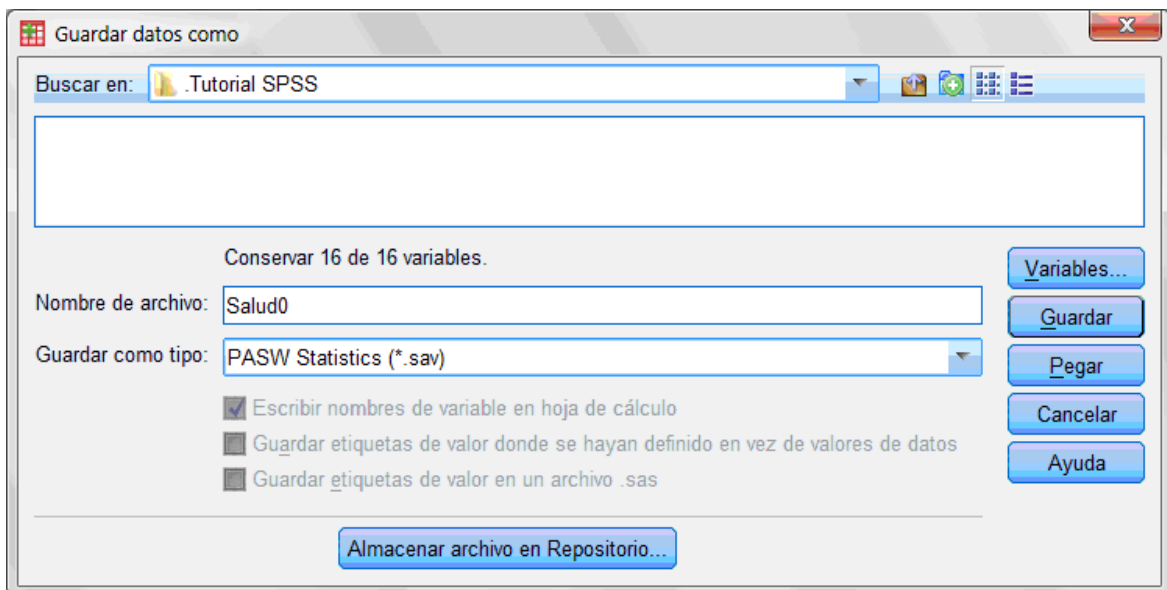
En caso que falte alguna de estas tres definiciones debe volver al **paso 11** (p. 18), seleccionar la(s) variable(s) sin la definición de los *users missing* y pulsar **Continuar**. Seguidamente se abre el cuadro de diálogo para definir las propiedades y se asignan los valores *user missing* tal como se explica en la pág. 20. Finalmente salir del cuadro pulsando **Pegar**.

Se ejecuta(n) el/los procedimiento(s) **Missing Values** pegado(s) en la ventana de sintaxis y con el menú **Utilidades | Variables...** se comprueba que los valores *user missing* se han asignado correctamente.

24. **Opcional.** Si en el punto anterior ha sido añadido algún procedimiento **Missing Values**, quedaría muy elegante que este/os procedimiento/s que se ha pegado al final de la ventana de sintaxis se desplazaran en el lugar donde se encuentran la definición de las otras propiedades de la variable afectada (con Cortar y Pegar).

3.3 Final de una sesión de trabajo: guardar las ventanas de datos y sintaxis

25. Una vez definidas las propiedades de las variables se puede guardar la información de la tabla de datos (Datos + Diccionario con las propiedades). Desde la ventana de datos, se escoge la opción de menú **Archivo | Guardar como...** y en la ventana *Guardar datos como* se selecciona la carpeta donde se guardan los datos y se escribe el nombre del archivo, **Salud0**. La extensión (**.sav**) la asigna *SPSS Statistics*.



26. De forma similar, podemos utilizar la opción de menú **Archivo | Guardar como...** desde el editor de sintaxis para guardar el archivo de sintaxis, que en este caso le llamaremos **Salud**.
27. Escoger la opción de menú **Archivo | Salir**. Es probable que pregunte ¿Quiere guardar el contenido de la ventana de Resultados? Contestar **No** y se cerrará directamente el programa porque previamente ya hemos guardado la ventana de Datos y la de Sintaxis.

4. Creación de nuevas variables

El archivo **Salud0.sav** contiene las **variables originales** recogidas en el cuestionario, pero en general para contestar las preguntas de la investigación de cualquier estudio es necesario, a partir de las variables originales, crear nuevas variables que recojan los contenidos de estas preguntas o que permitan contestarlas con más precisión. Para presentar los resultados de nuestro estudio a continuación necesitamos crear las siguientes variables:

a) Descripción de la muestra

Variable **Edad**, a partir de las fechas de nacimiento (**FN**) y de recogida de los datos (14 de abril de 2006).

Variable **CProv** numérica con el código de la provincia a partir del código postal (**CP**).

Variable **Sexo** numérica a partir de la variable original cadena **Sex**.

b) Análisis del consumo de tabaco

Variable categórica de Consumo de tabaco (**H6**) a partir de la variable original **Tab**.

Variable binaria de hábito de fumar (**HabFum**) a partir de la variable original **Tab**.

c) Análisis del peso

Variable cuantitativa con el cálculo del Índice de Masa Corporal (**IMC**) a partir de las variables originales **Peso** y **Talla**.

Variable binaria de sobrepeso **Obs** a partir de la variable **IMC**.

d) Análisis de los hábitos de salud

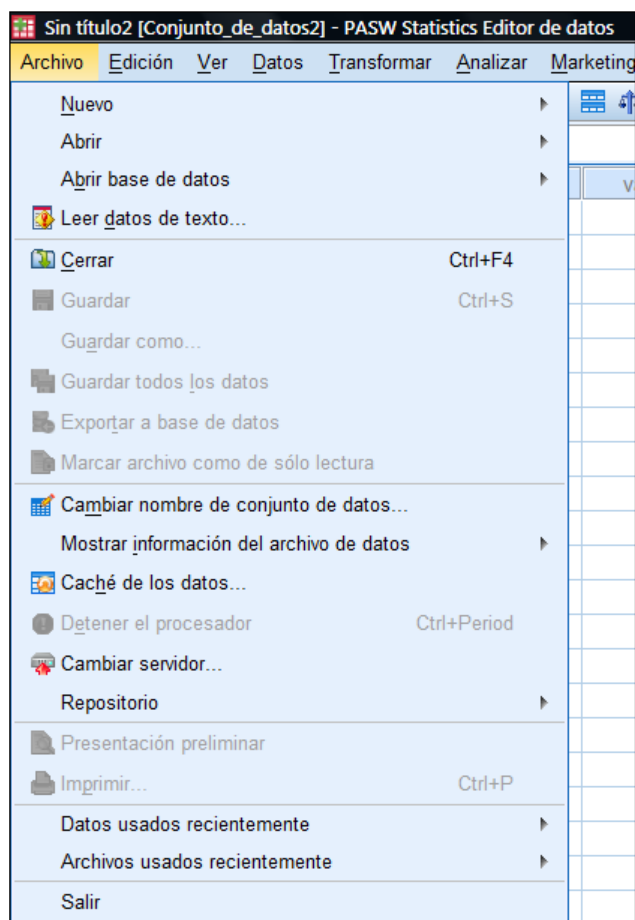
Variable binaria de hipertensión arterial (**HTA**) a partir de las variables originales **PAD** y **PAS**.

Variable cuantitativa **PS** con la puntuación de salud a partir de las variables **H1**, **H2**, **H3**, **H4**, **H5**, **H6**.

Para crear nuevas variables a partir de otras variables existentes se utiliza el menú **Transformar** que dispone de tres procedimientos básicos: **Calcular** (COMPUTE), **Recodificar** (RECODE) y **Contar** (COUNT). Estos procedimientos se ejecutan para todos los sujetos del estudio, pero también existe una estructura condicional que aplicada a ellos restringe el cálculo a los sujetos en los que la condición lógica es verdadera.

En este capítulo aprenderemos a crear todas estas variables.

4.1 Nueva sesión de trabajo: abrir las ventanas de datos y de sintaxis



Se trata de iniciar una nueva sesión. Para ello puede abrir directamente el archivo de datos **Salud0.sav** y se abrirá la ventana de datos con todas las propiedades de las variables que se han definido en el capítulo anterior. Pero también **deberemos abrir el archivo de sintaxis** para completarlo con los procedimientos de creación de variables que se ejecutarán en este capítulo.

28. Abrir los datos: Doble clic sobre **Salud0.sav** o elegir la opción Abrir del menú contextual.

29. Abrir la sintaxis: Elegir la opción de menú Archivo | Archivos usados recientemente y escoger **Salud.sps**. Seguidamente se abrirá la ventana de sintaxis con los procedimientos generados en el Capítulo anterior.

La opción de menú Archivo | Datos usados recientemente permite abrir archivos de datos utilizados en sesiones anteriores siempre que no se hayan cambiado de nombre y/o de carpeta.

← Abrir la lista y seleccionar **Salud.sps**

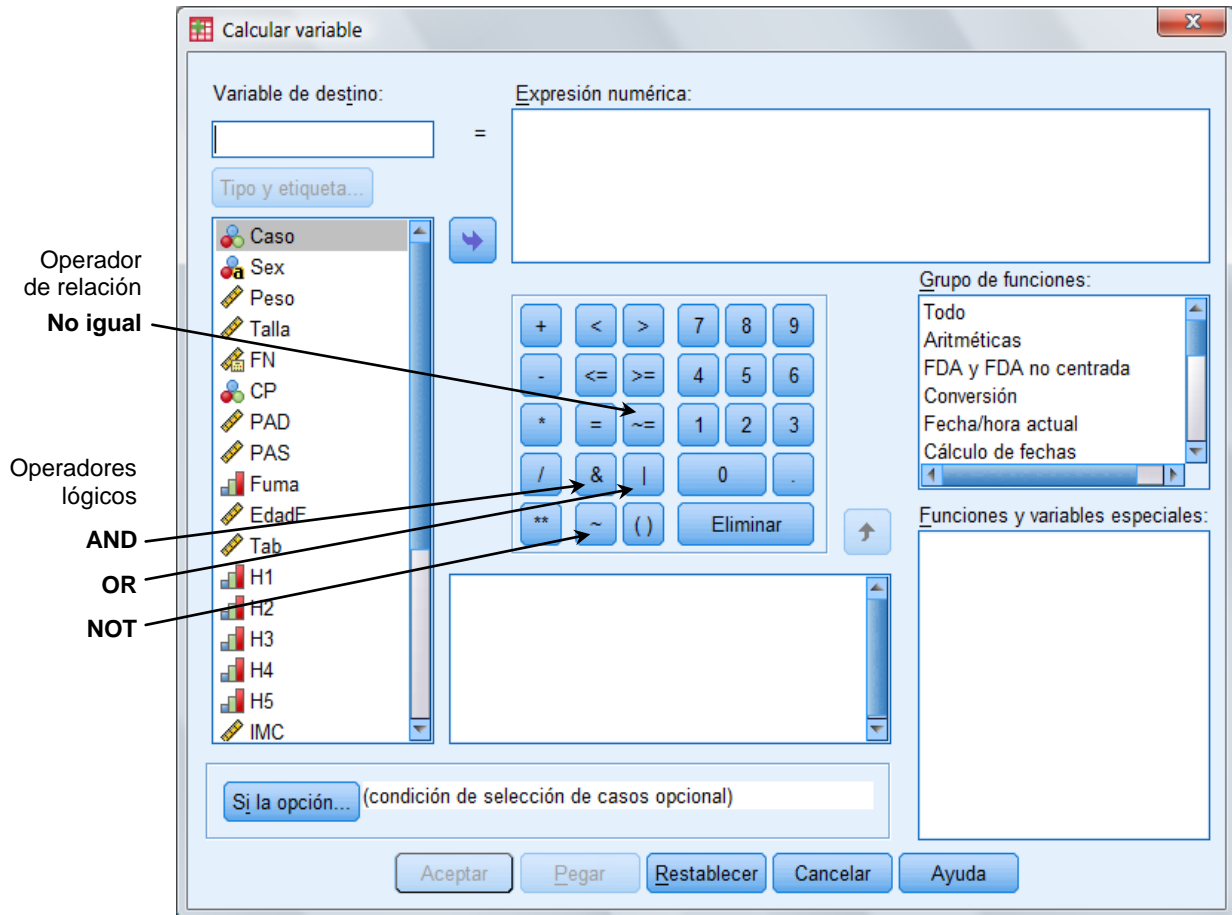
4.2 Calcular (COMPUTE)

Permite crear una **variable cuantitativa** a partir de una expresión aritmética (fórmula) o una **variable binaria** a partir de una expresión lógica (la variable toma el valor 1 en los sujetos en que es verdadera y 0 en los que es falsa).

Operaciones aritméticas: crear IMC

Crearemos el *Índice de Masa Corporal*, **IMC**, según la fórmula: $IMC (kg/m^2) = \text{Peso} / \text{Talla}^2$, donde el *Peso* se expresa en kilogramos y la *Talla* en metros.


30. Al escoger la opción de menú Transformar | Calcular variable... se abre el cuadro de diálogo *Calcular variable* que permite definir la expresión que queremos escribir.



- En el cuadro de texto **Variable de destino** se escribe el nombre de la variable a generar o modificar.
- En el cuadro de texto **Expresión numérica** se escribe la expresión que generará la nueva variable. Esta expresión se evaluará para cada sujeto y se asignará al valor de la variable para ese sujeto. Para escribir esta expresión disponemos de varias ayudas:
 - Desde la lista de la derecha podemos pasar los nombres de las variables que intervengan en la expresión.
 - Los botones permiten escribir números y los operadores aritméticos y lógicos.
 - La lista *Funciones y variables especiales* permite pasar una función a la expresión. Las funciones están agrupadas en los diferentes *Grupos de funciones*.

31. Para generar la variable **IMC** empezamos por escribir su nombre en el cuadro **Variable de destino**.


Variable de destino:
IMC

32. El botón  permite añadir los nombres de las variables de la tabla a la expresión. Lo utilizamos para añadir el *Peso* y la *Talla*. Como tenemos la Talla en centímetros y la fórmula del IMC trabaja en metros, es necesario dividir la *Talla* por 100.

Operador aritmético	Botón
Sumar	+
Restar	-
Multiplicar	*
Dividir	/
Potencia	**
Paréntesis para alterar la jerarquía de los operadores	()

33. Una vez tenemos escrita la expresión pulsamos el botón **Pegar** para añadir las instrucciones a la ventana del *Editor de sintaxis*.

COMPUTE IMC=Peso / (Talla / 100) ** 2.
EXECUTE.

34. Desde la ventana de sintaxis marcamos las instrucciones **COMPUTE** y **EXECUTE** con el ratón y pulsamos el botón ejecutar . Inmediatamente aparecerá en la ventana de datos la variable **IMC**.

Peso	Talla	IMC
64,7	169	22,65
101,5	170	35,12
55,0	.	.
66,7	175	21,78
75,5	170	26,12
71,0	175	22,48

35. A continuación hay que asignarle a esta variable las propiedades con la ventana *Definir propiedades de variables* (ver puntos 11 a 21; p. 18).

36. Se pulsa **Pegar** para añadir las instrucciones de definición de propiedades a la ventana de sintaxis y las ejecutamos para que tengan efecto.


* Definir propiedades de variables.
*IMC.
VARIABLE LABELS IMC 'Índice de mas corporal'.
FORMATS IMC(F4.1).
EXECUTE.

Operaciones aritméticas: crear CProv

La variable *Código Postal* **CP** contiene el *Código provincial*: los dos primeros dígitos del código postal indican la provincia a la que pertenece (ordenadas alfabéticamente y numeradas secuencialmente a partir de 1: la provincia 1 es *Álava*, 2 es *Albacete*, ... hasta 52 *Melilla*).

Algoritmo: Si un número tiene 5 dígitos y se quieren obtener los 2 primeros empezando por la izquierda, se debe dividir el número por 1 seguido de tantos 0 como cifras queramos eliminar, en este caso 1000, de forma que la parte entera del resultado se corresponde con los 2 dígitos que buscamos. Después se aplica la función **TRUNC** que retorna esta parte entera sin redondear.

37. Abrimos el diálogo *Calcular variable* desde el menú **Transformar | Calcular variable...** y se escribe el cociente CP/1000.

Seguidamente se marca este cociente **CP/1000**, se selecciona la función **Trunc(1)** que se encuentra dentro del grupo de funciones *Aritméticas* y se aplica a la expresión marcada pulsando el botón .

38. Pulsamos el botón **Pegar** para añadir las instrucciones a la ventana del *Editor de sintaxis*, las marcamos y las ejecutamos para obtener la nueva variable **CProv**.

COMPUTE Cprov=TRUNC(CP / 1000).
EXECUTE.

CP	Cprov
8430	8,00
8080	8,00
8190	8,00
8401	8,00
29700	29,00
25002	25,00

39. A continuación se definen las propiedades de **CProv** con la ventana *Definir propiedades de variables*, se pegan en la sintaxis y se ejecutan:

* Definir propiedades de variables.
*Cprov.
VARIABLE LEVEL Cprov(NOMINAL).
VARIABLE LABELS Cprov 'Código provincial'.
FORMATS Cprov(F2.0).
EXECUTE.

Cálculo del tiempo transcurrido a partir de dos fechas

Las fechas en SPSS se almacenan internamente como el número de segundos transcurridos desde las 0 horas del día 14 de octubre de 1582.

Unidades de medida del tiempo transcurrido

Para poder utilizar el mes y el año como unidades de medida del tiempo transcurrido es necesario definir su equivalencia en días. De las dos equivalencias siguientes, la estadística es la que se utiliza en Ciencias de la Salud:

Equivalencia administrativa:	1 mes = 30 días	1 año = 360 días
Equivalencia estadística:	1 mes = 30.4375 días	1 año = 365.25 días

Funciones para el tratamiento de fechas:

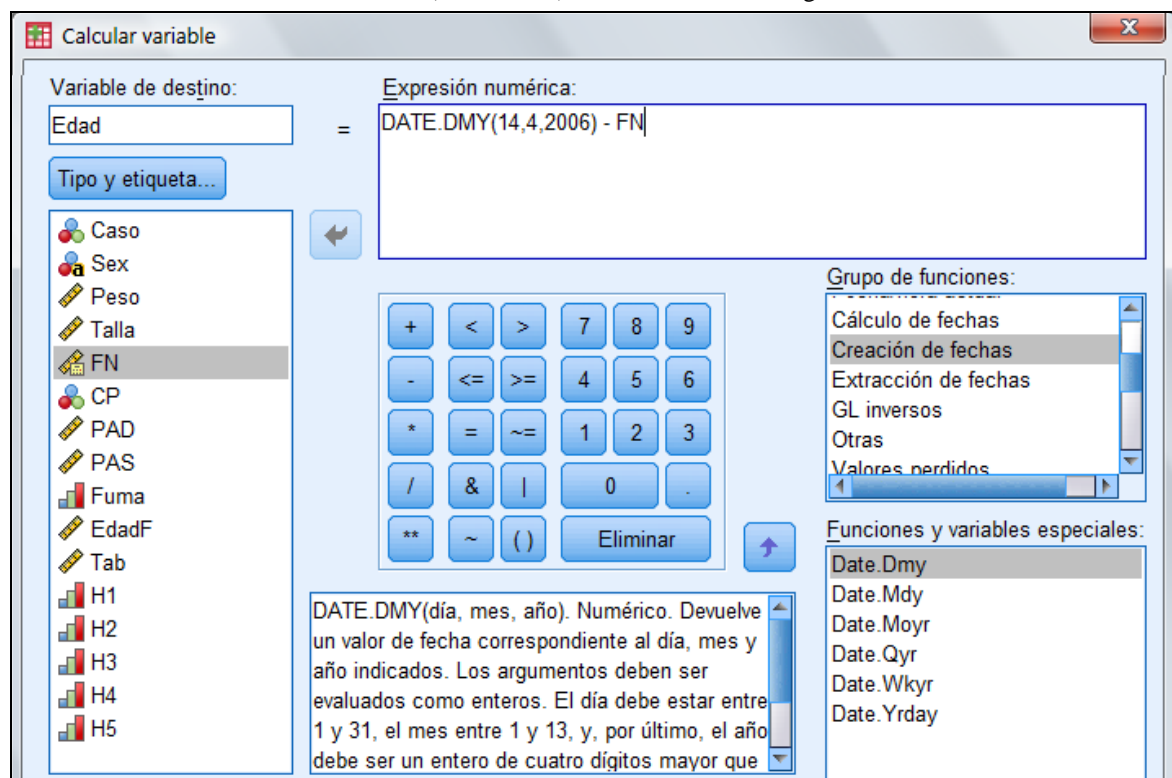
CTIME.DAYS(s)	Transforma en días un tiempo en segundos.
DATE.DMY(d,m,a)	A partir de un día, mes y año retorna la fecha.
XDATE.MDAY(f)	Retorna el día del mes de una fecha (f).
XDATE.MONTH(f)	Retorna el mes del año de una fecha (f).
XDATE.YEAR(f)	Retorna el año en cuatro dígitos de una fecha (f).

40. Queremos calcular la **Edad** del sujeto en el momento de recoger los datos, es decir, el **14 de abril de 2006**.

Se debe calcular el tiempo transcurrido entre esa fecha y la fecha de nacimiento de cada sujeto.

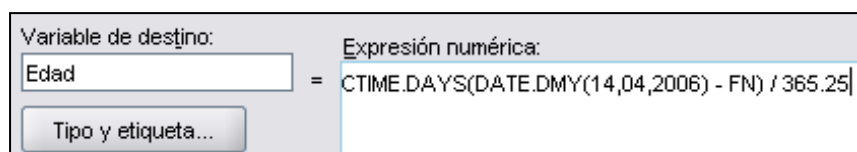
Para introducir una fecha constante en una expresión se usa la función DATE.DMY, que a partir del día, mes y año retorna la fecha.

Primero se transfiere la función **Date.Dmy** que está dentro del grupo de funciones de *Creación de fechas* y se sustituyen los interrogantes por 14, 4 y 2006. Seguidamente se resta la variable FN que contiene la fecha de nacimiento. La diferencia DATE.DMY(14,04,2006) – FN da la edad en segundos:



41. Para transformarla en días la se marca con el ratón DATE.DMY(14,04,2006) – FN, se selecciona la función **Ctime.Days** del grupo *Extracción de duración del tiempo* y se aplica a la expresión marcada pulsando el botón

42. La función CTIME.DAYS retorna el tiempo en días. Puesto que se quiere la edad en años se divide por 365.25.



Si se quisiera en semanas se dividiría por 7, y en meses por 30.4375.

43. Se pulsa el botón **Pegar** para añadir las instrucciones a la ventana del *Editor de sintaxis*, se marcan y se ejecutan para obtener la nueva variable *Edad*.

```
COMPUTE Edad=CTIME.DAYS(DATE.DMY(14,04,2006) - FN) / 365.25.
EXECUTE.
```

FN	Edad
07.04.1935	71,02
04.11.1933	72,44
27.07.1937	68,71
17.06.1949	56,82
.	.
17.03.1949	57,08
21.10.1933	72,48

44. A continuación se definen las propiedades de *Edad* con la ventana *Definir propiedades de variables*, se pegan en la sintaxis y se ejecutan.

Variable actual:	Edad	Etiqueta:	Edad (años decimales)	
Nivel de medida:	Escala	Sugerir	Tipo:	Numérico
Valores sin etiqueta:	49	Atributos...	anchura:	5
			Decimales:	2

```
* Definir propiedades de variables.
*Edad.
```

```
VARIABLE LABELS Edad 'Edad (años decimales)'.
FORMATS Edad(F5.2).
EXECUTE.
```

Si en lugar de una fecha constante tuviéramos la fecha de recogida (FR) como una variable de la tabla, calcularíamos la Edad restando la fecha de nacimiento a la fecha de recogida:

$$\text{Edad} = \text{CTIME.DAYS}(\text{FR} - \text{FN}) / 365.25.$$

La edad en años cumplidos (**EdadC**) se calcula truncando la edad en años decimales (Edad):

$$\text{EdadC} = \text{TRUNC}(\text{Edad}).$$

! Esta variable **no puede utilizarse en cálculos estadísticos** porque tiene un sesgo de -0.5 años. Los cálculos estadísticos con tiempos transcurridos deben realizarse con **tiempo decimal**, nunca con tiempo cumplido (truncado) para evitar introducir sesgos.

4.3 Calcular (COMPUTE) con expresiones lógicas

El procedimiento COMPUTE crea una **variable binaria** cuando se le asigna una expresión lógica. La variable toma el valor 1 en los sujetos en que la expresión es verdadera y 0 en los que es falsa.

Expresiones lógicas

Una expresión lógica puede considerarse como una variable interna que toma:

- valor 1 (verdad) si la expresión es verdadera
- valor 0 (falso) si es falsa
- *system missing* si no se puede evaluar

Las expresiones lógicas intervienen en la creación de variables binarias, en las transformaciones de datos condicionales y en la selección de sujetos.

Se construyen con los operadores de relación y lógicos resumidos en este cuadro.

Operadores de relación		Botón	Operadores lógicos		Botón
Más pequeño	LT	<	Intersección	AND	&
Más grande	GT	>	Reunión	OR	
Más pequeño o igual	LE	<=	Negación	NOT	~
Más grande o igual	GE	>=			
Igual	EQ	=			
No igual	NE	~=			

Una **expresión lógica simple** está formada por **dos expresiones aritméticas y un operador de relación**. Dos o más expresiones lógicas simples se pueden concatenar con los operadores lógicos AND y OR.

A continuación presentamos tres ejemplos de expresiones lógicas en las que intervienen la variable **Tab** (consumo de tabaco (c/d)), y el índice de masa corporal, IMC, definido como el cociente entre el peso (kg) y el cuadrado de la talla (m).

Condición de fumador:

Tab > 0

Condición de sobrepeso:

$\frac{\text{peso}}{\text{talla}^2} \geq 25$
 Expresión aritmética Operador de relación Expresión aritmética

Condición de fumador con sobrepeso:

Tab > 0 AND $\frac{\text{peso}}{\text{talla}^2} \geq 25$
 Expresión lógica Operador lógico Expresión lógica

Creación de una variable binaria: procedimiento de elección

Si queremos crear, por ejemplo, la variable **F15** (fumar antes de los 15 años) a partir de la variable **EdadF**, que registra la edad de inicio del hábito de fumar, tendremos que ejecutar la siguiente instrucción:

COMPUTE F15 = (EdadF < 15)

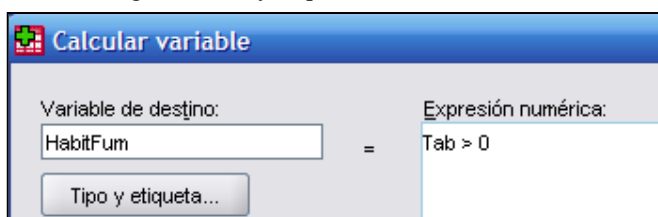
Esta tabla muestra, en seis sujetos hipotéticos, los valores que asignará esta expresión a la variable **F15**. Observad que para EdadF=99 (valor declarado *user missing*) la expresión no se puede evaluar y en consecuencia el resultado es *system missing*.

Caso	EdadF*	(EdadF < 15)
1	14	1 (verdad)
2	•	sysmis (no evaluable)
3	16	0 (falso)
4	99	sysmis (no evaluable)
5	13	1 (verdad)
6	15	0 (falso)

(*) El valor 99 se ha declarado *user missing*

45. Queremos añadir a nuestra tabla la variable **HabitFum**, *Hábito de fumar*, con dos categorías: 1 *Fumador* y 0 *No fumador* a partir de la variable original **Tab**, *Consumo de tabaco (c/d)*. Como hemos visto, **siempre** que tengamos que generar una variable binaria utilizaremos un COMPUTE con una expresión lógica. En este caso, si una persona fuma tendrá un consumo de cigarrillos mayor que 0, **Tab > 0**.

Utilizamos el diálogo *Calcular variable* para escribir la condición lógica del COMPUTE.



46. Se pulsa **Pegar** para añadir las instrucciones a la ventana del *Editor de sintaxis*, **COMPUTE** HabitFum=Tab > 0. se marcan y se ejecutan para obtener la nueva variable *HabitFum*. **EXECUTE**.

47. A continuación se define las propiedades de *HabitFum* con la ventana *Definir propiedades de variables*, sin olvidar el nivel de medida y las etiquetas de las categorías. Se pega en la sintaxis y las ejecutamos.

Variable actual: HabitFum Etiqueta: Hábito de fumar

Nivel de medida: Nominal Sugerir Tipo: Numérico dd-mmm...

Valores sin etiqueta: 0 Atributos... Anchura: 1 Decimales: 0

Rejilla eti. valores: Añada etiquetas a la rejilla o editelas. Puede añadir valores abajo.

	Cambiado	Perdidos	Recuento	Valor	Etiqueta
1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	29	0	No fumador
2	<input checked="" type="checkbox"/>	<input type="checkbox"/>	21	1	Fumador
3	<input type="checkbox"/>	<input type="checkbox"/>			

```
* Definir propiedades de variables.
*HabitFum.
VARIABLE LEVEL HabitFum(NOMINAL).
VARIABLE LABELS HabitFum 'Hábito de fumar'.
FORMATS HabitFum(F1.0).
VALUE LABELS HabitFum
  0 'No fumador'
  1 'Fumador'.
EXECUTE.
```

Tab	HabitFum
20	1
.	.
0	0
99	.
15	1
0	0

48. Queremos añadir a nuestra tabla la variable **Obs**, *Sobrepeso*, con dos categorías: 1 *Sí* y 0 *No* a partir de la variable **IMC**, *Índice de Masa Corporal*, que hemos calculado en los puntos 30 a 36. Consideraremos que una persona tiene sobrepeso si su IMC es igual o superior a 25, es decir, $IMC \geq 25$.

La mejor forma de calcular esta variable binaria es con **COMPUTE** con expresión lógica.

Se utiliza el diálogo *Calcular variable* para escribir la condición lógica que genera *Obs*.

Variable de destino: Obs Expresión numérica: = IMC >= 25

Tipo y etiqueta...

49. Pulsamos el botón **Pegar** para añadir las instrucciones a la ventana del *Editor de sintaxis*, las marcamos y las ejecutamos para generar la nueva variable *Obs*. **COMPUTE** Obs=IMC >= 25. **EXECUTE**.

50. A continuación se definen las propiedades de *Obs* con la ventana *Definir propiedades de variables*, sin olvidar el nivel de medida ni las etiquetas de las categorías. Se pegan en la sintaxis y las ejecutamos.

Variable actual: Obs Etiqueta: Sobrepeso

Nivel de medida: Nominal Sugerir Tipo: Numérico dd-mmm...

Valores sin etiqueta: 0 Atributos... Anchura: 1 Decimales: 0

Rejilla eti. valores: Añada etiquetas a la rejilla o editelas. Puede añadir valores abajo.

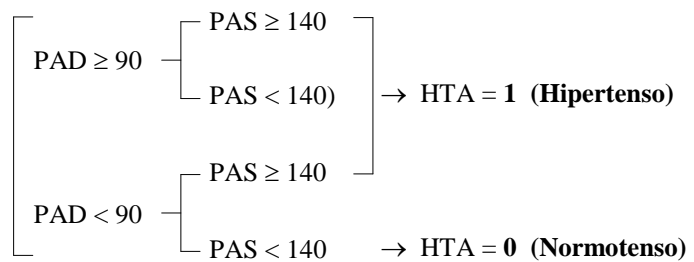
	Cambiado	Perdidos	Recuento	Valor	Etiqueta
1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	40	0	No
2	<input checked="" type="checkbox"/>	<input type="checkbox"/>	11	1	Sí
3	<input type="checkbox"/>	<input type="checkbox"/>			

```
* Definir propiedades de variables.
*Obs.
VARIABLE LEVEL Obs(NOMINAL).
VARIABLE LABELS Obs 'Sobrepeso'.
FORMATS Obs(F1.0).
VALUE LABELS Obs
  0 'No'
  1 'Sí'.
EXECUTE.
```

IMC	Obs
22,7	0
35,1	1
.	.
21,8	0
26,1	1
23,2	0
21,5	0

Creación de la variable Hipertensión: uso del operador NOT para definir una expresión lógica complementaria

A partir de los valores de las presiones arteriales diastólica y sistólica, **PAD** y **PAS** podemos clasificar a nuestros sujetos en hipertensos y normotensos según el siguiente diagrama de árbol que realiza todas las posibles combinaciones:



Por lo tanto, una hipotética variable **Normotensión** que tome como valores 1 cuando el sujeto es *Normotenso* y 0 cuando no lo es se podría escribir a partir de la reunión de las condiciones en PAD y PAS; el sujeto será normotenso si **PAD < 90 Y PAS < 140**:

Normotensión: PAD < 90 AND PAS < 140

Ahora bien, la variable que nos interesa para diagnosticar es la variable **HTA**, *Hipertensión Arterial*, que sea 1 *Hipertenso* y 0 *Normotenso*. Para generar la HTA podríamos escribir la **expresión lógica complementaria** de la que ya tenemos de la Normotensión:

HTA: PAD >= 90 OR PAS >= 140

Pero si ya tenemos escrita la condición de la Normotensión el **procedimiento de elección** para expresar la **condición lógica complementaria** es aplicar el operador NOT:

HTA: NOT(PAD < 90 AND PAS < 140)

El operador NOT niega expresiones lógicas: lo que era verdadero (1) se convierte en falso (0) y viceversa. Por lo tanto, negando la Normotensión obtendremos la Hipertensión.

⚡ Siempre es más sencillo negar una expresión lógica con el operador NOT que escribir la expresión complementaria para obtener el resultado deseado.

51. Crear la variable **HTA**, *Hipertensión Arterial* con valores 1 *Hipertenso* y 0 *Normotenso*. Para ello se usa el diálogo *Calcular variable* para escribir la condición lógica que genera HTA.

⚡ Se usan los botones del asistente para escribir los operadores lógicos & (AND) y ~ (NOT). Estos operadores se pueden escribir en forma de símbolo o con letras.

52. Pulsamos el botón **Pegar** para añadir las instrucciones a la ventana del *Editor de sintaxis*, las marcamos y las ejecutamos para generar la nueva variable **HTA**. **COMPUTE HTA= ~ (PAD < 90 & PAS < 140).**
EXECUTE.

53. A continuación se definen las propiedades de HTA con la ventana *Definir propiedades de variables*, sin olvidar el nivel de medida y las etiquetas de las categorías. Se pegan en la sintaxis y se ejecutan.

```

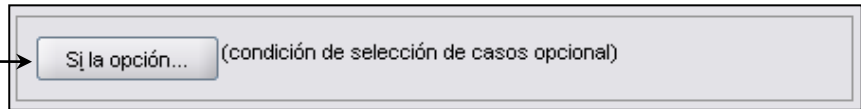
* Definir propiedades de variables.
* HTA.
VARIABLE LEVEL HTA(NOMINAL).
VARIABLE LABELS HTA 'Hipertensión Arterial'.
FORMATS HTA(F1.0).
VALUE LABELS HTA
  0 'Normotenso'
  1 'Hipertenso'.
EXECUTE.
  
```

PAD	PAS	HTA
70	120	0
92	130	1
86	140	1
.	150	1
80	126	0
100	.	1
84	.	.

4.4 Calcular Si (IF)

El procedimiento COMPUTE se ejecuta para todos los casos, pero es posible ejecutar la transformación de forma condicional utilizando el procedimiento IF. En un IF, la transformación establecida sólo se ejecuta para los sujetos en los que una determinada expresión lógica se cumple.

El procedimiento IF se define con el botón *Si la opción...* de la ventana *Calcular variable*.

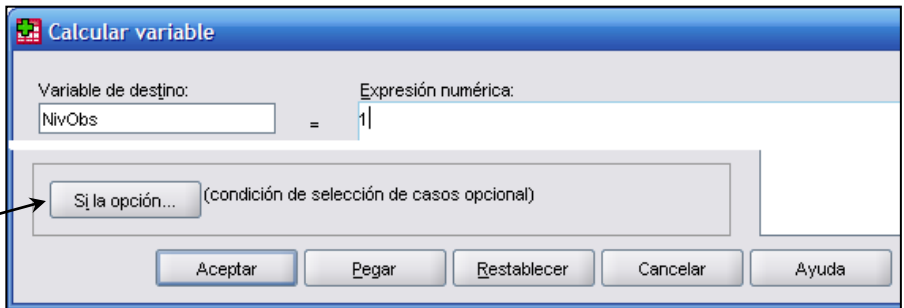


Se desea crear la variable **NivObs** (*Masa corporal categorizada*) a partir de la variable IMC ya calculada:

Si: $IMC < 18$	→	1 <i>Infrapeso</i>
Si: $IMC \geq 18$ y $IMC < 25$	→	2 <i>Normopeso</i>
Si: $IMC \geq 25$ y $IMC < 30$	→	3 <i>Sobrepeso leve</i>
Si: $IMC \geq 30$	→	4 <i>Sobrepeso grave</i>

La nueva variable **NivObs** no se define con una única instrucción, sino que tendremos que ejecutar cuatro **IF**, uno para cada rango de valores del IMC. Cada IF se aplicará únicamente a aquellos sujetos que lo cumplan, por lo que para generar correctamente *NivObs* será necesario ejecutar los cuatro IF simultáneamente.

54. Para definir **NivObs**, abrimos el diálogo *Calcular variable* y empezamos por la categoría 1 *Normopeso*. La variable **NivObs** será igual a 1 si $IMC < 18$. Lo primero que hacemos es escribir el nombre de la variable que estamos creando y asignarle como expresión numérica 1. A continuación pulsamos el botón *Si la opción...* para definir los casos en que **NivObs** será 1.

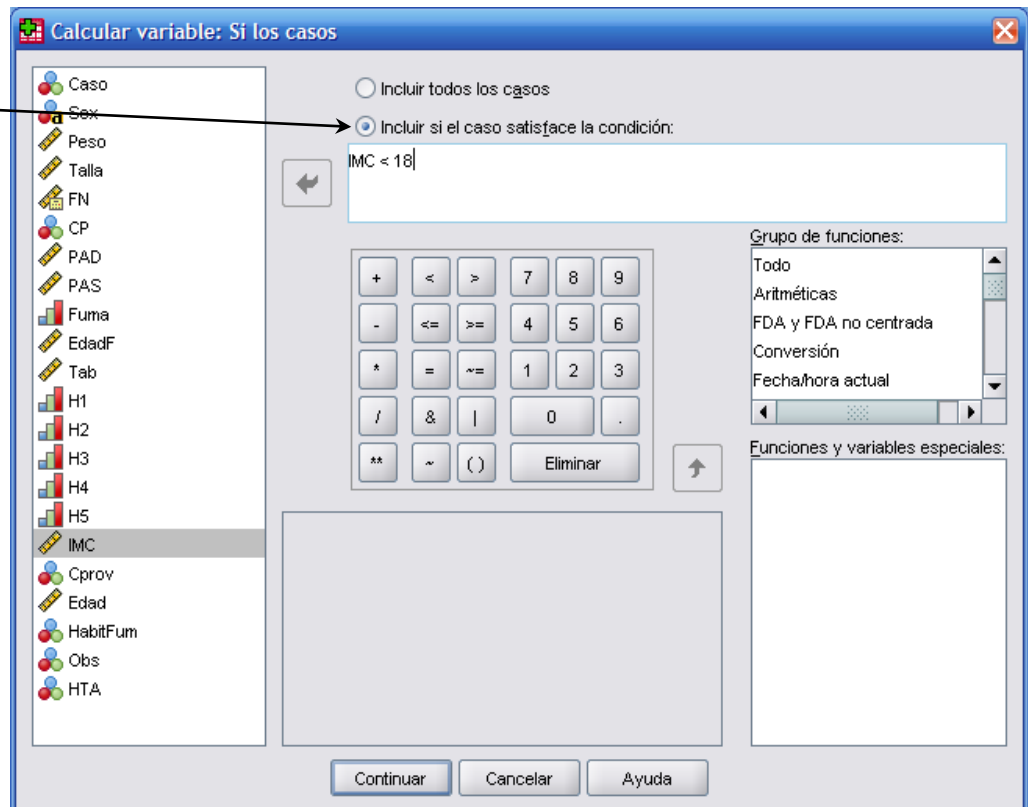


55. Seguidamente se abre la ventana *Calcular variable: Si los casos*. SPSS Statistics aplica las transformaciones por defecto a todos los casos, y por lo tanto la opción *Incluir todos los casos* está seleccionada por defecto.

Se escoge la opción *Incluir si el caso satisface la condición*.

Para definir las condiciones esta ventana funciona de forma similar a la de *Calcular variable*: hay disponibles los botones con los operadores y las funciones.

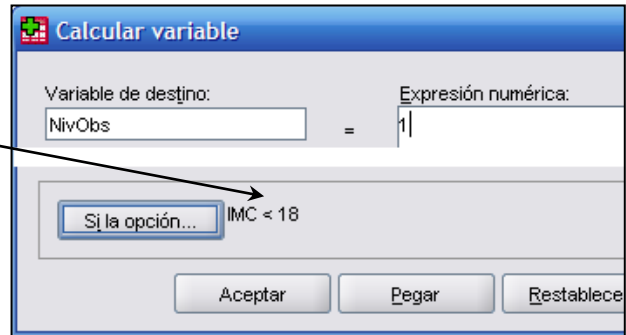
Se escribe la condición de los casos en que **NivObs** = 1 y se pulsa el botón **Continuar**.



56. La ventana *Calcular variable* se actualiza con la condición lógica que deben cumplir los registros para que *NivObs* sea 1.

57. Pulsamos **Pegar** para añadir el primer IF a la ventana de sintaxis, pero **no lo ejecutamos**. La variable *NivObs* sólo estará completa cuando hayamos definido los 4 IF correspondientes a las 4 categorías.

```
IF (IMC < 18) NivObs=1.
EXECUTE.
```



58. Repetimos los pasos 54 a 57 para cada una de las restantes categorías (2, 3 y 4) de *NivObs*, cambiando en cada caso la *Expresión numérica* y las condiciones lógicas con el botón *Si la opción...*
Pegamos cada expresión y finalmente obtenemos los cuatro IF que definen completamente la variable **NivObs**:

```
IF (IMC < 18) NivObs=1.
EXECUTE.
```

```
IF (IMC >= 18 & IMC < 25) NivObs=2.
EXECUTE.
```

```
IF (IMC >= 25 & IMC < 30) NivObs=3.
EXECUTE.
```

```
IF (IMC >= 30) NivObs=4.
EXECUTE.
```

59. Para obtener *NivObs*, se marca desde el primer IF hasta el último EXECUTE y se crea la nueva variable en la ventana de datos. A continuación definimos las propiedades de *NivObs* con la ventana *Definir propiedades de variables*, sin olvidar el nivel de medida y las etiquetas de las categorías. Las pegamos en la sintaxis y las ejecutamos.

	Cambiado	Perdidos	Recuento	Valor	Etiqueta
1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	1	1	Infrapeso
2	<input checked="" type="checkbox"/>	<input type="checkbox"/>	39	2	Normopeso
3	<input checked="" type="checkbox"/>	<input type="checkbox"/>	10	3	Sobrepeso leve
4	<input checked="" type="checkbox"/>	<input type="checkbox"/>	1	4	Sobrepeso grave
5	<input type="checkbox"/>	<input type="checkbox"/>			

* Definir propiedades de variables.

*NivObs.

```
VARIABLE LEVEL NivObs(ORDINAL).
```

```
VARIABLE LABELS NivObs 'Masa corporal categorizada'.
```

```
FORMATS NivObs(F1.0).
```

```
VALUE LABELS NivObs
```

```
1 'Infrapeso'
```

```
2 'Normopeso'
```

```
3 'Sobrepeso leve'
```

```
4 'Sobrepeso grave'.
```

```
EXECUTE.
```

IMC	Obs	NivObs
22,7	0	2
35,1	1	4
.	.	.
21,8	0	2
26,1	1	3
23,2	0	2
21,5	0	2

4.5 Recodificar (RECODE)

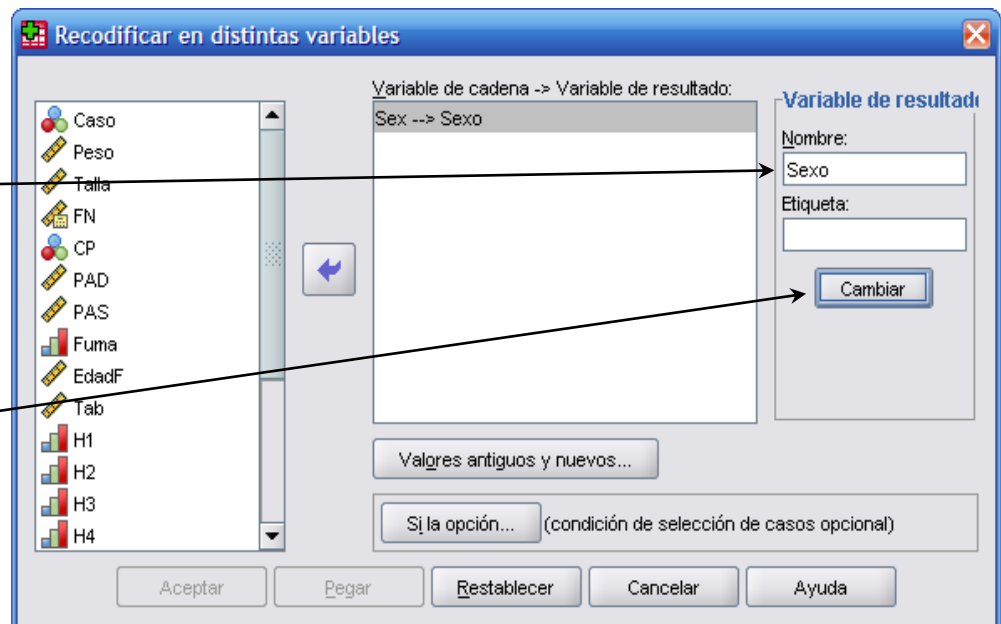
El procedimiento **Recodificar** (RECODE) cambia los valores originales de una variable por otros valores. Permite recodificar la misma variable (Transformar | Recodificar | En las mismas variables...) o crear una nueva variable recodificada manteniendo la original (Transformar | Recodificar en distintas variables...) El cuadro de diálogo de **Recodificar** permite realizar la misma recodificación a un conjunto de variables.

El cuadro también dispone del botón *Si la opción...* que permite hacer la recodificación sólo en los sujetos que cumplan una determinada condición lógica, en lugar de para todos los casos.

Recodificación de una variable cadena en una numérica

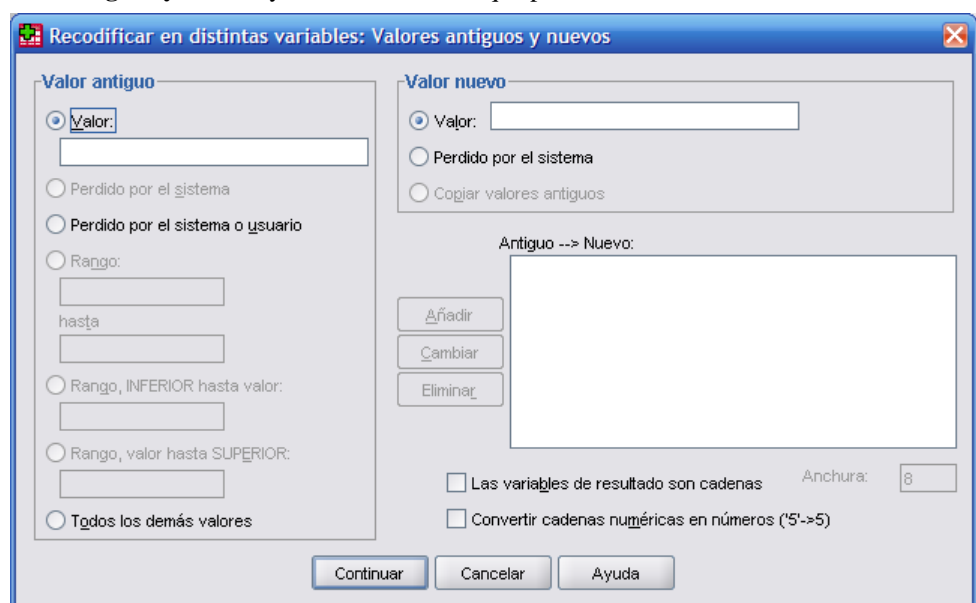
60. Queremos crear una variable numérica **Sexo**, con valores 0 *Masculino*, 1 *Femenino* a partir de la variable original **Sex** cadena. Para ello utilizaremos el RECODE INTO, es decir, recodificar en distintas variables. Escogemos la opción de menú Transformar | Recodificar en distintas variables... y se abre la ventana *Recodificar en distintas variables*.

- En el cuadro de texto *Nombre* se escribe el nombre de la nueva variable que vamos a crear, *Sexo*.
- De la lista de variables de la izquierda se pasa la variable original *Sex*.
- Es **imprescindible** pulsar el botón **Cambiar** para que en la lista se actualice:
Sex --> Sexo.



61. Pulsamos el botón **Valores antiguos y nuevos** y se abre la ventana que permite definir los cambios.

⚠ Esta ventana permite definir rangos de valores numéricos a recodificar, pero están desactivados porque es una variable cadena. Sólo están disponibles las opciones con sentido al trabajar con cadenas.



62. En esta ventana tenemos que definir las recodificaciones. Se empieza por cambiar M por 0: escribimos en *Valor antiguo* M y en *Valor nuevo* 0.

63. Pulsamos el botón **Añadir** para agregar esta recodificación a la lista de valores *Antiguo --> Nuevo*.

64. Se repite el proceso para añadir la recodificación F (Valor antiguo) --> 1 (Valor nuevo). No olvide pulsar **Añadir** para agregar la recodificación.

65. Para que la recodificación sea exhaustiva, es decir, cubra todos los posibles valores de *Sex*, marcamos *Todos los demás valores* en *Valor antiguo*, al que le asociamos *Perdido por el sistema* en *Valor nuevo* y se pulsa **Añadir**: todo valor que no sea M o F se convierte en *sysmis*.

66. Ya se ha acabado de definir las recodificaciones; se pulsa **Continuar** para volver al diálogo *Recodificar en distintas variables*, y luego se pulsa **Pegar** para añadir la sintaxis del RECODE al *Editor de sintaxis*.

67. Se ejecuta la sintaxis para crear la variable. `RECODE Sex ('M'=0) ('F'=1) (ELSE=SYSMIS) INTO Sexo.
EXECUTE.`

68. A continuación se definen las propiedades de *Sexo* con el menú *Definir propiedades de variables*, sin olvidar el nivel de medida y las etiquetas de las categorías. Se pega la sintaxis y se ejecuta.

	Cambiado	Perdidos	Recuento	Valor	Etiqueta
1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	24	0	Masculino
2	<input checked="" type="checkbox"/>	<input type="checkbox"/>	26	1	Femenino
3	<input type="checkbox"/>	<input type="checkbox"/>			

* Definir propiedades de variables.

*Sexo.

`VARIABLE LEVEL Sexo(NOMINAL).`

`VARIABLE LABELS Sexo 'Sexo'.`

`FORMATS Sexo(F1.0).`

`VALUE LABELS Sexo`

`0 'Masculino'`

`1 'Femenino'.`

`EXECUTE.`

Sex	Sexo
F	1
M	0
	.
F	1
M	0

Recodificación de una variable continua con valores discretos

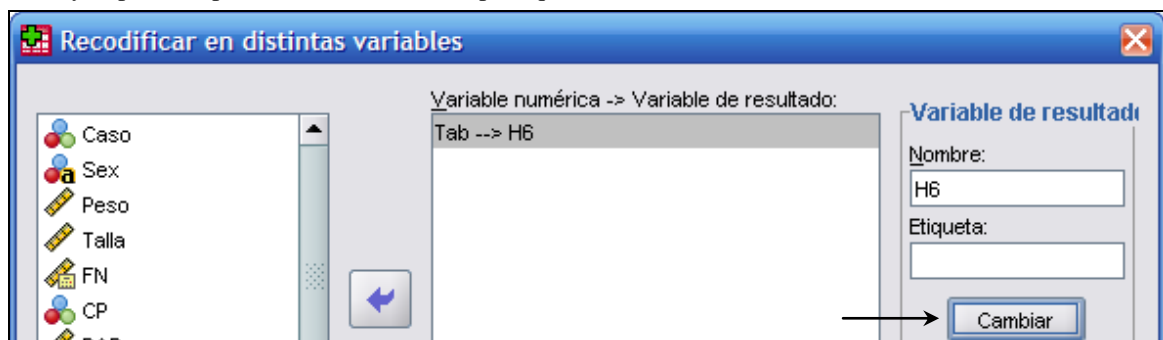
Para completar nuestras variables sobre hábitos de salud nos falta una variable **H6**, *Consumo de tabaco*, que crearemos a partir de la variable continua **Tab**, *Consumo de tabaco (c/d)* según la siguiente categorización. Puesto que los valores de **Tab** son discretos (0, 1, 2, ...) la categorización se expresará así:

Fumar entre 0 y 5 c/d:	Si: $\text{Tab} \geq 0$ y $\text{Tab} \leq 5$	→ 1 <i>Bajo</i>
Fumar entre 6 y 15 c/d:	Si: $\text{Tab} \geq 6$ y $\text{Tab} \leq 15$	→ 2 <i>Medio</i>
Fumar 16 o más c/d:	Si: $\text{Tab} \geq 16$	→ 3 <i>Alto</i>
Valores user missing:	Si: $\text{Tab} = 99$	→ 9 <i>No contesta</i>

En la recogida de datos hemos introducido el concepto de *user missing* en la variable **Tab**, asignándole valor 99, que quiere decir *No contesta*; se desea mantener esta codificación en la variable **H6**, asignándole valor 9, que también quiere decir *No contesta*. No obstante, si esta información no fuera interesante para la investigación lo recomendable sería asignarle valor *system missing*.

! Dado que el RECODE trabaja sobre todos los valores posibles de la variable original, incluidos los *system missing* y los *user missing*, **es necesario recodificar estos valores en primer lugar** para evitar errores en el algoritmo. Así, en este caso, la transformación $99 = 9$ deberá ser la primera que escribamos.

69. Abrimos el diálogo *Recodificar en distintas variables* con la opción de menú Transformar | Recodificar | En distintas variables... y escogemos como variable original **Tab** y como variable nueva que se crea **H6**. Es muy importante pulsar el botón *Cambiar* para que se actualice la transformación $\text{Tab} \rightarrow \text{H6}$.

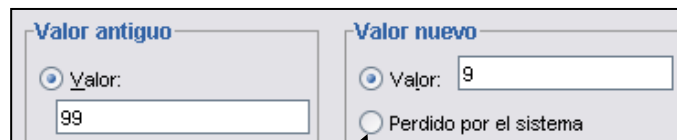


70. Se pulsa el botón **Valores antiguos y nuevos** y se debe empezar por definir la recodificación de valores *missing*:

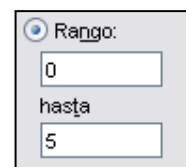
99 (Valor antiguo) -> 9 (Valor nuevo).

Se pulsa **Añadir** para agregarla a la lista de transformaciones.

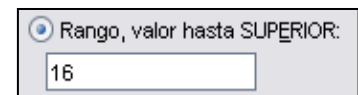
! Si no se quisiera conservar el *user missing* en la nueva variable se cambiaría a *sysmis* marcando *Perdido por el sistema*.



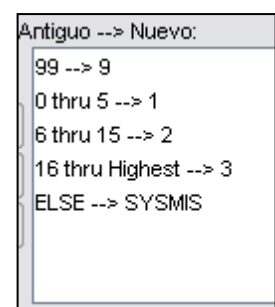
71. A continuación definimos la recodificación 0 a 5 (Valor antiguo) -> 1 (Valor nuevo). En este caso podemos utilizar la opción *Rango* en el *Valor antiguo*. Codificamos 1 en el *Valor nuevo* y pulsamos **Añadir** para agregarla a la lista de transformaciones. Para la recodificación 6 a 15 (Valor antiguo) -> 2 (Valor nuevo) podemos utilizar la misma opción.



72. Para la recodificación mayor que 16 (Valor antiguo) -> 3 (Valor nuevo) utilizamos la opción *Rango, valor hasta SUPERIOR*. (Highest)



73. Finalmente, para que la **transformación sea totalmente exhaustiva** se marca *Todos los demás valores (ELSE)* en *Valor antiguo*, al que le asociamos *Perdido por el sistema* en *Valor nuevo*. Esta recodificación cambiaría a *sysmis* cualquier valor negativo, ya que las anteriores recodificaciones incluyen todos los valores mayores o iguales a 0. Así se obtiene la lista de recodificaciones completa.



74. Ya hemos acabado de definir la recodificación de **Tab**. Pulsamos el botón **Continuar** para volver al diálogo *Recodificar en distintas variables*, y en éste pulsamos **Pegar** para añadir la sintaxis del RECODE al *Editor de sintaxis*.

```
RECODE Tab (99=9) (0 thru 5=1) (6 thru 15=2)
(16 thru Highest=3) (ELSE=SYSMIS) INTO H6.
```

EXECUTE.

⚠ La palabra clave *Highest* del RECODE significa *el mayor valor*, e incluye posibles valores *user missing* de la variable. Por lo tanto, el rango (16 thru Highest=3) incluiría el valor 99, lo que convertiría a los sujetos que *No contestan* cuántos cigarrillos fuman en fumadores con un consumo *Alto*. Para evitar este error hemos añadido la primera transformación (99=9), que lo impide: cuando *SPSS Statistics* ejecuta un RECODE lo hace leyendo de **izquierda a derecha** hasta que encuentra una transformación que se cumpla para un registro, la ejecuta y pasa al siguiente; cada registro **sólo se recodifica una vez**.

75. Se ejecuta la sintaxis para crear la variable y seguidamente se definen las propiedades de *H6*.

¡Atención! Puesto que esta variable tiene propiedades similares a *H5*, en el asistente *Definir propiedades de variables* se pasan las variables *H5* y *H6*. Una en la ventana de definición de las propiedades, se selecciona la variable *H6* y con el botón *Copiar propiedades de otras variables...* se copian las propiedades de *H5*; finalmente bastará escribir la etiqueta del nombre y, para el valor 9, la etiqueta e indicar que corresponde a un valor desconocido haciendo clic sobre la casilla de la columna **Perdidos**. Salir pegando la sintaxis y se ejecuta.

```
VARIABLE LEVEL H6(ORDINAL).
VARIABLE LABELS H6 'Consumo de tabaco'.
MISSING VALUES H6 (9).
FORMATS H6(F1.0).
VALUE LABELS H6
  1 'Bajo'
  2 'Medio'
  3 'Alto'
  9 'No contesta'.
EXECUTE.
```

Tab	H6
20	3
.	.
0	1
99	9
15	2
0	1

Recodificación de una variable continua con valores continuos

Las variables cuantitativas continuas que se registran en un estudio prácticamente siempre **presentan valores discretos**. Por ejemplo, si se registra la hemoglobina en sangre (g/dl) durante el embarazo y se introducen los valores con 1 decimal (... 10.9; 11.0; 11.1; 11.2; 11.3; ...), para crear la variable **Anemia** con las siguientes tres categorías, la recodificación se realizaría con esta sintaxis:

```
Ausente: Hb > 11 g/dl      → (11.1 thru Highest = 0)
Moderada: Hb entre 10 y 11 g/dl → (10 thru 11 = 1)
Severa: Hb < 10 g/dl      → (Lowest thru 9.9 = 2)
```

En este caso, aunque la variable es continua, no habría problema con los puntos de corte ya que al estar medida con exactitud de 0.1 g/dl (1 decimal) entre 11 y 11.1g/dl y entre 9.9 y 10.0 g/dl no habría ningún otro valor.

Sin embargo esto no ocurre con la variable continua IMC (kg/m²). Al haber sido generada a partir de la división Peso/Talla^2 el resultado es un número con muchos decimales (*SPSS Statistics* da el cálculo con 16 cifras exactas).

En el apartado 4.4 Calcular Si (IF) se generó la variable **NivObs** con las siguientes instrucciones:

```
IF (IMC < 18) NivObs=1.
IF (IMC >= 18 & IMC < 25) NivObs=2.
IF (IMC >= 25 & IMC < 30) NivObs=3.
IF (IMC >= 30) NivObs=4.
```

Con un solo RECODE se puede generar de una forma mucho más compacta la variable **NivObs** creada con estos cuatro procedimientos IF ¿pero para indicar, por ejemplo < 18, deberíamos escribir 17.9999999999999999?

En el caso de variables con valores continuos el RECODE se escribe de la siguiente manera. Puesto que los **límites superiores** de los intervalos están **excluidos** el RECODE se debe empezar por los valores más altos y poner como límite superior el límite inferior del intervalo anterior:

```
RECODE IMC (30 thru Highest = 4)
           (25 thru 30 = 3)
           (18 thru 25 = 2)
           (Lowest thru 18 = 1)
           (ELSE = SYSMIS) INTO NivObs.
```

¿Qué ocurriría si un sujeto tuviera exactamente valor 30? Puesto que la recodificación se realiza en el orden especificado, el valor 30 se encuentra en el primer intervalo (30 thru Highest = 4) de manera que se asignaría a este sujeto NivObs=4 y se pasaría a recodificar el siguiente sujeto. Lo mismo ocurriría con un sujeto que tuviera exactamente valor 25: primero mira si está en el primer intervalo, luego mira si está en el segundo intervalo y como pertenece al segundo intervalo asignaría a este sujeto NivObs=3 y se pasaría a recodificar el siguiente sujeto, es decir, un sujeto con IMC=25 no pasaría nunca al tercer intervalo.

Este RECODE genera igualmente la variable *NivObs* con las 4 categorías de manera mucho más compacta: un solo procedimiento en lugar de los cuatro IF.

4.6 Contar (COUNT)

El procedimiento **Contar valores dentro de los casos** (COUNT) crea una nueva variable que toma como valor el número de veces que un valor o lista de valores se encuentra en un determinado conjunto de variables del sujeto. Así pues, el valor máximo que puede tomar la variable creada por el COUNT es el número total de variables a contar, y el valor mínimo es **0** (esto sucede en los sujetos en que las variables a contar no contienen ninguno de los valores de la lista o en los que todos son valores desconocidos).

Por ejemplo, si en un estudio se recogen la presencia de 8 síntomas característicos de una enfermedad en las variables binarias S1, S2, ... S8, codificando 1 si el síntoma está presente y 0 si está ausente, el procedimiento COUNT permite crear una nueva variable (NumSint) con el número total de síntomas que presenta cada sujeto. También permitiría crear una nueva variable (NumMis) con el número de síntomas que presentan valor *missing* (es decir, que no se han registrado) en cada sujeto.

Si se realizan estos cálculos con el cuadro de diálogo del procedimiento **Contar** se obtendrá la siguiente sintaxis para el número de síntomas presentes (que contienen el código 1):

```
COUNT NumSint = S1 S2 S3 S4 S5 S6 S7 S8 (1) .
```

y la siguiente sintaxis para contar el número de síntomas no registrados (con valor *system missing*):

```
COUNT NumMis = S1 S2 S3 S4 S5 S6 S7 S8 (SYSMIS) .
```

Veamos otro ejemplo. Se dispone de la presión arterial (mmHg) medida a la entrada del estudio (PAS0, PAD0) y de los tres seguimientos posteriores (PAS1, PAD1, PAS2, PAD2, PAS3, PAD3) y se desea crear la variable nPAS con el número de veces que a lo largo del estudio la presión sistólica es igual o superior a 140 mmHg y la variable nPAD con el número de veces que a lo largo del estudio la presión diastólica es igual o superior a 90 mmHg. La sintaxis que permite crear estas variables es:

```
COUNT nPAS = PAS0 PAS1 PAS2 PAS3 (140 THRU Highest) .
```

```
COUNT nPAD = PAD0 PAD1 PAD2 PAD3 (90 THRU Highest) .
```

! La palabra clave **Highest** significa **hasta el mayor valor**: no se puede emplear si las variables tuvieran definidos valores *user missing* (p.e. 999) ya que los incluiría en el recuento. Si esto sucede (cosa poco probable) **Highest** se debería sustituir por un valor que nadie lo superara (p.e. 250).

Estas variables creadas tomarán valor 0 si ninguna de las presiones supera el límite establecido y tomarán valor 4 si las cuatro presiones superan los límites establecidos.

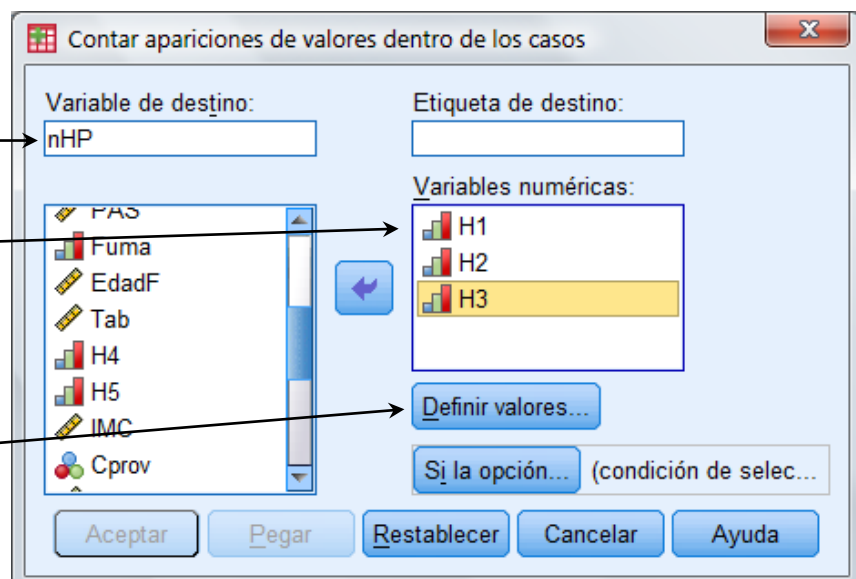
📁 El cuadro de diálogo del procedimiento **Contar** dispone también del botón *Si la opción...* que permite hacer el recuento sólo en los sujetos que cumplan una determinada condición lógica, en lugar de para todos los casos.

76. Se desea saber cuántos hábitos positivos practican de forma ocasional o habitual los sujetos del estudio. Para ello deberemos contar cuantas de las variables H1, H2 y H3 contienen los códigos 1 (ocasional) o 2 (habitual). Abrimos el cuadro de diálogo *Contar apariciones de valores dentro de los casos* con la opción de menú Transformar | Contar valores dentro de los casos...

En el cuadro de texto *Variable de destino* se escribe el nombre de la variable que contendrá el resultado del recuento.

Se pasan a la lista *Variables numéricas* las variables H1 H2 H3 que contienen los valores a contar.

Se pulsa el botón **Definir valores...** para especificar los valores a contar.



77. Se abre la ventana *Contar valores dentro de los casos: Contar los valores*. Este cuadro es parecido al del RECODE, y permite definir los valores que vamos a contar, ya sea uno o varios valores concretos, un rango de valores, desde el menor valor hasta un valor, a partir de un valor hasta el mayor, perdidos de sistema (*sysmis*) o usuario (*missing*). Sólo se deben contar los valores 1 y 2. Para ello se escribe 1 en el cuadro *Valor* y pulsar **Añadir**. A continuación se escribe 2 y se pulsa **Añadir**.

78. Pulsamos el botón **Continuar** y a continuación **Pegar** para obtener la sintaxis:

```
COUNT nHP= H1 H2 H3 (1) H1 H2 H3 (2) .
EXECUTE.
```

H1	H2	H3	H4	H5	H6	nHP
2	0	2	3	3	3	2,00
0	1	0	.	1	.	1,00
2	.	2	1	1	1	2,00
.	9	,0
0	1	0	3	2	2	1,00
0	1	2	1	3	1	2,00

79. Ejecute la sintaxis y compruebe que se añade la variable nHP a la matriz de datos.
80. Definir las propiedades de *nHP* con la ventana *Definir propiedades de variables*. Tenga presente que esta variable contiene un recuento que varía entre 0 y 3.

Se sale de la ventana pegando la sintaxis y se ejecuta:

```
VARIABLE LABELS nHP 'Número de hábitos positivos que practica'.
FORMATS nHP (F1.0) .
```

81. Crear la variable @M que contenga el número de hábitos de la escala de salud que no se han contestado. Para ello se abre de nuevo el cuadro de diálogo *Contar apariciones de valores dentro de los casos*, se pulsa **Restablecer** para vaciar el cuadro (ya que contiene el recuento anterior) y se pasan las variables H1 a H6 a la lista *Variables numéricas* y se pulsa **Definir valores...** para especificar los valores a contar.

82. Se abre la ventana *Contar valores dentro de los casos: Contar los valores*.

Puesto que la variable H6 también tiene valores *user missing*, se elije en el cuadro *Valor* la opción *Perdido por el sistema o usuario* y se pulsa **Añadir**.

83. Pulsamos el botón **Continuar** y a continuación **Pegar** para obtener la sintaxis:

```
COUNT @M=H1 H2 H3 H4 H5 H6(MISSING) .  
EXECUTE.
```

84. Ejecute la sintaxis y compruebe que se añade la variable @M a la matriz de datos.

! El nombre de la variable empieza por el carácter @ para recordar que es una variable auxiliar para realizar cálculos posteriores. Por lo tanto no debe definir las propiedades de esta variable porque una vez utilizada se borrará de la ventana de datos.

H1	H2	H3	H4	H5	H6	@M
2	0	2	3	3	3	,0
0	1	0	.	1	.	2,00
2	.	2	1	1	1	1,00
.	9	6,00
0	1	0	3	2	2	,0
0	1	2	1	3	1	0

4.7 Construcción del indicador de salud

El procedimiento que se explica en este ejercicio puede utilizarse para obtener la puntuación total de cualquier escala (calidad de vida, satisfacción, etc.) o test (alcoholismo, depresión, ansiedad, etc.).

Se trata de crear, a partir de las variables sobre hábitos de salud de nuestro estudio (H1 a H6), una nueva variable **PS** (Puntuación salud) que contenga una puntuación cuantitativa sobre cómo de saludable es el estilo de vida de los sujetos. Para ello se decide asignar a cada categoría de H1 a H6 una puntuación según los criterios que se especifican en esta tabla:

! No se debe confundir el **código** de una respuesta con los **puntos** asignados a dicha respuesta. En las variables H1 a H3 la puntuación coincide con los códigos, pero en las variables H4 a H6 la puntuación funciona de forma inversa a los códigos.

Cuestionario sobre hábitos de salud: códigos de respuesta y puntuación	
Hábitos positivos	Hábitos negativos
H1: Práctica deportiva	H4: Consumo de alcohol
H2: Dieta equilibrada	H5: Consumo de cafeína
H3: Descanso regular	H6: Consumo de tabaco
(0) Nunca = 0 puntos	(1) Bajo = 2 puntos
(1) Ocasional = 1 punto	(2) Medio = 1 punto
(2) Habitual = 2 puntos	(3) Alto = 0 puntos

La puntuación mínima que se puede obtener con esta escala es 0 puntos: es el caso de los sujetos con los peores hábitos de salud. La puntuación máxima es 12 puntos: la obtienen los sujetos que contestan todos los ítems en el sentido máximo de salud (2 puntos por ítem). Las puntuaciones intermedias dan idea de lo sanos que son los hábitos de los sujetos.

Por ejemplo, la puntuación del sujeto 103 se calcula tal como indica la tabla:

$$PS = 0+1+2+2+1+2 = 7 \text{ puntos}$$

Este sujeto ha contestado las 6 preguntas (ítems), pero **lo más habitual** en un estudio es que haya sujetos que dejen algunos ítems por contestar.

Por ejemplo el sujeto 91, cuya puntuación se calcula como indica la tabla, ha dejado 1 ítem sin contestar. Observe que todos los hábitos del sujeto 91 son en sentido de salud (cada uno de ellos contribuye con 2 puntos), pero ha obtenido sólo 10 puntos porque ha dejado H2 sin contestar.

! Esta forma de calcular la puntuación total equivale a asignar 0 puntos a los ítems sin respuesta, es decir, como si lo hubiese contestado en el sentido de “no salud”.

Corrección manual del cuestionario del sujeto 103:

	H1	H2	H3	H4	H5	H6	Total
Respuesta	0	1	2	1	3	1	
Puntuación	0	1	2	2	0	2	7

Corrección manual del cuestionario del sujeto 91:

	H1	H2	H3	H4	H5	H6	Total
Respuesta	2	.	2	1	1	1	
Puntuación	2		2	2	2	2	10

Un ejemplo que ilustra el problema de los ítems sin contestar

Vamos a explicar este problema con un ejemplo diferente del de la escala de Salud de este Tutorial. Suponga que ha aplicado un test de depresión que tiene 30 ítems y cada ítem contestado en sentido patológico contribuye con 1 punto. El punto de corte a partir del cual se considera probable un síndrome depresivo es **19 puntos**.

Suponga que un sujeto ha obtenido una puntuación total de 16 puntos, lo que descarta la posibilidad de presentar una depresión. Sin embargo, este sujeto ha dejado 10 ítems sin contestar (una tercera parte del test), por lo tanto, si hubiera contestado estos ítem en el mismo sentido que los otros obtendría una puntuación muy superior al punto de corte 19 y se le hubiese detectado un posible síndrome depresivo.

Este ejemplo evidencia la **absoluta necesidad** de corregir la puntuación total cuando haya ítems sin contestar. Los algoritmos informáticos para crear la puntuación de un test **siempre deben incluir esta corrección** porque cuando se administra un test **es frecuente** que los sujetos dejen algún ítem sin contestar.

¿Cómo debe realizarse esta corrección?

Debería constar en el manual de administración del instrumento pero, lamentablemente, no es habitual encontrarlo. El procedimiento más aceptado consiste en realizar los dos siguientes pasos:

- 1º) Poner un punto de corte en el número de ítems sin contestar a partir del cual se invalida la prueba. A pesar de que para muchos es demasiado laxo, la comunidad científica lo acostumbra a situar en el 50%.
- 2º) Si el número de ítems contestados es superior al punto de corte establecido, se asigna a cada ítem que no contesta la puntuación media de los ítems contestados.

Si K designa el número total de ítems del test, @PT la puntuación total sin contabilizar los @M ítems sin respuesta y PT la puntuación corregida, a cada ítem no contestado por el sujeto se le asigna la puntuación media:

$$@PT / (K - @M)$$

En el ejemplo del test de depresión de K=30 ítems, el sujeto con puntuación @PT=16, que ha dejado @M=10 ítems sin contestar, tiene una puntuación media por ítem contestado de 0.8 puntos/ítem:

$$@PT / (K - @M) = 16 / (30 - 10) = 16 / 20 = 0.8 \text{ puntos/ítem}$$

Si se asigna a cada ítem no contestado esta puntuación media, la puntuación total corregida PT se obtiene multiplicando la puntuación media por el número total de ítems:

$$PT = K \times [@PT / (K - @M)]$$

Observe que si no hay ningún ítem sin contestar @M=0 y por lo tanto PT no se corrige ya que multiplica por K y luego se divide por K.

Continuando con el ejemplo del test con K= 30 ítems, la puntuación total corregida PT del sujeto con puntuación @PT=16 que ha dejado @M=10 ítems sin contestar, vale:

$$PT = K \times [@PT / (K - @M)] = 30 \times 0.8 = 24 \text{ puntos}$$

Observe que esta fórmula equivale a añadir, a los 16 puntos obtenidos, 0.8 puntos por cada uno de los 10 ítems no contestados:

$$PT = 16 + 10 \times 0.8 = 16 + 8 = 24 \text{ puntos}$$

Seguidamente aplicaremos estos conceptos para calcular la puntuación total de nuestra escala de Salud.

¿Qué algoritmo se puede usar para calcular PT?

Hay diferentes formas de calcular PT, pero la más habitual es **recodificar los ítems** para que los códigos coincidan con la puntuación de cada respuesta y luego calcular la puntuación total sumando todos los ítems **con la función SUM** del procedimiento COMPUTE. Esta función permite invalidar la suma cuando el número de ítems contestados es inferior al **número mínimo de ítems válidos establecido** y darle valor *system missing*.

El hecho de que cada ítem contenga su puntuación permite directamente aplicar las técnicas psicométricas para validar el test (análisis de ítems, cálculos de consistencia interna, etc.).

Cálculo de la puntuación de salud (PS)

A continuación se aplicarán los conceptos expuestos para calcular la puntuación PS que cada sujeto ha obtenido en la escala de salud de nuestro estudio. El punto de corte para invalidar la escala lo situaremos en el 75%, lo que equivale a admitir **como mínimo 5 ítems contestados**, es decir, se admitirá sólo 1 ítem sin contestar, ya que 4 ítems contestados representarían sólo el 67% ($4/6=0.67$) de respuestas.

En primer lugar se recodifican los 3 hábitos negativos en 3 nuevas variables en las que los códigos coinciden con las puntuaciones.

⚡ La recodificación **no se realiza sobre las mismas variables** para evitar los errores que se producirían si por cualquier razón se ejecutara más de una vez el procedimiento RECODE:

```
RECODE H4 H5 H6 (1=2) (2=1) (3=0) INTO H4r H5r H6r.
```

En segundo lugar calcularemos la puntuación obtenida sumando los 6 ítems con la restricción que se necesitan como mínimo 5 valores válidos para realizar la suma. Para ello se ejecutará el siguiente COMPUTE:

```
COMPUTE @PS = SUM.5 (H1, H2, H3, H4r, H5r, H6r) . ← .5 (mínimo número de valores válidos)
```

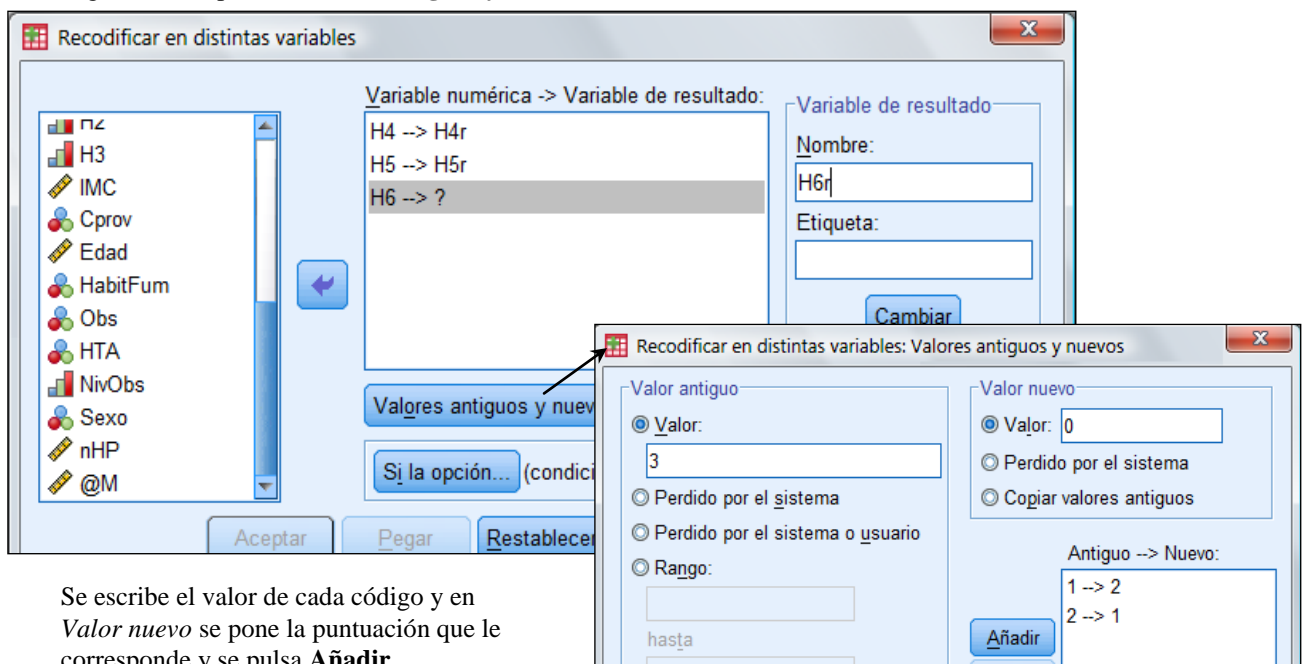
En tercer lugar se aplicará la corrección por ítems no contestados. Para ello es necesario crear la variable @M con el número de valores desconocidos de cada sujeto con el procedimiento COUNT. A partir de esta variable, que ya ha sido creada, la puntuación de salud se obtendrá así:

```
COMPUTE PS = 6 * @PS / (6 - @M) .
```

Y finalmente asignaremos las propiedades a la variable PS teniendo en cuenta que puede tener valores decimales.

85. Para recodificar los tres hábitos negativos se escoge la opción de menú Transformar | Recodificar en distintas variables...

Se pasan las variables H4 H5 H6 a la lista *Variables numéricas*. Seguidamente marcar la variable H4, en el cuadro de texto *Nombre* se escribe el nuevo nombre H4r y se pulsa **Cambiar**. Se repite la misma operación para H5 y luego para H6 sin olvidar pulsar Cambiar para transferir el nombre. Seguidamente pulsar **Valores antiguos y nuevos...**



Se escribe el valor de cada código y en *Valor nuevo* se pone la puntuación que le corresponde y se pulsa **Añadir**.


Se sale pulsando **Continuar** y luego se pulsa **Pegar** para añadir el RECODE a la ventana de sintaxis.

86. Se ejecuta la siguiente sintaxis para recodificar las variables H4 H5 H6:
- ```
RECODE H4 H5 H6 (1=2) (2=1) (3=0) INTO H4r H5r H6r.
EXECUTE.
```

⚡ No vamos a definir las propiedades de estas tres nuevas variables porque en este Tutorial sólo las usaremos como auxiliares para crear la puntuación total.

Sin embargo, si se tratara de un estudio para validar una escala, deberíamos definir sus propiedades porque estas variables recodificadas son las que se usan para realizar los diferentes análisis psicométricos de la escala, y en este caso las etiquetas son especialmente útiles.

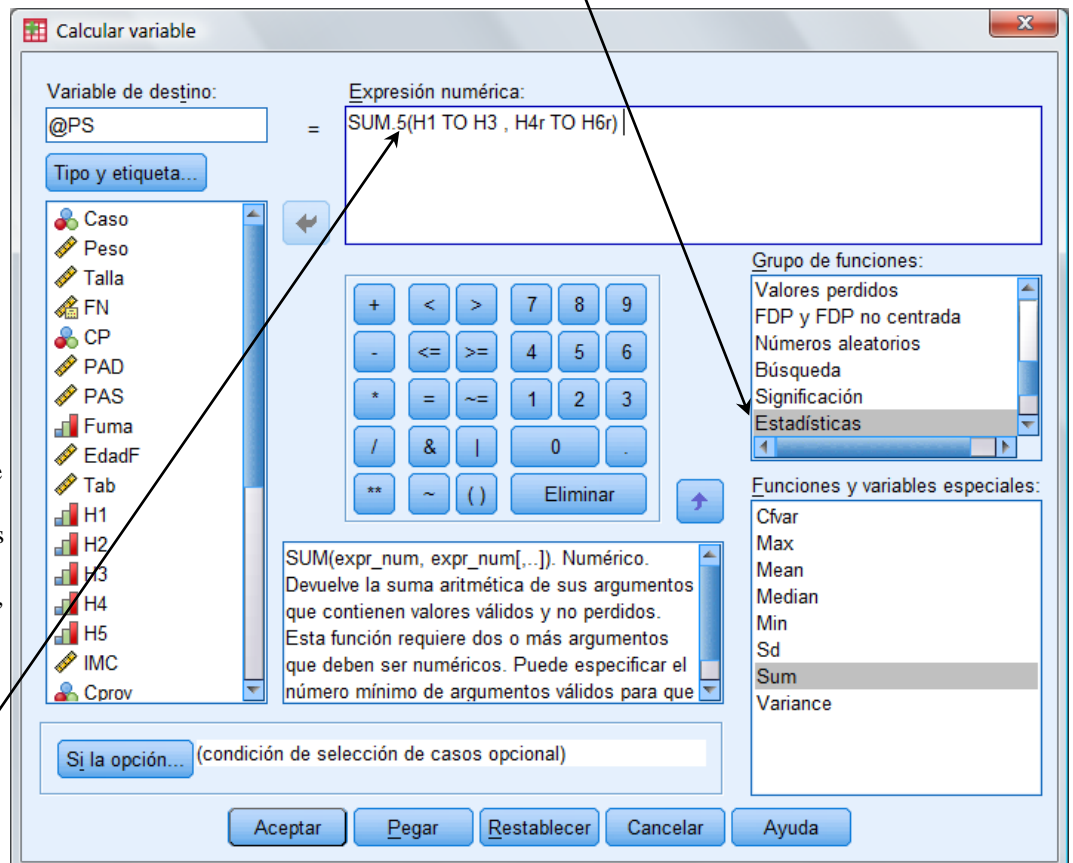
87. Ahora vamos a calcular la puntuación de salud @PS sumando las variables H1 a H3 y H4r a H6r.

Abrir el menú **Transformar | Calcular...** y escribir en *Variable destino* @PS. Seguidamente se marca el *Grupo de funciones Estadísticas* y se selecciona la función **Sum** y se transfiere con el botón 

El primer ? se sustituye por H1 TO H3 (las variables comprendidas entre H1 y H3)

El segundo ? se sustituye por H4r TO H6r (las variables comprendidas entre H4r y H6r) tal como ilustra esta imagen.

La restricción de necesitar como mínimo 5 valores válidos para efectuar la suma, se debe escribir .5 entre SUM y el paréntesis sin dejar ningún espacio en blanco.



Pulsar **Pegar** para añadir el COMPUTE a la ventana de sintaxis.

```
COMPUTE @PS=SUM.5 (H1 TO H3,H4r TO H6r) .
EXECUTE.
```



La función SUM requiere que los nombres de las variables estén separados por comas; esto es muy engorroso cuando el test tiene muchos ítems ya que además de pasar el nombre de cada ítem hay que escribir una coma para separarlo. El **truco** de escribir las listas de variables consecutivas con la palabra clave TO es **fundamental** cuando el test tiene muchos ítems.

88. Se ejecuta la sintaxis anterior para crear la puntuación @PS:

La imagen presenta los valores obtenidos por los 4 primeros sujetos de la matriz de datos y permite comprobar que @PS se crea correctamente.

| H1 | H2 | H3 | H4r | H5r | H6r | @M   | @PS   |
|----|----|----|-----|-----|-----|------|-------|
| 2  | 0  | 2  | 0   | 0   | 0   | .0   | 4,00  |
| 0  | 1  | 0  | .   | 2   | .   | 2,00 | .     |
| 2  | .  | 2  | 2   | 2   | 2   | 1,00 | 10,00 |
| .  | .  | .  | .   | .   | .   | 6,00 | .     |
| 0  | 1  | 0  | 0   | 1   | 1   | .0   | 3,00  |

← Sujeto con todos los ítems contestados  
 ← Sujeto no puntuado (4 ítems contestados)  
 ← Sujeto puntuado (5 ítems contestados)  
 ← Sujeto no puntuado (0 ítems contestados)



89. Abrimos el cuadro de diálogo *Calcular variable* para crear la variable PS con el valor de la puntuación @PS corregida por ítems sin respuesta. Para escribir el denominador pulsar **()** para insertar los paréntesis.

90. Pulsamos **Pegar** para añadir el COMPUTE a la ventana de sintaxis.

```
COMPUTE PS=6*@PS / (6-@M) .
EXECUTE.
```

Se ejecuta y en la ventana de datos verificamos en los primeros sujetos que PS se ha creado correctamente. Observe que la puntuación del tercer sujeto que había contestado todos los ítems en el sentido de salud excepto uno de ellos, se corrige de 10 al máximo valor (PS=12).

| @M   | @PS   | PS    |
|------|-------|-------|
| ,0   | 4,00  | 4,00  |
| 2,00 | .     | .     |
| 1,00 | 10,00 | 12,00 |
| 6,00 | .     | .     |
| ,0   | 3,00  | 3,00  |
| 0    | 7,00  | 7,00  |

En la ventana de resultados aparece el siguiente aviso:

```
>Warning # 511
>A division by zero has been attempted on the indicated command.
>The result has been set to the system-missing value.
>Command line: 640 Current case: 4 Current splitfile group: 1
```

Advierte que al crear la variable PS se ha producido una división por 0 en el caso que está en cuarto lugar y que el resultado se le ha dado valor *system missing*. Observe que este sujeto no había contestado ningún ítem (@M=6) y por lo tanto el denominador de la fórmula que realiza la corrección vale 0:  $(6 - @M) = 6 - 6 = 0$ .

91. Sólo quedan por definir las propiedades de PS con la ventana *Definir propiedades de variables*. Puesto que PS está corregida puede tomar valores decimales. Así pues, le asignaremos 1 decimal.

Salimos pegando la sintaxis y se ejecuta:

```
VARIABLE LABELS PS 'Puntuación de salud'.
FORMATS PS (F4.1) .
```

92. Finalmente borraremos las variables auxiliares @M y @PS porque no se necesitarán más. Puesto que el procedimiento de borrar variables **no está disponible a través de menús**, se debe abrir la ventana de sintaxis y escribir:

```
DELETE VARIABLES @M @PS. ← No olvide el punto final
```

- ! No deben escribirse los nombres de las variables para evitar errores. Se deben **pegar desde la ventana Utilidades | Variables** de esta forma: se pone el cursor al final de la instrucción DELETE VARIABLES, se selecciona @M y @PS y se pulsa **Pegar**. Las dos variables se insertan en el lugar donde estaba el cursor. Escriba el **punto al final**, ejecute la sintaxis y compruebe que estas dos variables se han borrado de la ventana de datos.

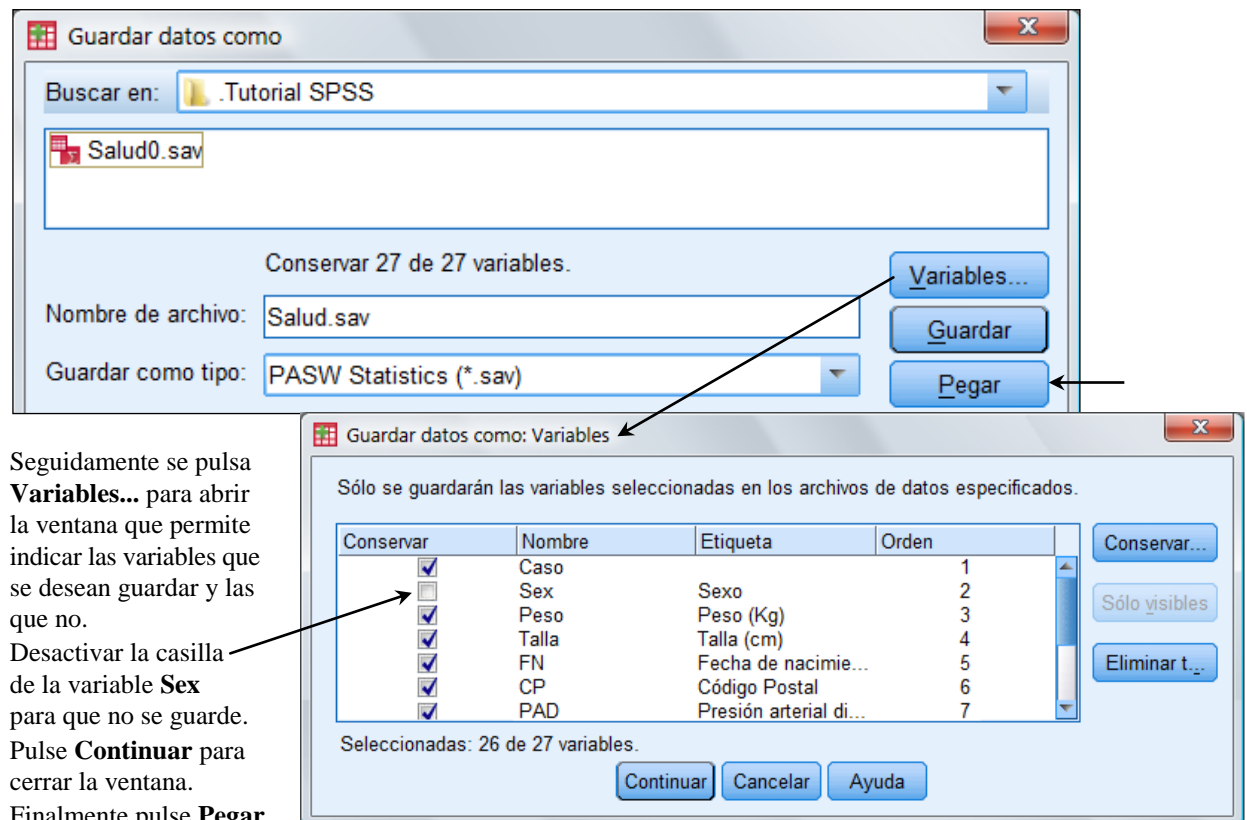
Seleccionar las variables y pulsar Pegar

## 4.8 Final de la etapa de creación de variables: guardar las ventanas de datos y sintaxis

Se ha terminado la importante etapa de preparación de los datos para el análisis estadístico. Terminada la **Fase 1** (descrita en el apartado 2.7 Esquema de un proceso de datos con *SPSS Statistics*, p. 13), se deben guardar todos los datos con el nombre **Salud.sav** (excepto la variable **Sex**, duplicada al haber sido recodificada en la variable Sexo) y la sintaxis con el nombre **Salud.sps**.

93. Finalizado el proceso de creación de las variables necesarias para contestar las hipótesis de nuestro estudio, se deberá guardar la información de la tabla de datos (Datos + Diccionario con las propiedades).

Desde la ventana de datos, se escoge la opción de menú Archivo | Guardar como... y en la ventana *Guardar datos como* se selecciona la carpeta donde se guardan los datos y se escribe el nombre del archivo **Salud** (la extensión **.sav** la asigna *SPSS Statistics* automáticamente porque identifica que es un archivo de datos).



Seguidamente se pulsa **Variables...** para abrir la ventana que permite indicar las variables que se desean guardar y las que no.

Desactivar la casilla de la variable **Sex** para que no se guarde. Pulse **Continuar** para cerrar la ventana. Finalmente pulse **Pegar**.

94. En la ventana de sintaxis se ha pegado:
- ```
SAVE OUTFILE='C:\...\Tutorial SPSS\Salud.sav'  
/DROP=Sex  
/COMPRESSED.
```
- Observe que contiene la instrucción **DROP=Sex** para indicar que se deben guardar todas las variables excepto **Sex**. Ejecútelos y compruebe que se ha guardado el archivo **Salud.sav** en la carpeta especificada.

Atención. La variable **Sex** continúa en la ventana de datos. Lo que ha hecho esta instrucción es **no guardarla** en el archivo **Salud.sav**, de manera que la próxima vez que se abra **Salud.sav** ya no aparecerá **Sex**.

95. Seguidamente se elige la opción de menú Archivo | Guardar desde el editor de sintaxis para guardar la sintaxis completa. En este caso no es necesario dar nombre al archivo porque ya tiene el nombre de **Salud.sps**.

96. Finalmente escoja la opción de menú Archivo | Salir. Es probable que pregunte ¿Quiere guardar el contenido de la ventana de Resultados?

Contestar No y se cerrará directamente el programa porque previamente ya hemos guardado la ventana de Datos y la de Sintaxis.

4.9 Comprobación de la sintaxis creada

Se ha terminado la etapa de preparación de los datos para el análisis estadístico. Terminada la **Fase 1** (descrita en la pág. 13) y una vez guardado el archivo de sintaxis **Salud.sps** deberemos comprobar que no contiene ningún error.

Para ello debe hacer doble clic sobre **Salud.sps**: se abrirá *SPSS Statistics* con la ventana de datos vacía.

Seguidamente, en la ventana de sintaxis, debe elegir el menú **Ejecutar | Todo**. Si la sintaxis es correcta el primer procedimiento es el GET que leerá el Excel con los datos y el resto de procedimientos generarán de nuevo en la ventana de datos todas las variables definidas y si la sintaxis es correcta en la ventana de resultados no aparecerá ningún mensaje de error.

97. Abrir el archivo **Salud.sps**. Situar en la ventana de sintaxis y elegir el menú **Ejecutar | Todo**.

98. En primer lugar compruebe si en la ventana de datos aparecen datos. En caso de que esté vacía es probable que no haya encontrado los datos porque el archivo Excel está en una ubicación distinta a la que se especifica en el GET DATA. Si esto ocurre arregle la ruta, borre los mensajes de la ventana de resultados dejándola vacía y repita el paso anterior hasta que se lean los datos.

99. Una vez se ha llenado la ventana de datos, compruebe que en la ventana de resultados no aparece ningún mensaje de error, excepto el *Warning* de la división por 0 (véase punto 90, pág. 46).

Si aparecen errores debe localizarlos, arregle la sintaxis errónea, borre los mensajes de la ventana de resultados dejándola vacía y repita el proceso hasta lograr eliminar todos los errores.

Una vez eliminados todos los errores guarde la ventana de sintaxis con el mismo nombre.

Fin de Fase 1

5. Ordenación y Selección de registros

5.1 Ordenación de registros

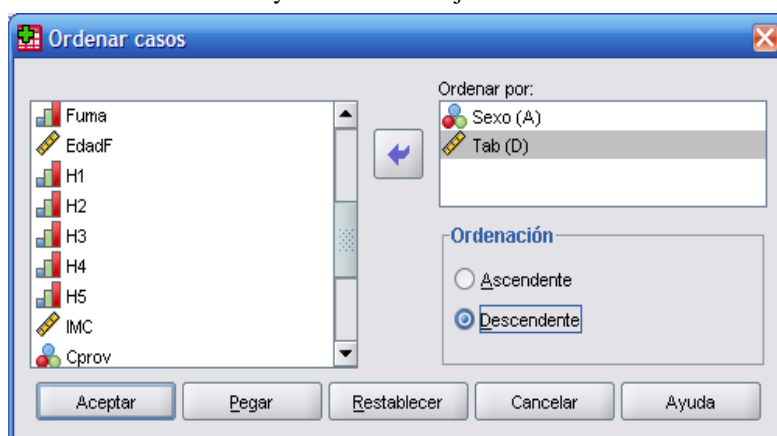
El procedimiento SORT CASES permite ordenar los registros de una tabla de datos de *SPSS Statistics*. Puede ser de interés ordenar los registros para generar un listado con los datos. Además, algunos procedimientos requieren que los datos estén ordenados previamente.

Las ordenaciones se realizan por defecto en *orden ascendente*, es decir, de A a Z, o desde el valor más bajo al más alto. También pueden ser en *orden descendente*, es decir, de Z a A, o desde el valor más alto al más bajo.

100. Abrir el archivo **Salud.sav** y comprobar que se ha incorporado en la ventana de datos. No abra el archivo de sintaxis **Salud.sps** porque a partir de este punto se guardará en un nuevo archivo que se llamará **SaludE.sps**.

101. Ordenar los registros de la tabla primero por *Sexo* en orden ascendente y dentro de los sujetos del mismo sexo por consumo de tabaco (*Tab*) en orden descendente.

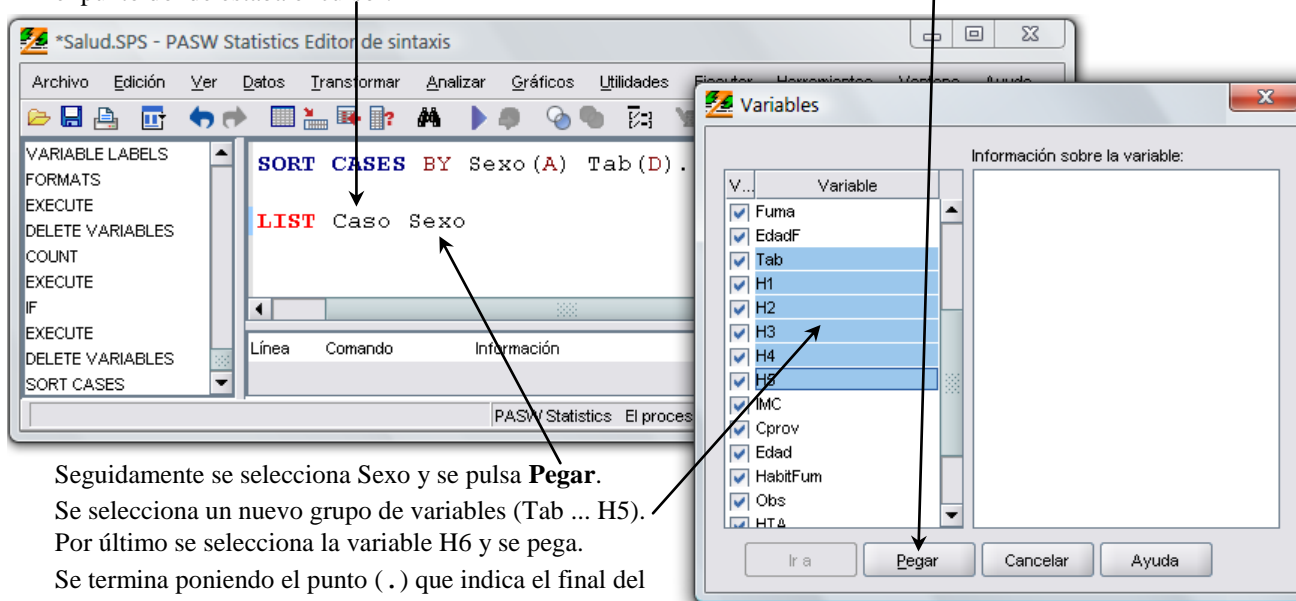
Seleccionar Datos | Ordenar casos... de la ventana de datos, y en la ventana *Ordenar casos* se pasan las dos variables de ordenación a la lista *Ordenar por*. El grupo de opciones *Ordenación* permite definir el orden de forma independiente para cada variable.



102. Se pulsa **Pegar** para obtener la sintaxis: **SORT CASES BY Sexo(A) Tab(D)**. Se ejecuta y se comprueba que se han ordenado los registros de la tabla de datos.

103. Se desea obtener un listado con los datos de las variables H1 a H6 incluyendo previamente las variables Caso, Sexo y Tab. Se realiza con el procedimiento LIST, que genera listados de variables en la ventana de resultados. Este procedimiento **no está disponible por menús** y por lo tanto **se debe escribir** en la ventana de sintaxis: se escribe el nombre del procedimiento (LIST) seguido de la lista de nombres de variables a listar y con un punto final (.) que es el finalizador de la instrucción.

Escriba en la ventana de sintaxis LIST (que es el nombre del procedimiento) y **deje el cursor** a continuación de LIST. Para evitar equivocarse al escribir los nombres de las variables se pueden "pescar"; para ello abra el cuadro Utilidades | Variables, escoja la primera variable (Caso) y al pulsar **Pegar** su nombre se escribe en el punto donde estaba el cursor:



Seguidamente se selecciona *Sexo* y se pulsa **Pegar**. Se selecciona un nuevo grupo de variables (*Tab ... H5*). Por último se selecciona la variable *H6* y se pega. Se termina poniendo el punto (.) que indica el final del procedimiento LIST:

LIST Caso Sexo Tab H1 H2 H3 H4 H5 H6.

104. Se ejecuta el procedimiento LIST y se obtiene el siguiente listado ordenado.

Caso	Sexo	Tab	H1	H2	H3	H4	H5	H6
162	.	25	1	2	.	2	2	3
91	.	0	2	.	2	1	1	1
145	0	35	1	1	2	2	2	3
604	0	35	1	1	2	3	2	3
743	0	30	0	2	2	1	2	3
302	0	25	0	1	0	3	3	3
622	0	25	0	1	0	3	1	3
200	0	20	0	0	2	2	3	3
97	0	15	0	1	0	3	2	2
132	0	15	0	2	2	2	2	2
836	0	15	1	2	0	3	2	2
239	0	10	1	1	1	2	2	2
129	0	0	2	0	2	3	1	1
130	0	0	0	0	1	1	1	1
327	0	0	0	1	0	1	2	1
330	0	0	2	0	2	1	1	1
395	0	0	0	0	1	1	1	1
406	0	0	0	2	2	3	2	1
447	0	0	1	1	2	2	2	1
497	0	0	1	2	.	2	2	1
539	0	0	0	0	2	2	3	1
640	0	0	0	1	0	1	2	1
662	0	0	2	0	2	1	1	1
668	0	0	0	0	1	1	1	1
792	0	0	1	1	2	2	2	1
84	0	.	0	1	0	.	1	.
92	1	99	9
161	1	40	2	2	1	1	1	3
214	1	30	0	1	1	2	3	3

! Los valores *sysmis* de *Sexo* quedan al principio al ordenar de forma ascendente. El valor 99 en *Tab* es el mayor valor posible aunque sea un *user missing* y queda al principio al ordenar de forma ascendente.

! El procedimiento LIST permite generar listados de forma rápida, p.e. para verificar el resultado de una transformación. El procedimiento SUMMARIZE (Analizar | Informes | Resúmenes de casos...) permite generar listados en forma de tablas.

105. Repetimos los puntos 101 y 102 para volver a ordenar los registros por Caso.

Sort Cases by Caso(A) .

Después de ejecutar esta sintaxis los registros de la ventana de datos recuperan el orden inicial.

5.2 Selección de registros

SPSS Statistics permite seleccionar qué registros de una tabla de datos están disponibles para los procedimientos estadísticos. Las alternativas más utilizadas son la selección **temporal** (procedimiento FILTER) y la selección **permanente** (procedimiento SELECT IF).

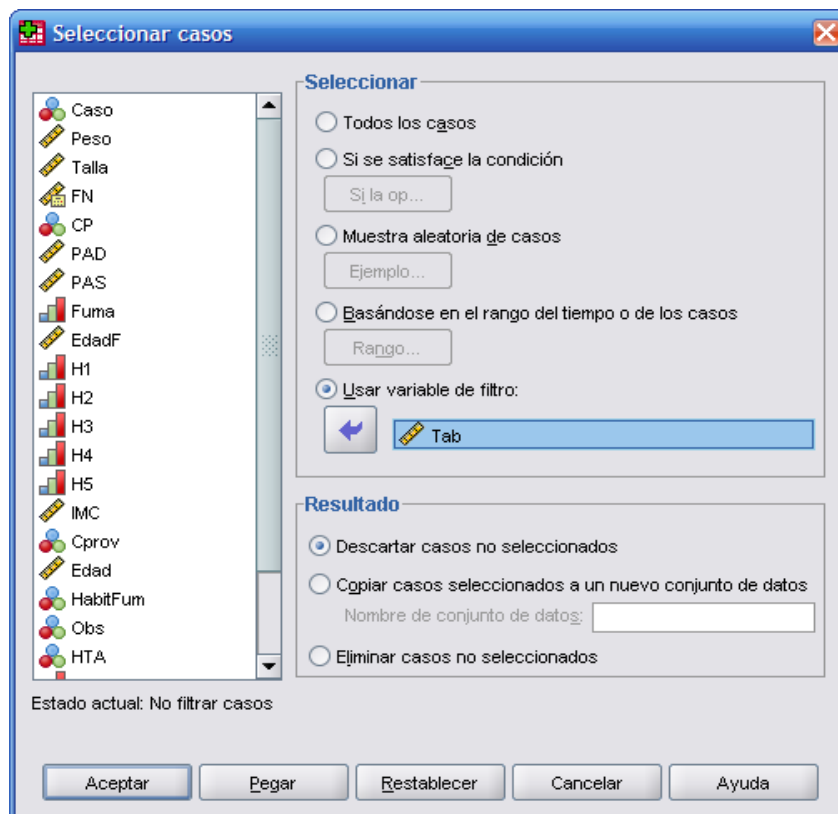
Selección temporal

El procedimiento FILTER permite activar una selección temporal de casos mediante un filtro basado en los valores de una variable de selección. Cuando el filtro está activado, todos los casos con valores 0 o *sysmis* en la variable de selección no intervienen en los procedimientos estadísticos que se ejecutan.

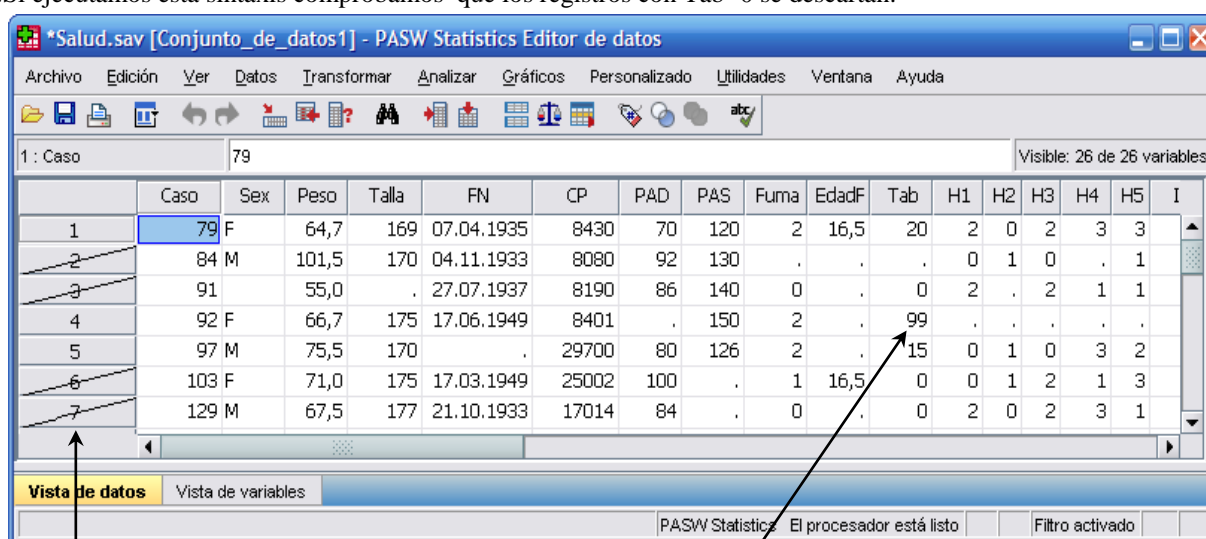
106. Si en la tabla hay una variable que es 0 o *sysmis* en los casos a filtrar, se puede usar como variable de filtro. Por ejemplo, si se desea filtrar la tabla para analizar sólo los fumadores, se usa la variable *Tab* como variable de selección porque vale 0 para los no fumadores y diferente de 0 para que fuman. Se elige la opción de menú Datos | Seleccionar casos... para abrir la ventana *Seleccionar casos*, se marca la opción *Usar variable de filtro* y se escoge *Tab* como variable de filtro.

107. Para realizar una selección temporal, marcar la opción **Descartar casos no seleccionados**. Pulsar el botón **Pegar** para escribir el procedimiento en la ventana de sintaxis:

```
USE ALL.
FILTER BY Tab.
EXECUTE.
```



108. Si ejecutamos esta sintaxis comprobamos que los registros con *Tab*=0 se descartan:



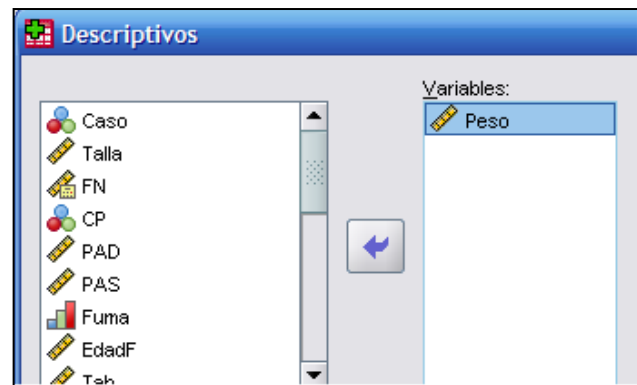
Los registros no seleccionados se descartan y quedan tachados

⚠ Los valores *usermis* no se filtran: el registro con *Tab*=99 no se filtra

Cuando hay un filtro activo está indicado en la barra de estado

109. Con un filtro activo, solo los registros no filtrados se utilizan para los cálculos estadísticos.
Para comprobarlo escoger: Analizar | Estadísticos descriptivos | Descriptivos... seleccionar la variable *Peso*, pegar y ejecutar la sintaxis:

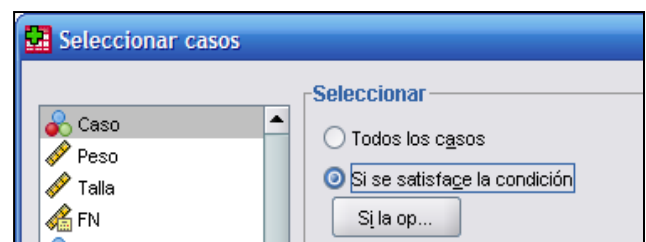
```
DESCRIPTIVES VARIABLES=Peso
/STATISTICS=MEAN STDDEV MIN MAX.
```



- ! SPSS Statistics ha realizado el cálculo sólo con los 21 fumadores. El filtro anterior continúa activado.

Estadísticos descriptivos					
	N	Mínimo	Máximo	Media	Desv. tip.
Peso (Kg)	21	57,5	92,3	69,024	9,8149
N válido (según lista)	21				

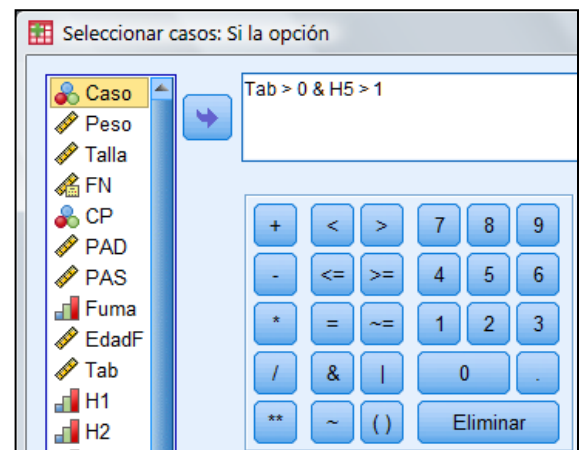
110. Si no existe una variable que directamente sirva como variable de filtro, el cuadro de diálogo permite escribir la condición que deben cumplir los registros para no ser filtrados. Abrir la ventana de *Seleccionar casos* (Datos | Seleccionar casos...) y marcar la opción *Si se satisface la condición*. Pulsar el botón *Si la op...* y se abre una ventana muy similar a la del COMPUTE para escribir la condición.



Vamos a seleccionar a los fumadores y bebedores de café; se escribe la condición:
 $Tab > 0 \text{ \& } H5 > 1$.

111. Cuando hemos escrito la condición pulsamos el botón **Continuar** y el botón **Pegar** para obtener el filtro en la ventana de sintaxis:

```
USE ALL.
COMPUTE filter_$=(Tab > 0 & H5 > 1).
VARIABLE LABELS filter_$ 'Tab > 0 & H5 > 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
```

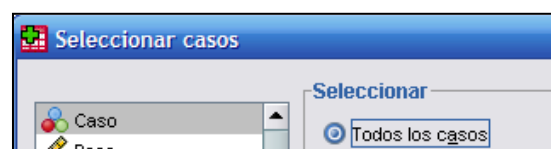


112. El asistente genera una nueva variable **filter_\$** que será 1 cuando se cumpla la condición y 0 cuando no se cumpla; esta variable es la que se utiliza para filtrar. Las líneas correspondientes a las instrucciones **VARIABLE LABEL**, **VALUE LABELS**, **FORMAT** definen las propiedades de la nueva variable **filter_\$** que no son realmente necesarias. Si se ejecuta la sintaxis y el anterior **DESCRIPTIVES** del *Peso* se obtiene la siguiente tabla que con los estadísticos de *Peso* de los 13 fumadores y bebedores de café:

Estadísticos descriptivos					
	N	Mínimo	Máximo	Media	Desv. tip.
Peso (Kg)	13	58,4	85,2	70,462	8,9897
N válido (según lista)	13				

113. Para desactivar el filtro, se abre la ventana *Seleccionar casos*, se escoge la opción *Todos los casos*, se pega y se ejecuta el procedimiento **FILTER OFF**.

```
FILTER OFF.
USE ALL.
EXECUTE.
```

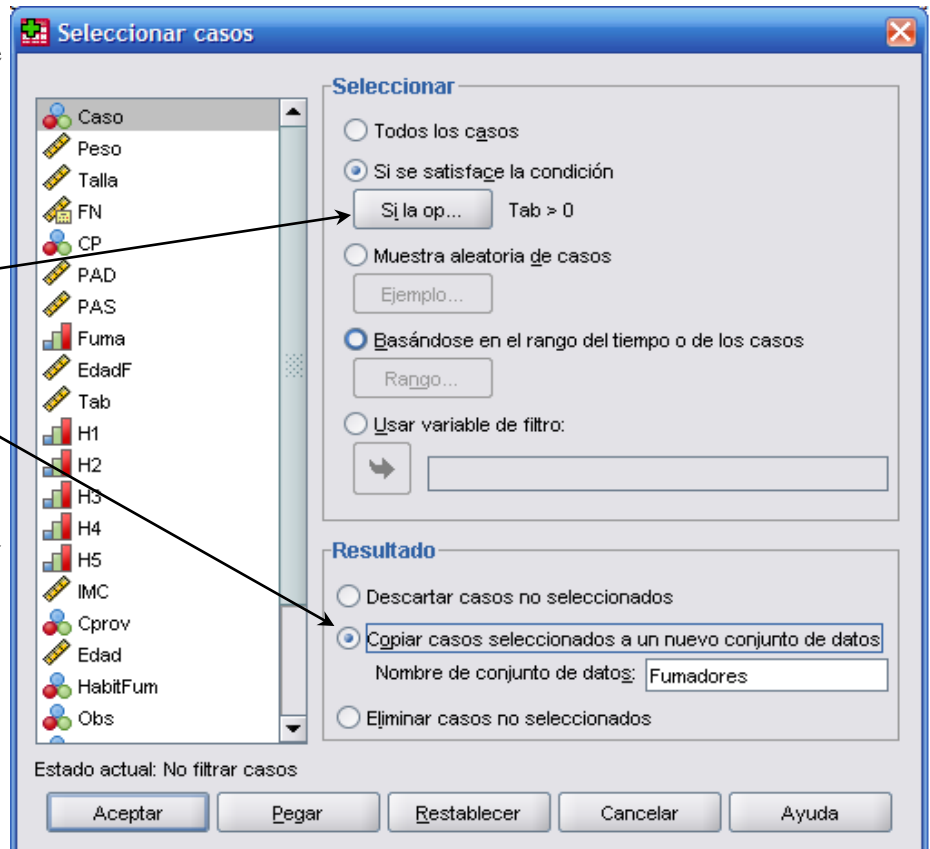


Selección permanente

El procedimiento SELECT IF selecciona de forma permanente el subconjunto de casos en los que una expresión lógica es verdadera. Los registros que no cumplen la expresión lógica se eliminan de la tabla de datos.

114. Seleccionar de forma permanente el subconjunto de fumadores. Para ello se abre el cuadro de diálogo **Datos | Seleccionar casos...** y se escoge la opción *Si se satisface la condición*. Pulsar el botón **Si la op...** y en la ventana que se abre se escribe la condición de fumador: *Tab > 0*.

115. Escoger la opción *Copiar casos seleccionados a un nuevo conjunto de datos* y poner el nombre *Fumadores*. Esta opción genera una nueva ventana de datos con los registros que cumplen la expresión lógica. La opción *Eliminar casos no seleccionados* también genera un SELECT IF, pero realiza una **selección permanente** con el riesgo de perder los datos. Es más segura la opción intermedia.



116. Pulsar **Pegar** y ejecutar la sintaxis.

Las instrucciones DATASET permiten gestionar varios conjuntos de datos abiertos a la vez. El procedimiento DATASET COPY crea una copia del conjunto de datos actual en una nueva ventana. Acto seguido DATASET ACTIVATE pasa el foco a esta ventana sobre la que se ejecuta el procedimiento SELECT IF.

```
DATASET ACTIVATE Conjunto_de_datos1.
DATASET COPY Fumadores.
DATASET ACTIVATE Fumadores.
FILTER OFF.
USE ALL.
SELECT IF (Tab > 0).
DATASET ACTIVATE Conjunto_de_datos1.
EXECUTE.
```

117. Tras ejecutar aparecerá una nueva ventana con el conjunto de datos *Fumadores* que contiene sólo los sujetos que fuman.

118. Elegir la opción de menú **Archivo | Guardar** para grabar el nuevo conjunto de datos como *Fumadores.sav* en el disco, pegar y ejecutar la sintaxis:

```
SAVE OUTFILE='C:\...\Fumadores.sav' /COMPRESSED.
```

Cerrar la ventana *Fumadores.sav* y seguimos trabajando con nuestros datos de *Salud* completos.

119. Ya ha finalizado este Capítulo y debe guardar el archivo de sintaxis y el de resultados.

Desde la ventana de sintaxis escoger la opción de menú Archivo | Guardar como..., seleccionar la carpeta **Tutorial SPSS** y guardarlo con el nombre **SaludE** pulsando el botón **Guardar**.

Seguidamente se accede a la ventana de Resultados, se escoge Archivo | Guardar como..., se selecciona la carpeta **Tutorial SPSS** y se guarda con el nombre **SaludE** pulsando el botón **Guardar**.

Finalmente se sale de *SPSS Statistics* con el menú Archivo | Salir sin guardar el archivo de datos porque no se desea incluir la variable **filter_\$** que se ha añadido en un ejercicio de este capítulo.

6. Descripción de datos

En este capítulo se tratarán los diferentes procedimientos estadísticos de descripción de datos.

6.1 Clasificación de los procedimientos descriptivos de SPSS Statistics

En función del tipo de variables (cuantitativas o categóricas) y de si se deben describir en toda la población o en subpoblaciones, tenemos diferentes procedimientos estadísticos disponibles:

	Variables a describir	
	Categóricas	Cuantitativas
Descripción en el conjunto de la población	FREQUENCIES	FREQUENCIES EXAMINE MEANS DESCRIPTIVES
Descripción en diferentes subpoblaciones	CROSSTABS	EXAMINE ...BY... MEANS ... BY...

120. Abrir el archivo **Salud.sav** y comprobar que se ha incorporado en la ventana de datos. **Abrir la sintaxis** con la opción de menú Archivo | Archivos usados recientemente y escoger primero **SaludE.sps** y luego **SaludE.spv**. Seguidamente tendrá abiertas las tres ventanas del Capítulo anterior para continuar con el Tutorial.

6.2 Procedimiento FREQUENCIES

Ordena los valores de las variables especificadas y produce tablas con frecuencias, porcentajes y porcentajes acumulados, representaciones gráficas y cálculo de índices estadísticos.

Descripción de una variable categórica ordinal

121. Vamos a describir la variable ordinal **Fuma** con el procedimiento FREQUENCIES.

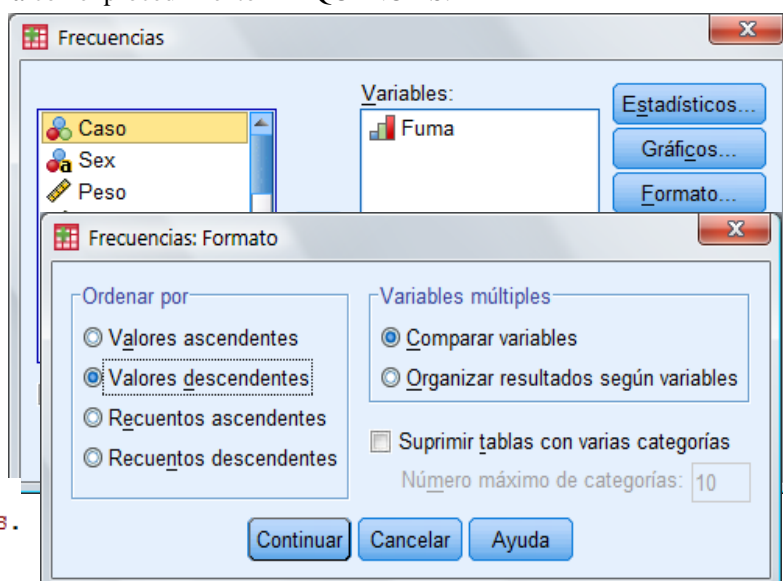
Abrir el cuadro de diálogo *Frecuencias* con la opción de menú Analizar | Estadísticos descriptivos | Frecuencias...

- seleccionar la variable **Fuma**.
- Pulsar **Formato** y en la ventana escoger *Ordenar por Valores descendentes* ya que se desea mostrar en primer lugar la categoría **Fumador** (codificada con el mayor de los tres códigos).

122. Pulsar **Continuar** para salir del cuadro *Formato* y finalmente pulsar **Pegar** para añadir el siguiente procedimiento

```
FREQUENCIES VARIABLES=Fuma  
/FORMAT=DVALUE /ORDER=ANALYSIS.
```

FREQUENCIES a la ventana de sintaxis:



123. Ejecute este procedimiento y compruebe que se obtiene la siguiente tabla:

¿Fuma o ha fumado?



Si no especifica *Orden descendente* en **Formato**, las categorías estarían en orden inverso y la categoría de riesgo **Fumador** (que interesa destacar) estaría en la tercera fila.

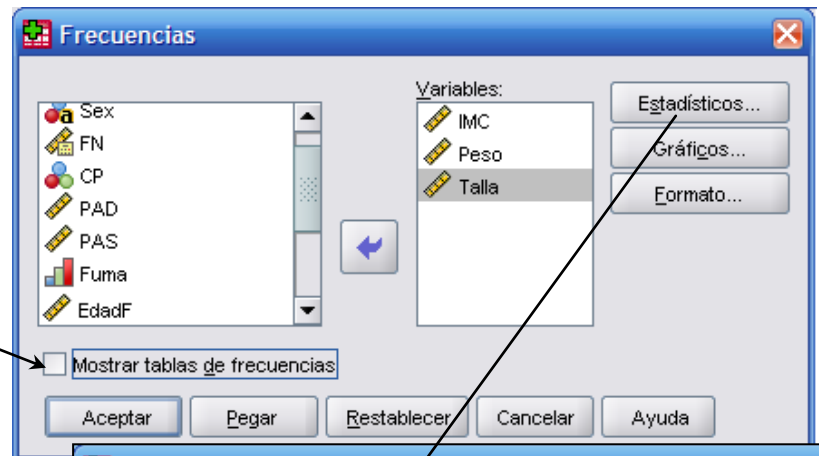
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Fumador	22	42,3	43,1	43,1
	Ex fumador	9	17,3	17,6	60,8
	No fumador	20	38,5	39,2	100,0
	Total	51	98,1	100,0	
Missing	System	1	1,9		
Total		52	100,0		

Descripción de una variable cuantitativa

124. También podemos utilizar el FRECUENCIES para describir variables cuantitativas.

Abrir el cuadro de diálogo *Frecuencias* desde el menú *Análizar | Estadísticos descriptivos | Frecuencias...* y seleccionar las variables IMC, Peso y Talla.

⚡ Con variables cuantitativas se debe **desactivar** la casilla *Mostrar tablas de frecuencias*.



125. Pulsar el botón **Estadísticos** y escoger los siguientes índices que corresponden a los **estadísticos de elección** para una buena descripción de las características estadísticas de variables cuantitativas: *Cuartiles*, *Media*, *Desviación típica*, *Mínimo* y *Máximo*.

126. Pulsa **Continuar** para salir del cuadro *Estadísticos* y finalmente pulsar **Pegar** para añadir este procedimiento a la ventana de sintaxis:

```
FRECUENCIES VARIABLES=IMC Peso Talla
/FORMAT=NOTABLE
/NTILES=4
/STATISTICS=STDDEV MINIMUM MAXIMUM MEAN
/ORDER=ANALYSIS.
```



127. Ejecutar este procedimiento y comprobar que se obtiene la siguiente tabla en la ventana de resultados:

Statistics				
		Índice de masa corporal (kg/m2)	Peso (Kg)	Talla (cm)
N	Valid	51	52	51
	Missing	1	0	1
Mean		22,783	67,852	172,82
Std. Deviation		2,9443	9,7839	6,147
Minimum		17,6	48,6	159
Maximum		35,1	101,5	189
Percentiles	25	21,078	61,550	168,00
	50	21,597	66,600	173,00
	75	24,475	71,950	177,00

Descripción de variables que contengan tiempos cumplidos

Las edades y tiempos transcurridos registrados en años, meses o semanas cumplidas introducen un sesgo de -0.5 unidades (años, meses o semanas) cuando se utilizan en cálculos estadísticos.

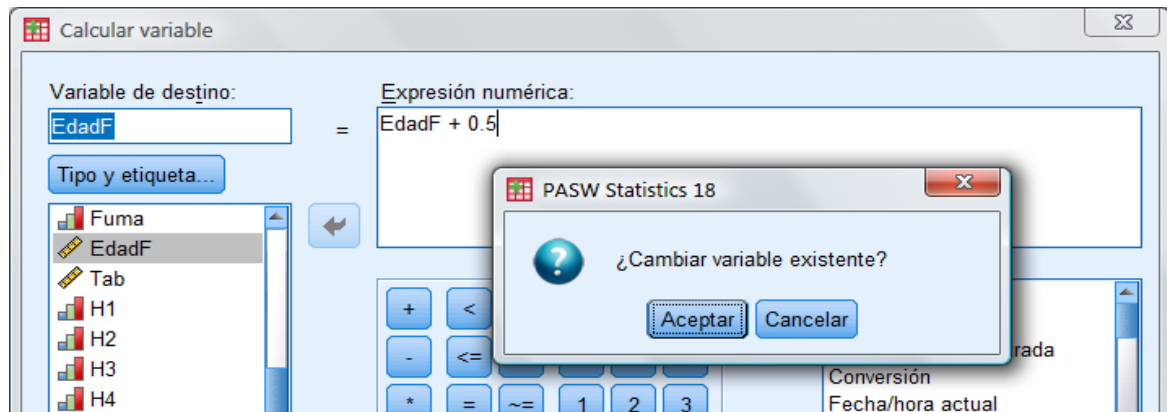
En efecto, la *edad de inicio de fumar* se ha registrado en años cumplidos (12; 13; 14; 15; ...) de manera que para un sujeto que empezó a fumar a los 14 años la variable contiene el valor 14; sin embargo, empezar a **fumar a los 14 años** cubre el intervalo de edades que van desde el día que cumplió los 14 hasta el día anterior a cumplir los 15. Puesto que no se conoce la fecha exacta, se considera que, en promedio, han empezado a fumar a los 14.5 años.

Para eliminar este sesgo que **afecta a todos los índices de posición** (media, mediana, percentiles, etc.), todas las variables medidas en tiempos cumplidos **se deben corregir** incrementándolas $+0.5$ unidades para calcular sin sesgo los índices de posición.

Para realizar esta corrección se puede corregir el sesgo incrementando en $+0.5$ los valores de la variable de forma permanente con el procedimiento COMPUTE. Otra opción es incrementarla de forma temporal para calcular los índices descriptivos; se realiza con un COMPUTE precedido del procedimiento TEMPORARY (no disponible a través de menús), tal como se explica en el siguiente ejercicio.

128. Para corregir el sesgo de la variable **EdadF** le sumaremos $+0.5$ años de forma temporal. Para ello debe escribir en la ventana de sintaxis: **TEMPORARY .** (no olvide el punto final).

Abra el cuadro de diálogo *Calcular variable* desde el menú Transformar | Calcular variable... Se escribe en *Variable de destino:* **EdadF** y en *Expresión numérica:* **EdadF + 0.5**. En este caso se modifica una variable que existe y al pulsar **Pegar** nos pedirá confirmación para cambiar la variable y pulsamos **Aceptar**.



129. Acceda a la ventana de Sintaxis y encontrará estos procedimientos:

Para que el TEMPORARY tenga efecto **debe borrar** el EXECUTE:

TEMPORARY .

COMPUTE EdadF=EdadF + 0.5 .

EXECUTE . ← **Borrarlo**

130. Abrir el menú Analizar | Estadísticos descriptivos | Frecuencias...,

seleccione la variable **EdadF**, desactive la casilla *Mostrar tablas de frecuencias*, pida los mismos estadísticos del ejercicio anterior y salga del cuadro con el botón **Pegar**.

131. Acceda a la ventana de Sintaxis y encontrará estos procedimientos. Compruebe que no haya errores (los nombres deben estar en azul). Márquelos todos, pulse ▶ para ejecutarlos y obtendrá la siguiente tabla en la ventana de resultados:

Edad a la que empezó a fumar (años)

N	Valid	26
	Missing	26
Mean		15,58
Std. Deviation		1,468
Minimum		13
Maximum		19
Percentiles	25	14,50
	50	15,50
	75	16,50

TEMPORARY .

COMPUTE EdadF=EdadF + 0.5 .

FRECUENCIES VARIABLES=EdadF

/FORMAT=NOTABLE

/NTILES=4

/STATISTICS=STDDEV MINIMUM MAXIMUM MEAN

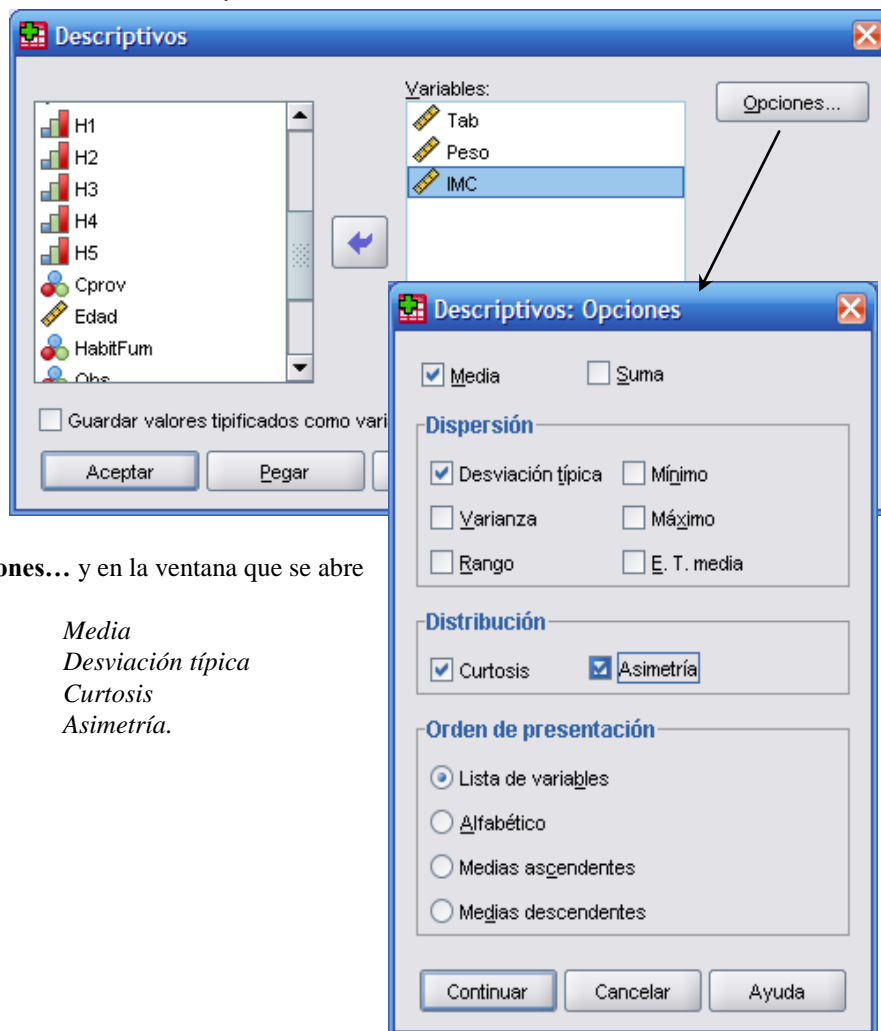
/ORDER=ANALYSIS .

⚠ **Atención.** Si se olvida de marcar el **TEMPORARY** o no está en color azul, el **COMPUTE** modificará de forma permanente la variable **EdadF**.

6.3 Procedimiento DESCRIPTIVES

Describe **variables cuantitativas** sólo con los índices estadísticos basados en momentos. A diferencia de la descripción anterior con FRECUENCIES, no incluye los estadísticos de orden (cuartiles y percentiles).

132. Escoger la opción de menú Analizar | Estadísticos descriptivos | Descriptivos ... para abrir el cuadro de diálogo y seleccionar las variables *Tab*, *Peso* y *IMC*.



Pulsamos el botón **Opciones...** y en la ventana que se abre marcar las casillas:

Media
Desviación típica
Curtosis
Asimetría.

133. Pulsa **Continuar** para salir del cuadro *Estadísticos* y luego pulsar **Pegar** para añadir este procedimiento a la ventana de sintaxis:

```
DESCRIPTIVES VARIABLES=Tab Peso IMC
  /STATISTICS=MEAN STDDEV KURTOSIS SKEWNESS.
```

134. Ejecutar este procedimiento y comprobar que se obtiene la siguiente tabla en la ventana de resultados:

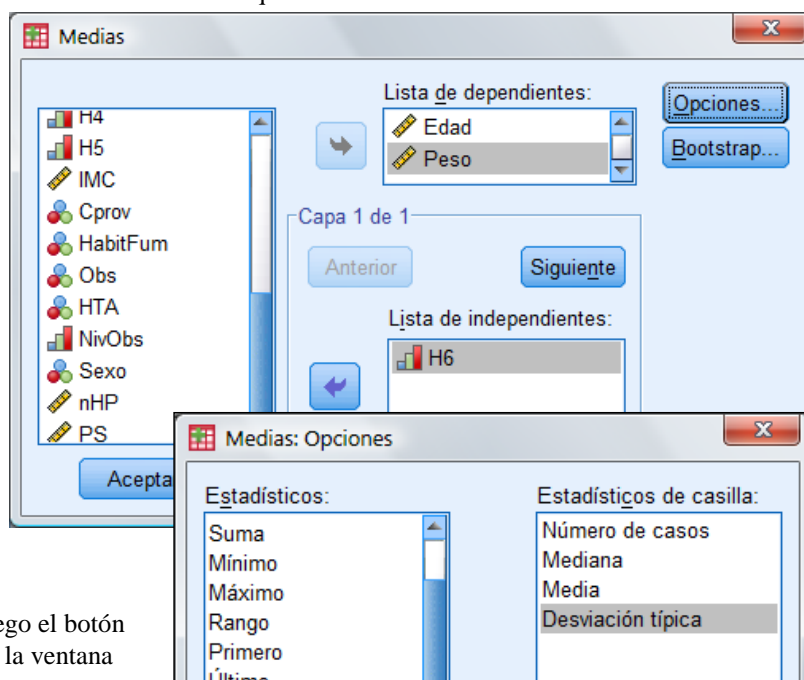
Descriptive Statistics									
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Tabaco (c/d)	50	0	40	9,40	12,562	,922	,337	-,593	,662
Peso (Kg)	52	48,6	101,5	67,852	9,7839	1,054	,330	2,093	,650
Índice de masa corporal (kg/m2)	51	17,6	35,1	22,783	2,9443	1,823	,333	5,085	,656
Valid N (listwise)	49								

6.4 Procedimiento MEANS

Calcula los estadísticos descriptivos de las variables especificadas. Si se especifica una o más variables de segmentación, los descriptivos se calcularán en cada uno de los grupos definidos por dichas variables.

135. Describir las variables *Edad* y *Peso* en las submuestras que define la variable *H6 Consumo de tabaco*.

Para ello abrir el cuadro de diálogo *Medias* con la opción de menú *Analizar | Comparar medias | Medias...*
 Seleccionar las variables *Edad*, y *Peso* en la *Lista de dependientes*.
 Seleccionar *H6* en la *Lista de independientes*.



136. Pulsar el botón **Opciones...** y escoger los estadísticos: *Número de casos*, *Mediana*, *Media* y *Desviación típica*.

137. Pulsar el botón **Continuar** y luego el botón **Pegar** para añadir el MEANS a la ventana de sintaxis:

```
MEANS TABLES=Edad Peso BY H6
/CELLS COUNT MEDIAN MEAN STDDEV.
```

138. Ejecutar este procedimiento y comprobar que se obtiene la siguiente tabla en la ventana de resultados:

Esta tabla se puede editar y trasponer filas por columnas para que las variables se sitúen en las filas y los estadísticos en las columnas.
 La forma de editarla se explica en el Capítulo 7 (p. 65) de este Tutorial.

La tabla una vez editada

Report

Consumo de tabaco		N	Median	Mean	Std. Desv.
Edad (años decimales)	Alto	15	69,58	68,13	4,82
	Bajo	4	68,64	68,04	3,85
	Medio	29	65,09	65,31	5,56
	Total	48	66,04	66,42	5,31
Peso (Kg)	Alto	15	66,90	69,29	10,88
	Bajo	5	71,10	69,22	7,82
	Medio	30	65,85	65,82	7,62
	Total	50	66,30	67,20	8,72

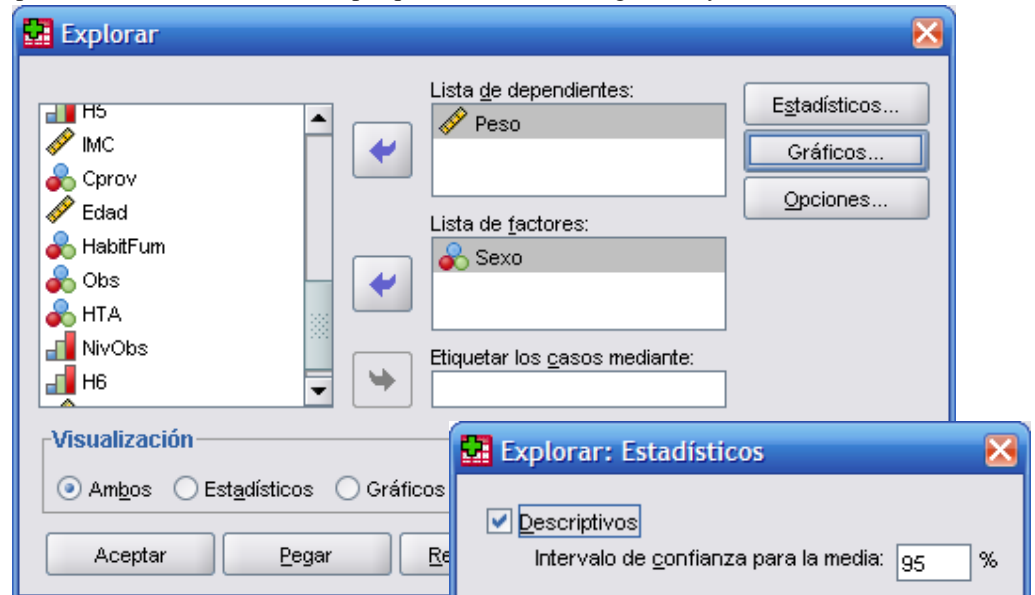
Report

Consumo de tabaco		Edad (años decimales)	Peso (Kg)
Alto	N	15	15
	Median	69,5825	66,900
	Mean	68,1349	69,287
	Std. Desv.	4,81729	10,8781
Bajo	N	4	5
	Median	68,6448	71,100
	Mean	68,0363	69,220
	Std. Desv.	3,84532	7,8184
Medio	N	29	30
	Median	65,0869	65,850
	Mean	65,3080	65,823
	Std. Desv.	5,55585	7,6193
Total	N	48	50
	Median	66,0356	66,300
	Mean	66,4188	67,202
	Std. Desv.	5,30747	8,7217

6.5 Procedimiento EXAMINE

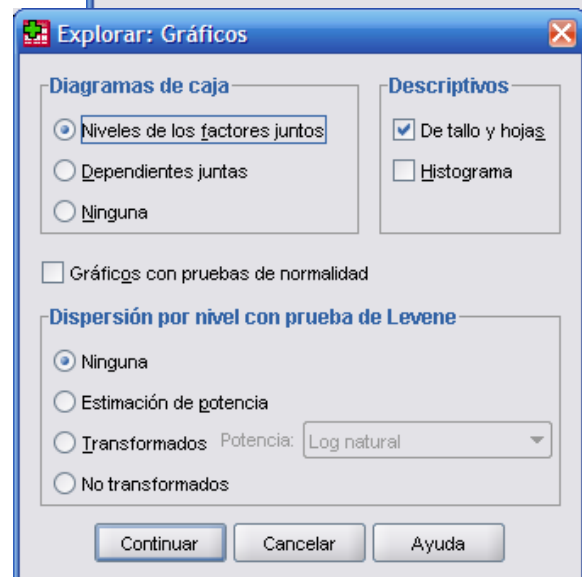
Describe los datos con técnicas de “Análisis exploratorio de datos” (EDA), entre las que se incluyen diferentes gráficos, los principales estadísticos basados en momentos y en ordenaciones, así como diferentes estadísticos robustos.

139. Describir la variable *Peso* en función de las submuestras que definen el *Sexo* de los sujetos utilizando el procedimiento EXAMINE. Para ello abrimos el cuadro de diálogo *Explorar* con la opción de menú Analizar | Estadísticos descriptivos | Explorar... y escogemos el *Peso* como variable dependiente y el *Sexo* como factor. Se marca la opción de *Visualización Ambos* porque se desea obtener gráficos y estadísticos:



140. Pulsar el botón **Estadísticos**. Por defecto se muestran los *Descriptivos* que es lo que nos interesa. Salir del cuadro pulsando **Continuar**.

141. Pulsar el botón **Gráficos**. Por defecto se generan los *Diagramas de caja* con los *Niveles de los factores juntos* y el gráfico de *Tallo y hojas*. **Desactivar** la casilla del gráfico de *Tallo y hojas*, porque no se quiere este gráfico. Salir del cuadro pulsando **Continuar**.



142. Pulsar el botón **Pegar** para obtener el EXAMINE en la ventana de sintaxis.

```
EXAMINE VARIABLES=Peso BY Sexo /PLOT BOXPLOT STEMLEAF /COMPARE GROUPS
/STATISTICS DESCRIPTIVES /INTERVAL 95 /MISSING LISTWISE
/NOTOTAL.
```

143. Ejecutar este procedimiento y comprobar que se obtiene la siguiente tabla en la ventana de resultados:

Descriptives				
Sexo			Statistic	Std. Error
Peso (Kg)	Masculino	Mean	73,817	1,9253
		95% Confidence Interval for Mean	Lower Bound	69,834
			Upper Bound	77,799
		5% Trimmed Mean	72,942	
		Median	71,300	
		Variance	88,962	
		Std. Deviation	9,4320	
		Minimum	63,1	
		Maximum	101,5	
		Range	38,4	
		Interquartile Range	9,8	
		Skewness	1,482	,472
		Kurtosis	2,125	,918
	Femenino	Mean	62,573	1,2673
		95% Confidence Interval for Mean	Lower Bound	59,963
			Upper Bound	65,183
		5% Trimmed Mean	62,447	
		Median	61,800	
		Variance	41,759	
		Std. Deviation	6,4621	
		Minimum	48,6	
		Maximum	79,6	
		Range	31,0	
		Interquartile Range	8,3	
		Skewness	,421	,456
		Kurtosis	1,018	,887

Gráficos de tallo y hojas

Peso (Kg) Stem-and-Leaf Plot for
Sexo= Masculino

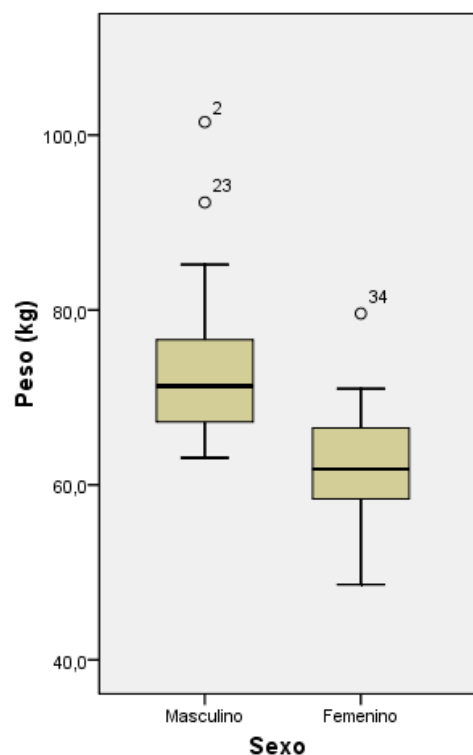
Frequency	Stem &	Leaf
1,00	6 .	3
10,00	6 .	5556677889
5,00	7 .	11234
3,00	7 .	567
2,00	8 .	24
1,00	8 .	5
2,00	Extremes	(>=92)

Stem width: 10,0
Each leaf: 1 case(s)

Peso (Kg) Stem-and-Leaf Plot for
Sexo= Femenino

Frequency	Stem &	Leaf
1,00	4 .	8
2,00	5 .	34
6,00	5 .	778899
10,00	6 .	0111123444
3,00	6 .	669
3,00	7 .	001
1,00	Extremes	(>=80)

Stem width: 10,0
Each leaf: 1 case(s)



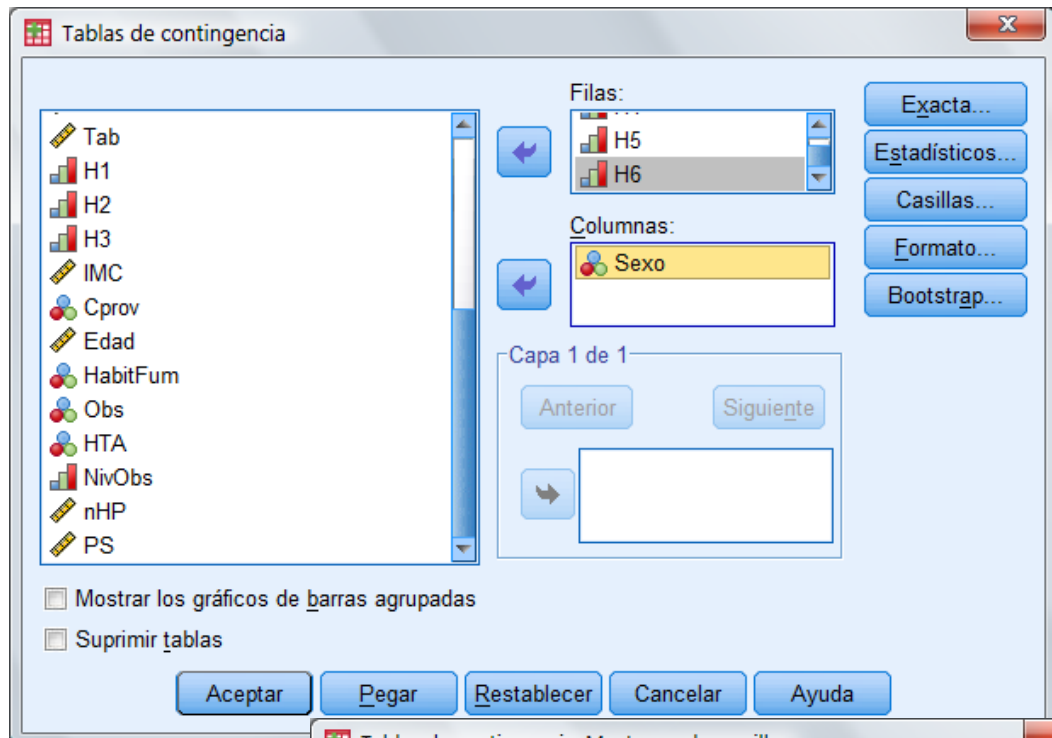
6.6 Procedimiento CROSSTABS

Construye tablas de contingencia que representan la distribución conjunta de dos (o más) variables categóricas.

144. Obtener las tablas de contingencia de las variables *H4 Consumo de alcohol*, *H5 Consumo de cafeína* y *H6 Consumo de tabaco* en función del *Sexo*. Para ello se debe escoger la opción de menú:

Analizar | Estadísticos descriptivos | Tablas de contingencia...

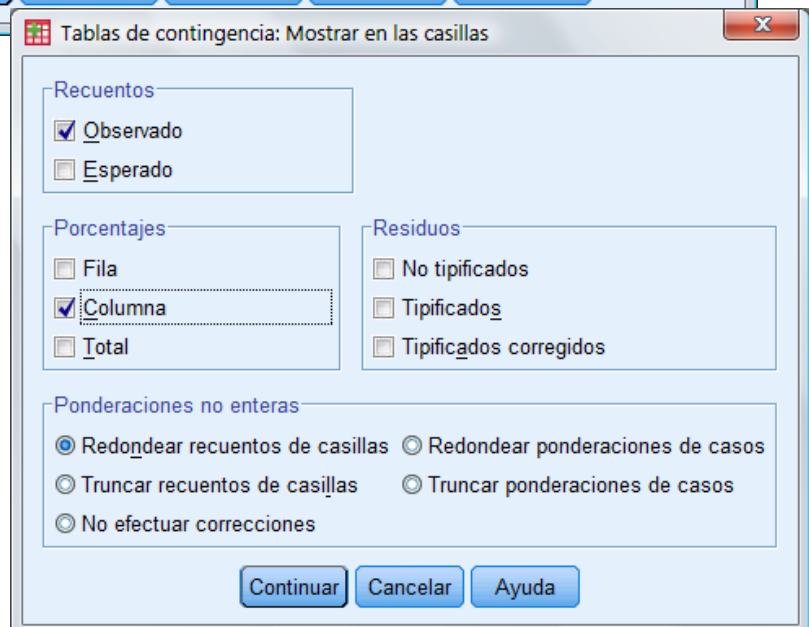
Situar las variables *H4*, *H5* y *H6* en las *Filas* y situar la variable *Sexo* en las *Columnas*:



El botón **Formato...** permite presentar en *orden descendente* las categorías de las variables situadas en las filas. En este caso no es necesario porque las categorías de riesgo que interesan destacar están codificadas con el código más bajo.

145. Pulsar el botón **Casillas...**

Por defecto la tabla presenta las frecuencias observadas. Marcar también la casilla *Columna* para que la tabla incluya los porcentajes de columna. Salir del cuadro pulsando **Continuar**.



146. Pulsar el botón **Pegar** para añadir el CROSSTABS a la ventana de sintaxis:

```
CROSSTABS
/TABLES=H4 H5 H6 BY Sexo
/FORMAT=AVALUE TABLES
/CELLS=COUNT COLUMN
/COUNT ROUND CELL.
```

147. Ejecutar este procedimiento y comprobar que se obtienen las siguientes tablas en la ventana de resultados:

Consumo de alcohol * Sexo Crosstabulation

			Sexo		Total
			Masculino	Femenino	
Consumo de alcohol	Bajo	Count	8	13	21
		% within Sexo	34,8%	52,0%	43,8%
	Medio	Count	8	4	12
		% within Sexo	34,8%	16,0%	25,0%
	Alto	Count	7	8	15
		% within Sexo	30,4%	32,0%	31,3%
Total	Count	23	25	48	
	% within Sexo	100,0%	100,0%	100,0%	

Consumo de cafeína * Sexo Crosstabulation

			Sexo		Total
			Masculino	Femenino	
Consumo de cafeína	Bajo	Count	8	10	18
		% within Sexo	33,3%	40,0%	36,7%
	Medio	Count	13	3	16
		% within Sexo	54,2%	12,0%	32,7%
	Alto	Count	3	12	15
		% within Sexo	12,5%	48,0%	30,6%
Total	Count	24	25	49	
	% within Sexo	100,0%	100,0%	100,0%	

Consumo de tabaco * Sexo Crosstabulation

			Sexo		Total
			Masculino	Femenino	
Consumo de tabaco	Bajo	Count	13	16	29
		% within Sexo	56,5%	64,0%	60,4%
	Medio	Count	4	1	5
		% within Sexo	17,4%	4,0%	10,4%
	Alto	Count	6	8	14
		% within Sexo	26,1%	32,0%	29,2%
Total	Count	23	25	48	
	% within Sexo	100,0%	100,0%	100,0%	

6.7 Copiar las tablas de la ventana de resultados en un documento Word

Para traspasarlas a Word basta con hacer clic con el botón derecho del ratón sobre la tabla pivote y escoger la opción *Copiar* (o Edición | Copiar) para pasarla al portapapeles de Windows; seguidamente situarse en el punto del documento Word donde se quiere insertar la tabla y elegir la opción Edición | Pegar (Ctrl+V) y la tabla se pegará como una **tabla de Word** y se podrá editar y modificar en el documento de Word.

En el Capítulo 7 (p. 72) se explica cómo pegar la tabla en forma de imagen escalable; en este caso conviene editar previamente la tabla para darle el formato de presentación deseado. La edición de tablas está explicada en el apartado 7.2 Edición de una tabla pivote (p. 66).

6.8 Final de la sesión: guardar las ventanas de sintaxis y resultados

Una vez realizados todos los cálculos estadísticos, se debe guardar la sintaxis con el nombre **SaludE.sps** y los resultados con el nombre **SaludE.spv**.

148. Desde la ventana de sintaxis se escoge la opción de menú Archivo | Guardar como..., se selecciona la carpeta **Tutorial SPSS** y se guarda con el nombre **SaludE** pulsando el botón **Guardar**.

Seguidamente se accede a la ventana de Resultados, se escoge Archivo | Guardar como..., se selecciona la carpeta **Tutorial SPSS** y se guarda con el nombre **SaludE** pulsando el botón **Guardar**.

Finalmente se sale de *SPSS Statistics* con el menú Archivo | Salir.

7. Edición de Resultados

En este capítulo se aprende a editar **tablas pivot** y **gráficos** que *SPSS Statistics* genera en la ventana de resultados.

7.1 Manipulación de tablas pivot y objetos gráficos

Las tablas pivot generadas por *SPSS Statistics* contienen los resultados de los análisis y son completamente editables y redimensionables. Vamos a ver cómo editar una tabla pivot estándar para adaptarla a nuestras necesidades para ser incluida en un informe o en un PowerPoint.

149. Abrir el archivo **Salud.sav** y comprobar que se ha incorporado en la ventana de datos. **Abrir la sintaxis** con la opción de menú Archivo | Archivos usados recientemente y escoger primero **SaludE.sps** y luego **SaludE.spv**. Seguidamente tendrá abiertas las tres ventanas del Capítulo anterior para continuar con el Tutorial.

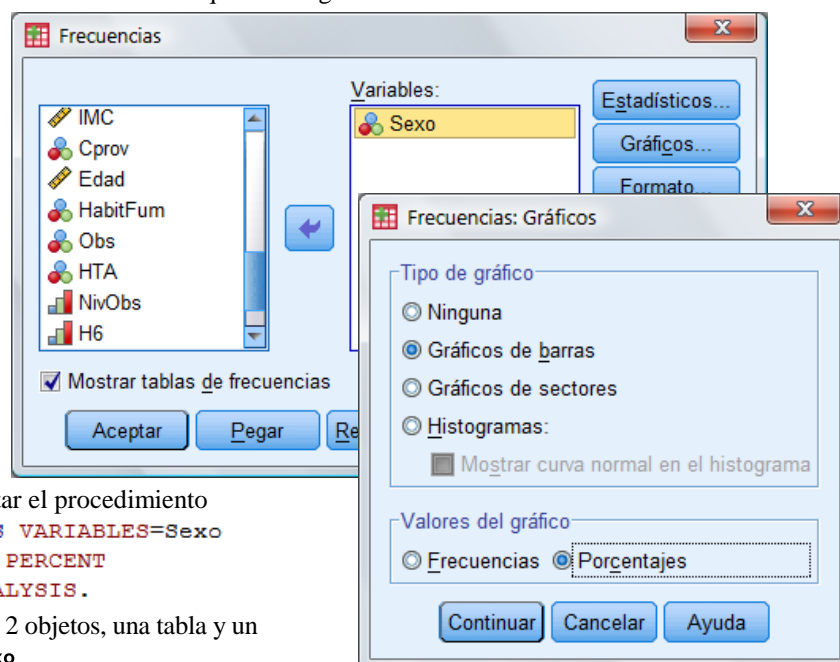
150. Crear con el procedimiento FREQUENCIES una tabla que contenga la distribución de sexos del estudio.

Abrir la opción de menú:
Analizar | Estadísticos
descriptivos | Frecuencias...

Transferir al cuadro *Variables:*
la variable **Sexo**.

Pulsar el botón **Gráficos...**
y elegir *Gráficos de barras*
y como valores *Porcentajes*.
Salir del cuadro pulsando
Continuar.

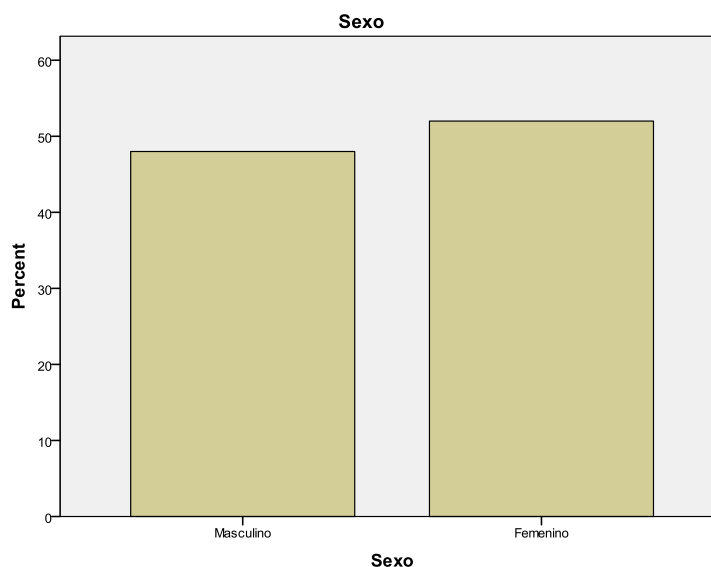
Salir del cuadro *Frecuencias*
pulsando **Pegar**.



151. Abrir la *ventana de sintaxis* y ejecutar el procedimiento
FREQUENCIES: **FREQUENCIES VARIABLES=Sexo**
/BARCHART PERCENT
/ORDER=ANALYSIS.

Se crean en la *Ventana de resultados* 2 objetos, una tabla y un
gráfico:

		Sexo			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Masculino	24	46,2	48,0	48,0
	Femenino	26	50,0	52,0	100,0
	Total	50	96,2	100,0	
Missing	System	2	3,8		
Total		52	100,0		



7.2 Edición de una tabla pivote

El objetivo de este apartado es editar la tabla del FRECUENCIES para transformarla en la tabla de la derecha:

Sexo					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Masculino	24	46,2	48,0	48,0
	Femenino	26	50,0	52,0	100,0
	Total	50	96,2	100,0	
Missing	System	2	3,8		
Total		52	100,0		

→

Sexo		
	Frequency	Percent
Masculino	24	48,0%
Femenino	26	52,0%
Total	50	100,0%

- 152.** Para empezar a editar la tabla, hacer doble clic sobre ella.
Notará que está en modo edición porque cambia el borde de la tabla. Para salir del modo edición, basta con hacer clic en cualquier punto de la ventana de resultados que esté fuera de la tabla.

Sexo					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Masculino	24	46,2	48,0	48,0
	Femenino	26	50,0	52,0	100,0
	Total	50	96,2	100,0	
Missing	System	2	3,8		
Total		52	100,0		

- 153.** Eliminar la columna **Percent**: Marque todos sus valores y pulse la tecla suprimir.

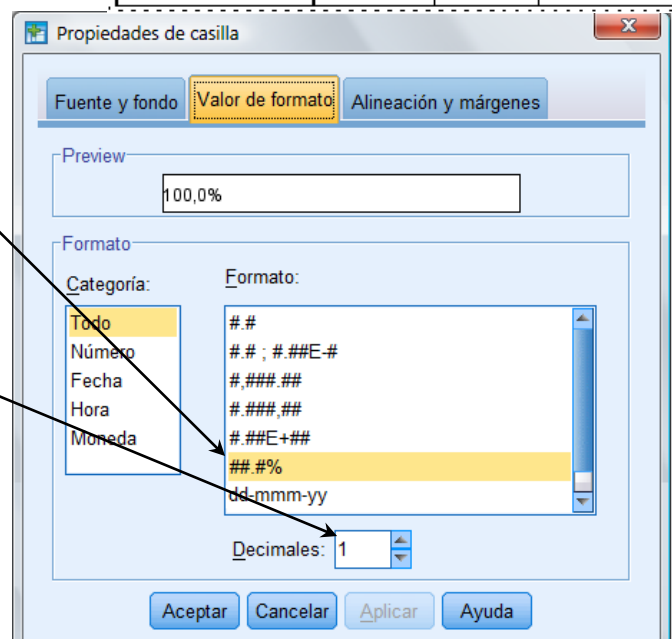
Sexo					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Masculino	24	46,2	48,0	48,0
	Femenino	26	50,0	52,0	100,0
	Total	50	96,2	100,0	
Missing	System	2	3,8		
Total		52	100,0		

- 154.** Modificar el texto de la columna **Valid Percent** borrando **Valid**. Para ello haga doble clic sobre la casilla y borre **Valid**

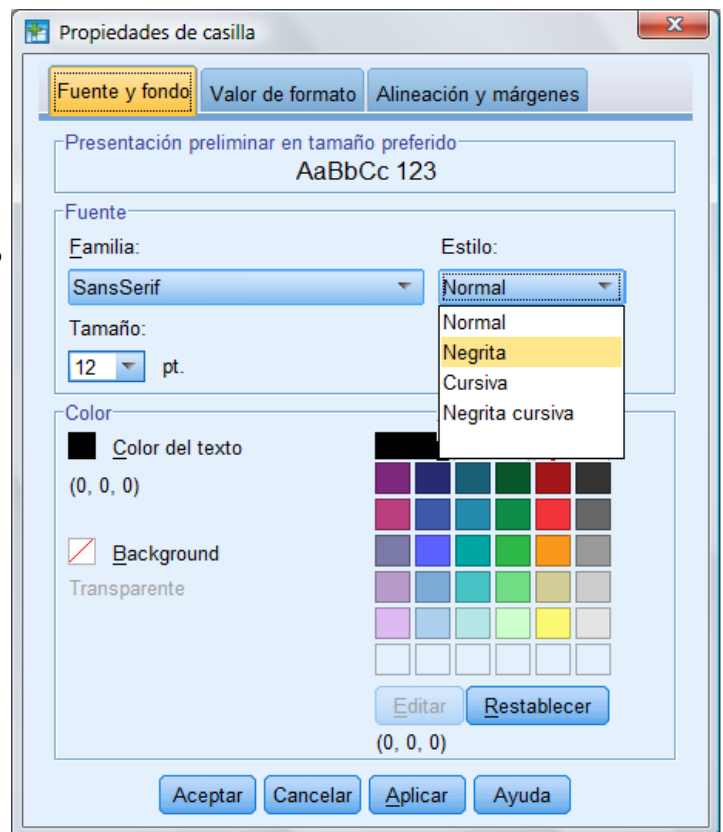
Sexe				
		Frequency	Percent	Cumulative Percent
Valid	Masculi	24	48,0	48,0
	Femeni	26	52,0	100,0
	Total	50	100,0	
Missing	System	2		
Total		52		

Sexe			
	Frequency	Percent	Cumulative Percent
Valid	Masculi	24	48,0
	Femeni	26	52,0
	Total	50	100,0
Missing	System	2	
Total	52		

- 155.** Asignar a los valores de la columna **Percent** formato %. Marque todos sus valores, abra con el botón derecho del ratón el menú contextual, elija la opción **Propiedades de casilla...**, asígneles formato **##.##%** y pulse **Aceptar**. Compruebe que los valores de los porcentajes incluyen el símbolo %.
- Esta pestaña permite escoger el formato de visualización del dato contenido en la celda; si es el resultado de un procedimiento estadístico, también permite aumentar o disminuir el número de decimales.



156. Marque la casilla correspondiente al porcentaje de varones (48,0%), abra el menú contextual, elija la opción **Propiedades de casilla...**, abra la pestaña **Fuente y fondo** y asígnele Estilo **Negrita**.
Esta pestaña permite cambiar la fuente y las propiedades del texto de la celda y el color, tanto del texto como del fondo.
Salga del cuadro pulsando **Aceptar**.



157. Eliminar la columna **Cumulative Percent**: Marque todos sus valores y pulse la tecla suprimir.
158. Eliminar las dos últimas filas (**Missing** y **Total**): Marque todos sus valores y pulse la tecla suprimir.

Sexo			
		Frequency	Percent
Valid	Masculino	24	48,0%
	Femenino	26	52,0%
	Total	50	100,0%
Missing	System	2	
Total		52	

159. Ocultar el texto **Valid**: Arrastre con el ratón el borde derecho de la casilla hasta que quede oculto.

Sexo			
		Frequency	Percent
Valid	Masculino	24	48,0%
	Femenino	26	52,0%
	Total	50	100,0%

Ocultar

Obtendrá finalmente la tabla con el formato deseado.

160. Vamos a copiar la tabla en un documento de Word.

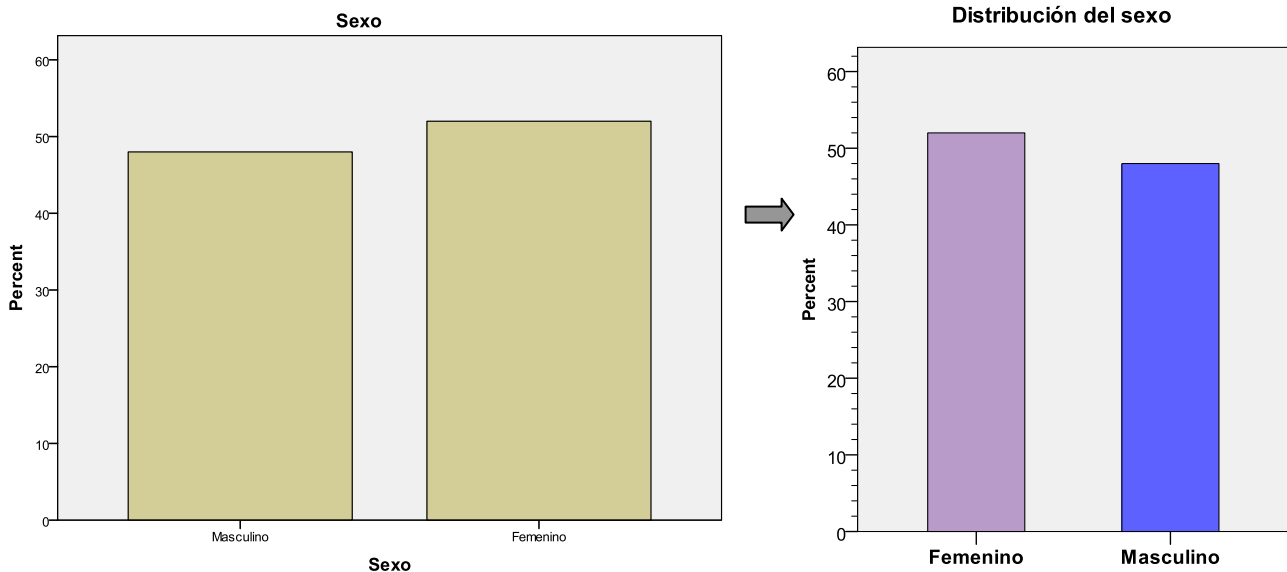
- Hacer clic con el botón derecho del ratón sobre la tabla pivote y escoger la opción *Copiar* (o Edición | Copiar) para pasarla al portapapeles de Windows.
- Abrir un documento de Word. Elija la opción Edición | Pegar (Ctrl+V) y la tabla se pegará como una tabla de Word y se podrá editar y modificar en el documento de Word.
- Ahora vamos a pegar la tabla en forma de imagen escalable. Abrir una nueva línea al final de la tabla y si es Word 2003 o anterior, escoger la opción Edición | Pegado especial... | Imagen (metarchivo mejorado). Con Word 2007, la secuencia es: Inicio | Pegar | Pegado especial... | Imagen (metarchivo mejorado).

Compruebe que se ha pegado una imagen exacta de la tabla y que se puede variar su tamaño.
Este objeto también se puede copiar al PowerPoint.

Sexo		
	Frequency	Percent
Masculino	24	48,0%
Femenino	26	52,0%
Total	50	100,0%

7.3 Edición de gráficos

Los gráficos se pueden editar para darles el aspecto deseado. El objetivo de este apartado es editar el gráfico de barra obtenido con el FRECUENCIAS para darle el aspecto del gráfico de la derecha:



161. Hacer doble clic sobre el gráfico para editarlo y aparecerá la ventana del *Editor de gráficos* que permite modificar los elementos de un gráfico ya creado utilizando los menús y la barra de botones.

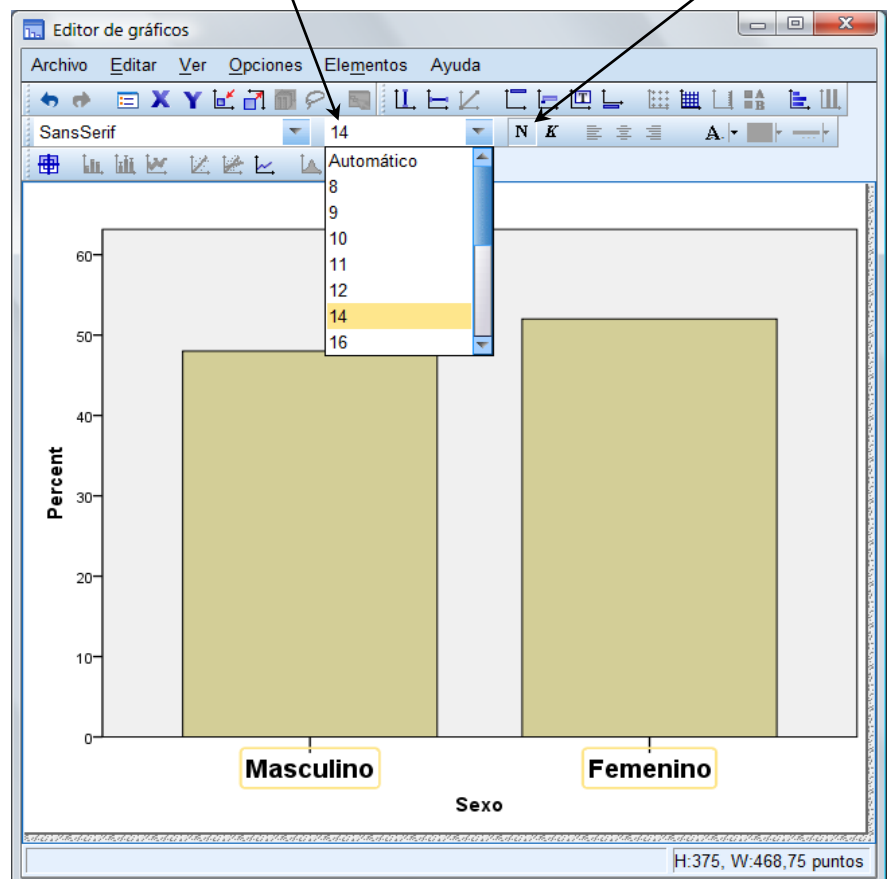
162. Vamos a aumentar el tamaño de las etiquetas. Seleccionar las categorías de Sexo, pulsar el botón negrita y desplegar la lista con el tamaño de fuente para asignarle 14pt.

Repetir la misma operación con la etiqueta **Percent** pero asignarle un tamaño de 13pt.

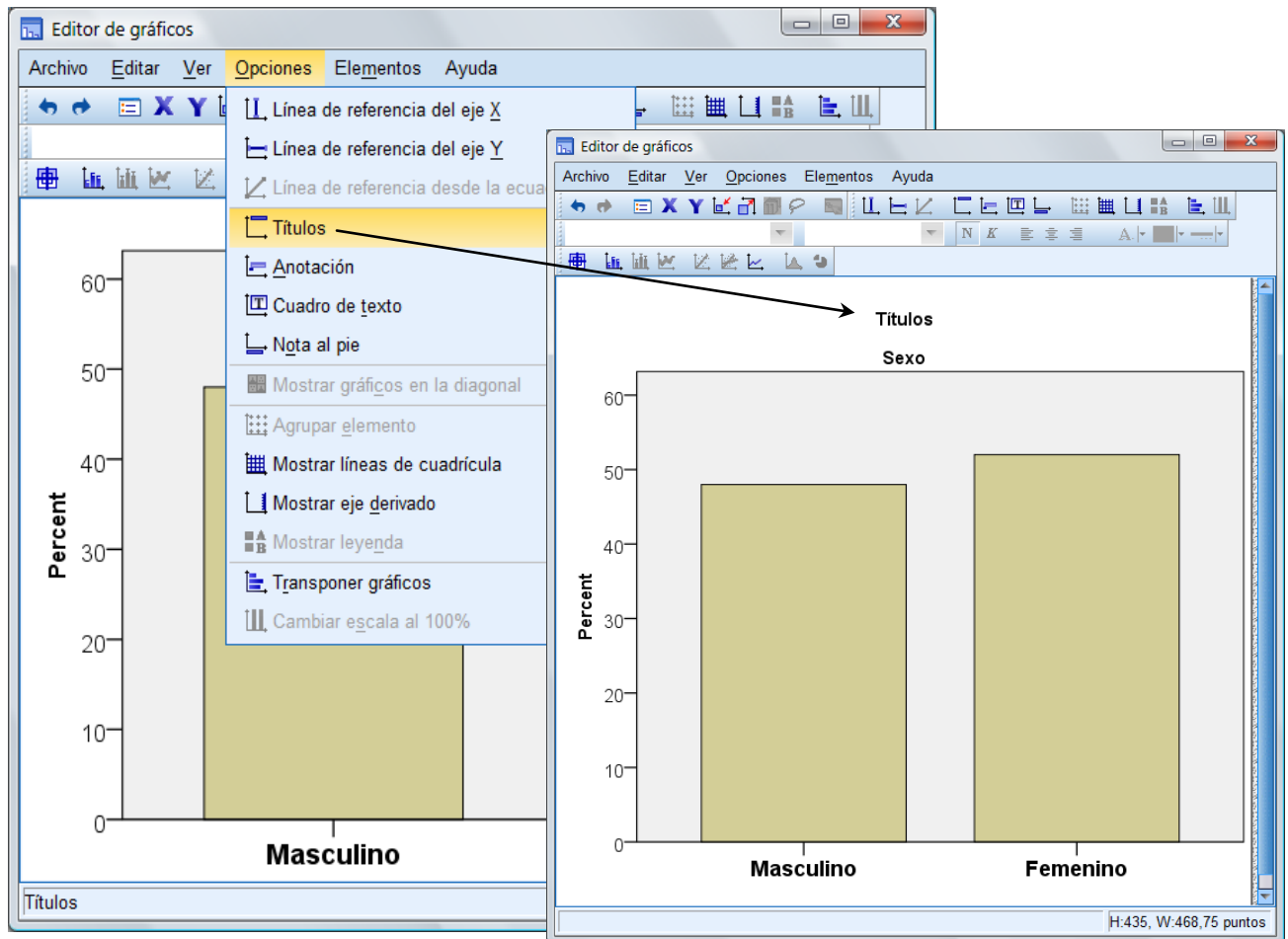
Seleccionar los números y asignarle un tamaño de 13pt (sin negrita).

Seleccionar el título de gráfico **Sexo** y asignarle tamaño de 16pt y negrita.

Finalmente se debe borrar el nombre de la variable que está en la parte inferior del gráfico. Hacer un clic para señalar la etiqueta **Sexo** y luego un segundo clic para pasar a forma edición del texto y borrar la etiqueta.



163. Cambiar el título del gráfico. Se puede editar el título (**Sexo**) con un clic para pasar a forma edición y poner como título **Distribución de sexo**; pero con la versión actual del editor, el título así modificado queda junto a las barras y es difícil separarlo de ellas. Lo más fácil es insertar un nuevo título con Opciones | Títulos.

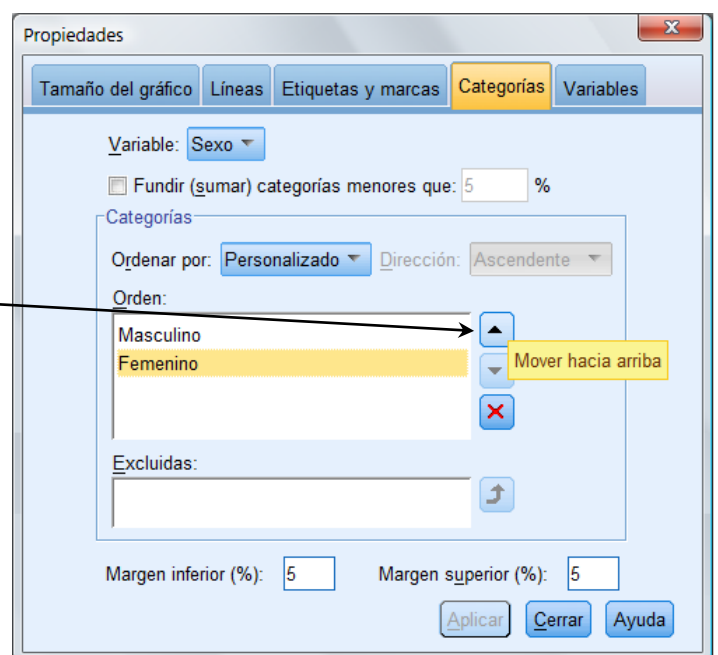



Hacer clic sobre el texto **Títulos** y sustituirlo por **Distribución de sexo**. Seguidamente asignarle tamaño 16pt. Finalmente se borra el título original **Sexo** tal como se explicó en el punto anterior.

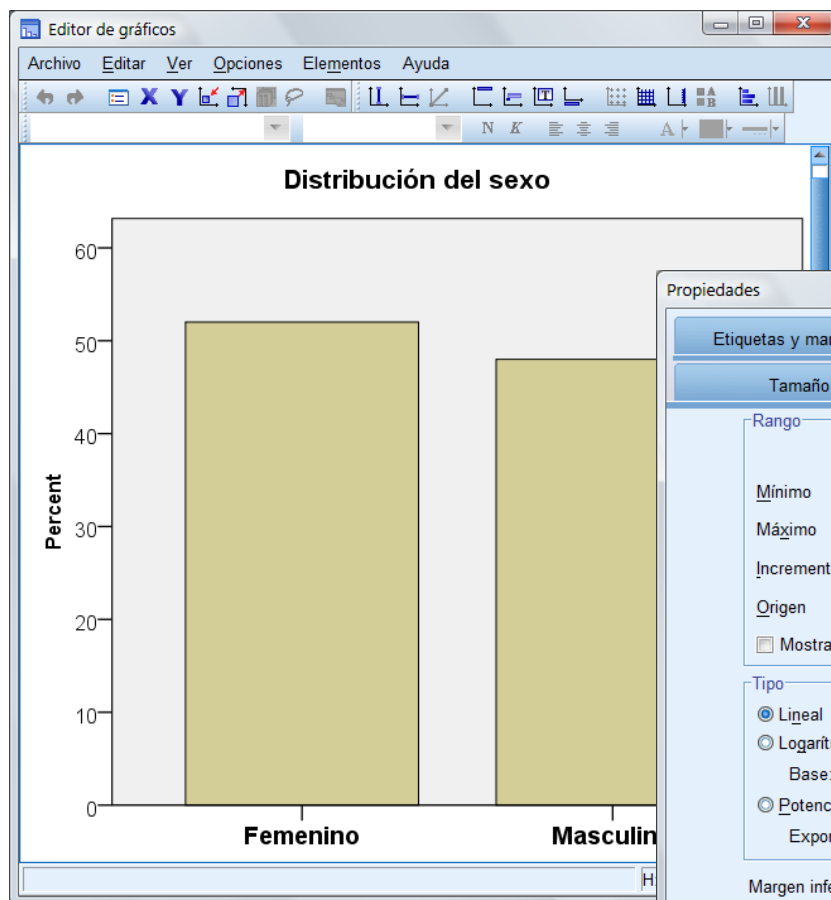
164. Pulsar el botón **X** para abrir la ventana con las propiedades del *Eje X* que permite definir: el tamaño, el tipo de líneas que dibuja, las etiquetas y marcas, los márgenes derecho e izquierdo del gráfico respecto a lo bordes del objeto, cambiar el orden de las categorías, etc.

Seleccione la categoría **Femenino** y póngala antes de **Masculino** pulsando el botón correspondiente.

Pulse **Aplicar** para cambiar el orden y luego **Cerrar** para salir de este cuadro.



165. Vamos a modificar la escala del Eje Y pulsando el icono  para abrir la ventana de *Propiedades* del eje Y. En la pestaña **Escala** se podría cambiar la graduación de manera que se incremente de 5 en 5 (Incremento mayor 5) y los valores mínimos y máximos que deben mostrar los ejes. También permite cambiar el formato de los números y la posibilidad de añadir marcas adicionales.



A la casilla Margen superior debe asignarle un 3%. Esto indica que el eje Y no acabe en el valor 60 sino que sea un 3% más largo. El margen inferior dejarlo en 0% porque interesa que el eje empiece exactamente en 0. Confirme pulsando **Aplicar**.

	Automático	Personalizado	Datos
Mínimo	<input checked="" type="checkbox"/>	0	48
Máximo	<input checked="" type="checkbox"/>	60	52
Incremento mayor	<input checked="" type="checkbox"/>	10	
Origen	<input checked="" type="checkbox"/>	0	

☐ Mostrar línea en origen

Tipo:
☒ Lineal
☐ Logarítmico (Base: 10, Seguro)
☐ Potencia (Exponente: 0,5, Seguro)

Margen inferior (%): 0 Margen superior (%): 3

Aplicar **Cancelar** **Ayuda**

☒ Mostrar título del eje Mostrar eje en: Por defecto

Etiquetas de incremento mayor:
☒ Mostrar etiquetas
 Orientación de la etiqueta: Automática

Ubicación de etiquetas de categoría:
☒ Automática
☐ Personalizado

Marcas mayores:
☒ Mostrar marcas
 Estilo: Fuera

Marcas menores:
☒ Mostrar marcas
 Estilo: Fuera
 Número de marcas menores por marcas mayores: 4

Aplicar **Cerrar** **Ayuda**

En la pestaña **Escala** se ha especificado que las marcas principales empiezan por 0% y se van incrementando un 10% hasta el 60%.

En la pestaña **Etiquetas y marcas** debe activar la casilla **Mostrar marcas** del grupo **Marcas menores** y especificar 4 marcas menores. Confirme pulsando **Aplicar**.



En la graduación del eje Y aparecerán 4 marcas intermedias que indicarán incrementos del 2%.


⚠ En SPSS Statistics 19 no aparecen las marcas menores, aunque se definan correctamente.

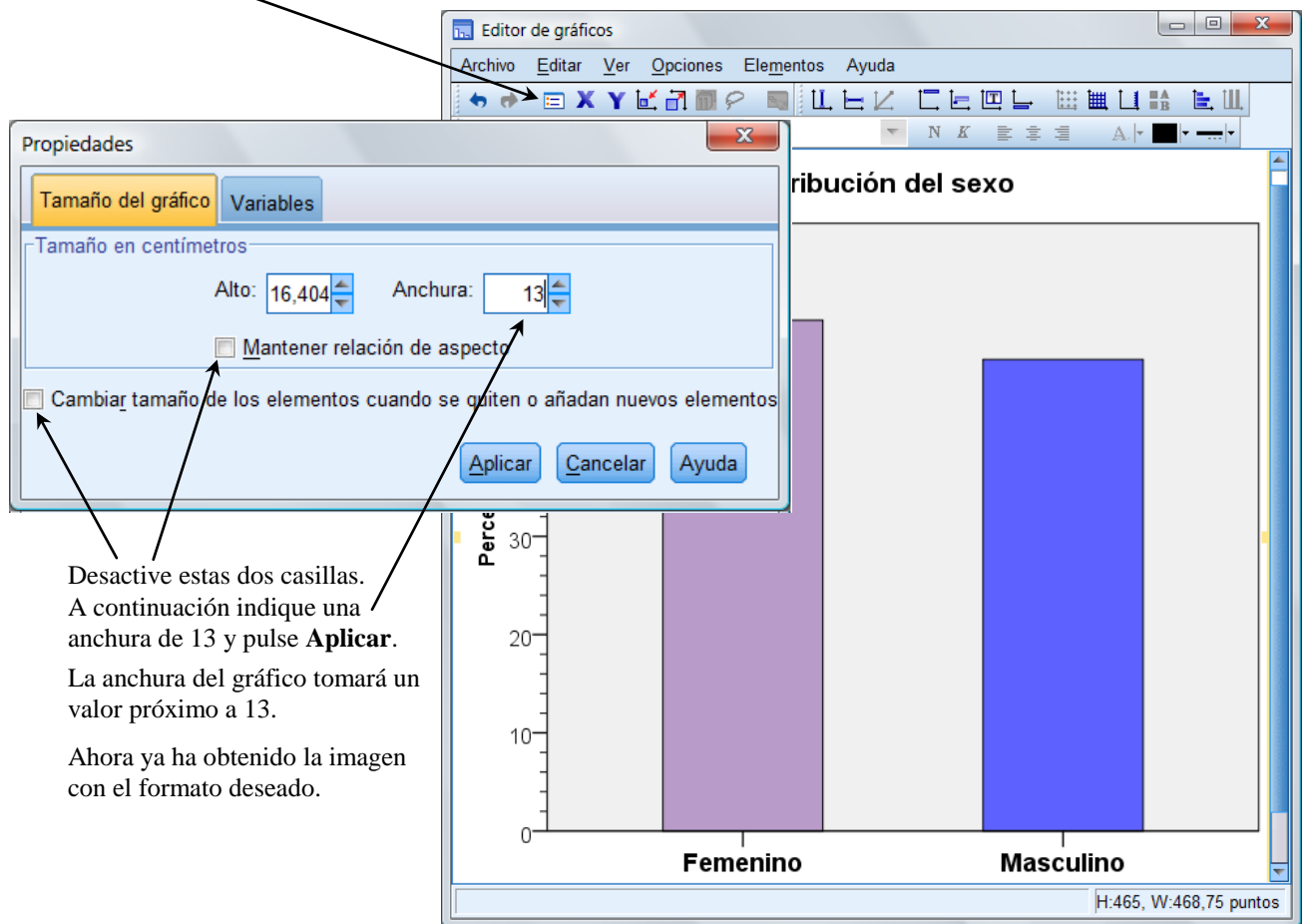
166. Ahora vamos a modificar el tamaño y el color de las barras (rosa a las mujeres y azul a los hombres).
 Seleccione la barra que representa a las mujeres: el primer clic las selecciona todas y un segundo clic sobre la barra deseada la selecciona. Seguidamente hacer doble clic para abrir el cuadro de Propiedades.

Para **cambiar el color** hacer clic sobre la casilla **Relleno** y luego elegir el color deseado. Para confirmar pulsar **Aplicar** y el color de la barra cambiará a rosa.

La ficha **Opciones de las barras** permite cambiar su anchura, Deslice el cursor hasta el 50%. Para confirmar pulsar **Aplicar** y las barras se estrecharán.

Cierre la ventana de propiedades y seleccione la barra **Masculino**. Doble clic sobre ella y se abrirá el cuadro de Propiedades. Asigne color azul (93, 97, 255), pulse **Aplicar** y cierre el cuadro.

167. Finalmente vamos a modificar el tamaño del gráfico para hacerlo más estrecho sin modificar su altura. Pulse el icono  de la barra de herramientas para abrir el cuadro de Propiedades.



Desactive estas dos casillas.
A continuación indique una anchura de 13 y pulse **Aplicar**.
La anchura del gráfico tomará un valor próximo a 13.
Ahora ya ha obtenido la imagen con el formato deseado.

168. Vamos a **copiar este gráfico en un documento de Word** con el mismo procedimiento de las tablas.

- Hacer clic con el botón derecho del ratón sobre el gráfico y elegir la opción *Copiar* (o Edición | Copiar) para pasarla al portapapeles de Windows.
- Abrir el documento Word y pegarla. Si se trabaja con Word 2003 o anterior, escoger la opción: Edición | Pegado especial... | Imagen (metarchivo mejorado).
Con Word 2007, la secuencia es: Inicio | Pegar | Pegado especial... | Imagen (metarchivo mejorado).
Compruebe que se ha pegado la imagen y que se puede variar su tamaño. Este objeto también se puede copiar al PowerPoint.

169. Desde la ventana de sintaxis se escoge la opción de menú Archivo | Guardar como..., se selecciona la carpeta **Tutorial SPSS** y se guarda con el nombre **SaludE** pulsando el botón **Guardar**. Seguidamente se accede a la ventana de Resultados, se escoge Archivo | Guardar como..., se selecciona la carpeta **Tutorial SPSS** y se guarda con el nombre **SaludE** pulsando el botón **Guardar**. Finalmente se sale de *SPSS Statistics* con el menú Archivo | Salir.

Fin del Tutorial