

Práctica 2.

Instalación del programa PSPP y obtención de una distribución de frecuencias.

Con esta práctica instalaremos el programa PSPP. El programa es un software específico para el análisis estadístico de datos que emplearemos para tratar encuestas sociológicas y de opinión. Una vez instalado, comenzaremos por grabar los datos de edad de un grupo de 80 personas y finalizaremos obteniendo la distribución de frecuencias de esta variable de intervalo. La realización de la práctica permitirá poner en marcha las herramientas básicas para el análisis estadístico.

Existen distintos paquetes comerciales de análisis estadístico. Dentro del campo específico del análisis de encuestas destacan por su mayor uso el ya clásico SPSS -Statistical Package for Social Sciences, y el STATA.

Desde la filosofía del Software libre (GNU) se ha desarrollado el proyecto PSPP. Obsérvese que son las siglas del SPSS en forma invertida. El software libre se caracteriza por la distribución gratuita y por la elaboración colaborativa entre distintos programadores. Dado que el funcionamiento es equivalente al SPSS y contiene las principales utilidades analíticas, en este curso se utilizará PSPP.

Para obtener el software acceda a la web: <https://www.gnu.org/software/pspp/get.html>

Busque su Sistema operativo y seleccione la última versión disponible, teniendo en cuenta si su ordenador es de 32 o 64 bits. El de 32 bits funciona en cualquier sistema Windows, el de 64 sólo con los sistemas operativos basados en este formato.

Si su sistema operativo es diferente (Mac, Linux...) o tiene conocimientos avanzados de informática. Diríjase a la página del proyecto de PSPP y busque la versión para OS X
Descarga desde la página del proyecto PSPP desde GNU



El proyecto PSPP cuenta con una página propia:

<http://www.gnu.org/software/pspp/>

Si tiene dificultad para bajar el software puede obtenerlo aquí:

<http://www2.uned.es/socioestadistica/pspp-20160927-daily-32bits-setup.exe>

<http://www2.uned.es/socioestadistica/pspp-20160927-daily-64bits-setup.exe>

Primera parte: Instalación de PSPP

Instrucciones de instalación para sistemas operativos Windows

Desde la página del proyecto, accedemos a downloads, y en la nueva página que se abre seleccionamos el Sistema operativo. Seleccione, dentro de la opción del sistema operativo Windows: "Installers" y después escoja la versión de 32 o 64 bits.

Los diferentes sistemas operativos de Windows pueden presentar ventanas con distinto aspecto, pero básicamente serán equivalentes a las que aparecen en este texto. A partir de ahora comienza el proceso de instalación. Posiblemente se abra un pequeño cuadro de diálogo para que seleccionemos el idioma: probablemente vendrá por defecto "Spanish" y pinchamos en "OK", en otros casos el instalador asume el idioma de la variable de Sistema del ordenador.

Ejecutamos el instalador, marcamos "Acepto los términos de la licencia" y "Siguiendo".

En cuanto a la carpeta de destino, se puede dejar la que ofrece el instalador por defecto (dependerá de cada ordenador) y pinchamos en "siguiendo".

Volvemos a aceptar con "Siguiendo". Windows instalará los archivos y se dará por terminada la instalación. Normalmente, en el escritorio encontraremos un nuevo icono, con el cual podremos acceder al programa:



Parte segunda: Introducción a la grabación manual de datos en PSPP

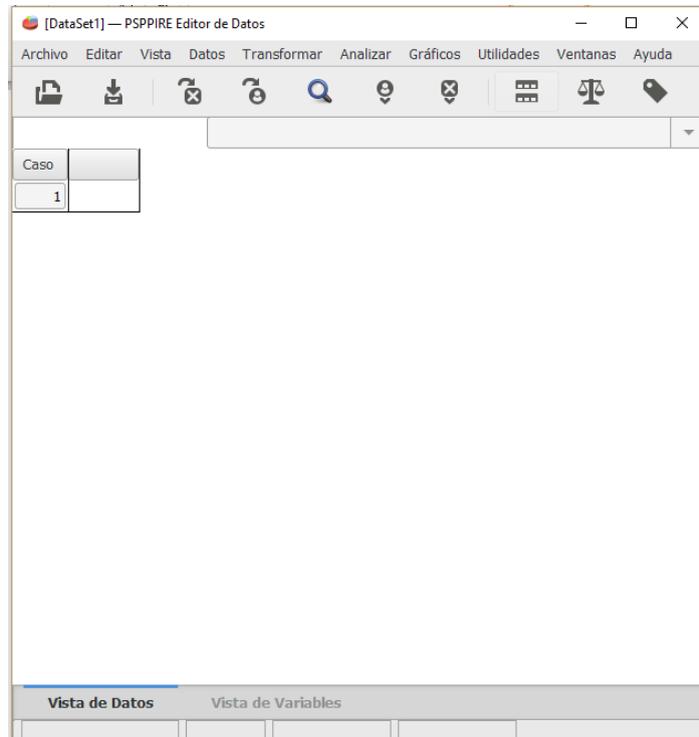
Antes de comenzar a utilizar el programa conviene que tenga en cuenta que PSPP utiliza tres ventanas simultáneamente (datos, resultados y sintaxis).

- Una, con aspecto de hoja de cálculo, para visualizar los datos. Son ficheros con extensión **.sav**

- Otra segunda ventana en la que se registran los resultados. Son ficheros con diferentes extensiones según exportemos el contenido como htm, pdf, txt, etc.

- Una tercera, denominada de sintaxis, en la que escribiremos comandos e instrucciones. Son ficheros con extensión **.sps**

Al ejecutar el programa PSPPP desde el icono del escritorio se abre la ventana de los datos apareciendo una matriz en blanco.



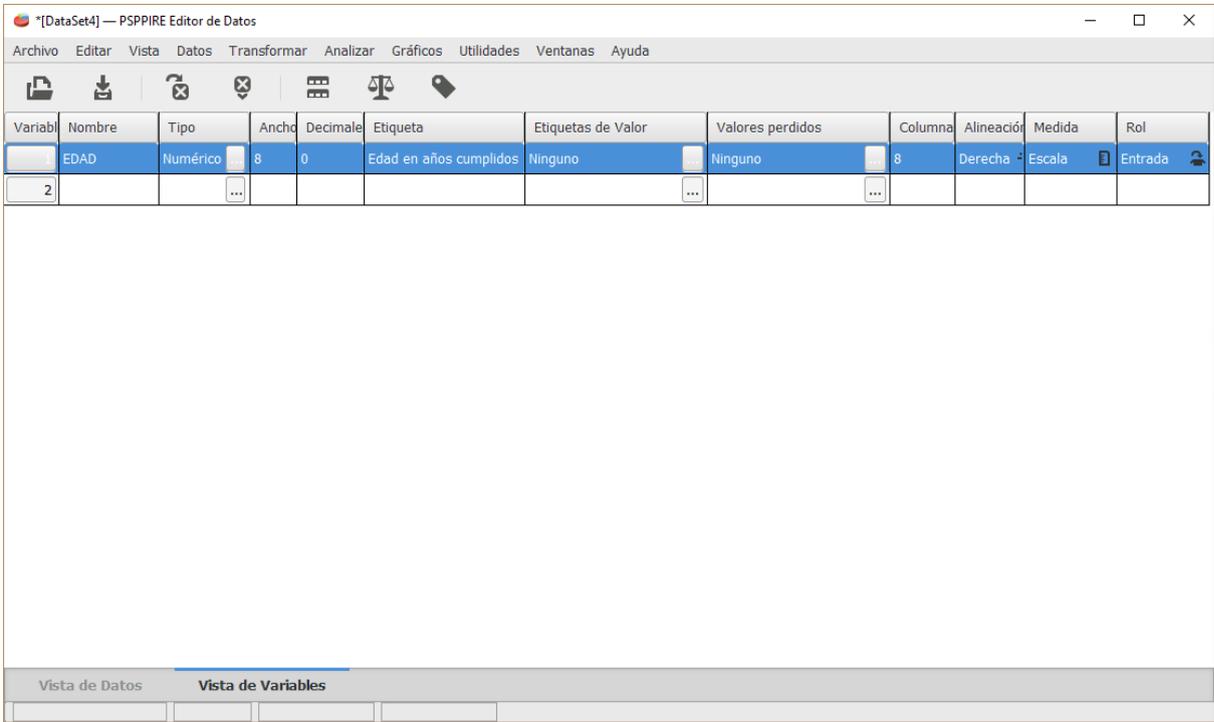
En las pestañas de la parte inferior izquierda de la ventana del programa encontramos activada por defecto la “**vista de datos**”, pero podemos cambiar a la “**vista de variables**”, pinchando en la pestaña correspondiente.

```
56 66 21 63 60 60 60 42 38 35 45 44 57 51 57
45 38 49 50 49 47 35 30 53 32 53 42 47 41 28
47 61 35 33 48 40 38 54 63 21 20 27 53 55 41
54 28 39 45 34 39 21 42 39 27 48 48 53 35 33
27 27 54 52 38 29 30 53 53 58 23 41 63 28 43
56 59 60 40 24
```

Para comenzar a conocer la estructura y los menús del programa, comenzaremos por introducir manualmente los datos que aparecen arriba. El recuadro contiene las edades de un grupo de 80 personas que queremos representar mediante una tabla de frecuencias.

Cambiamos a la Vista de variables y escribiremos en la primera casilla un nombre para la variable que vamos a introducir.

Llamaremos a esta variable “Edad” y pulsamos intro. Por defecto el software asigna determinadas características a las variables que se crean, pero las podremos cambiar de forma sencilla. A la derecha del nombre aparece el tipo de variable, en este caso es numérica y lo dejamos como está por defecto, pero en la característica decimales, podemos quitar los dos decimales que no necesitamos: en la casilla decimales ponemos el valor 0.



Téngase en cuenta que el nombre de la variable debe contener como máximo ocho caracteres sin espacios ni signos. Podremos sin embargo escribir sin limitaciones la etiqueta de esa variable, por ejemplo “Edad en años cumplidos”.

A continuación regresamos a la pestaña de datos para introducir las edades en la primera columna de la ventana de datos:

56 66 21 63 60 60 60 42 38 35 45 44 57 51 57
 45 38 49 50 49 47 35 30 53 32 53 42 47 41 28
 47 61 35 33 48 40 38 54 63 21 20 27 53 55 41
 54 28 39 45 34 39 21 42 39 27 48 48 53 35 33
 27 27 54 52 38 29 30 53 53 58 23 41 63 28 43
 56 59 60 40 24

Introducimos cada dato y seleccionamos la casilla siguiente con los cursores....

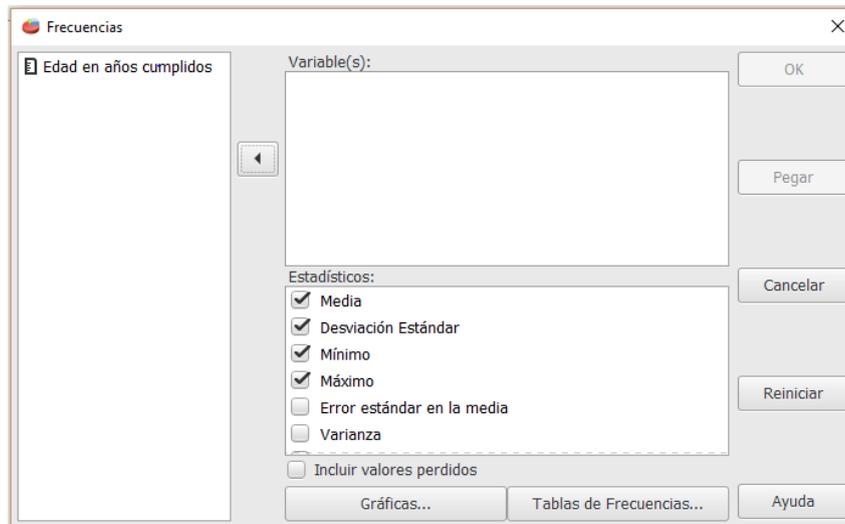
Caso	EDAD
1	56
2	66
3	21
4	63
5	60
6	60
7	42
8	38
9	35
10	45
11	

Una vez grabados los 80 casos es imprescindible grabar el fichero. Deberemos ponerle un nombre, -por ejemplo **PRACTICA-**, nos pedirá confirmación de la carpeta en la que queremos guardar el fichero y nos dará la opción para que el indiquemos, por defecto si el fichero es de sistema¹. Le indicamos que el fichero es de sistema y nos lo grabará con el nombre PRACTICA.SAV

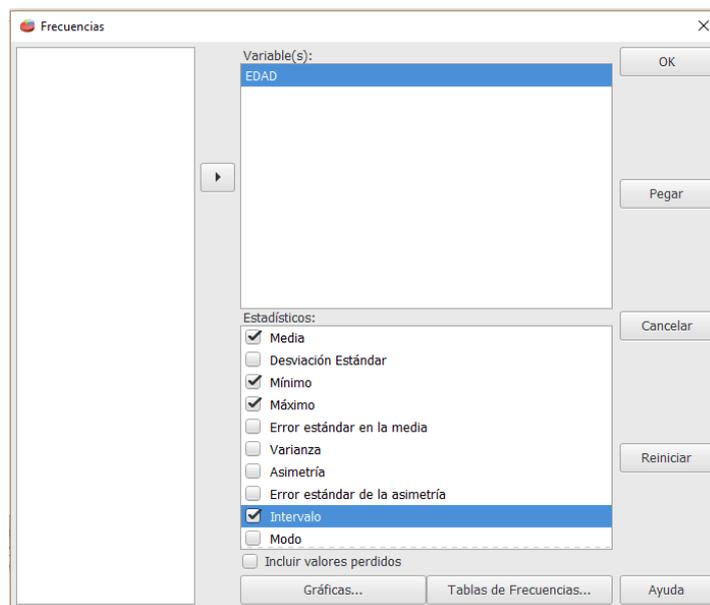
Finalizada la grabación vamos a comenzar el estudio de los datos. En el menu superior, seleccionamos “Analizar”, y desplegaremos, “estadística descriptiva”, y dentro de ésta “frecuencias”.

Caso	EDAD
1	56
2	66
3	21
4	63
5	60
6	60
7	42
8	38
9	35
10	45
11	44
12	57
13	51
14	57
15	45
16	38

Se abrirá una nueva ventana:



Una vez en **frecuencias**, seleccionamos la variable Edad en la caja de la izquierda y la enviamos a la caja de la derecha, pulsando en el icono correspondiente :



Seleccionamos los estadísticos de la media, el mínimo, el máximo y el intervalo. Aceptamos con OK y obtenemos la siguiente tabla en la ventana de resultados:

FRECUENCIAS
 /VARIABLES= EDAD
 /FORMAT=AVALUE TABLE
 /STATISTICS=MEAN MINIMUM MAXIMUM RANGE.

Edad en años cumplidos

Etiqueta de Valor	Valor	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
	20	1	1.25	1.25	1.25
	21	3	3.75	3.75	5.00
	23	1	1.25	1.25	6.25
	24	1	1.25	1.25	7.50
	27	4	5.00	5.00	12.50
	28	3	3.75	3.75	16.25
	29	1	1.25	1.25	17.50
	30	2	2.50	2.50	20.00
	32	1	1.25	1.25	21.25
	33	2	2.50	2.50	23.75
	34	1	1.25	1.25	25.00
	35	4	5.00	5.00	30.00
	38	4	5.00	5.00	35.00
	39	3	3.75	3.75	38.75
	40	2	2.50	2.50	41.25
	41	3	3.75	3.75	45.00
	42	3	3.75	3.75	48.75
	43	1	1.25	1.25	50.00
	44	1	1.25	1.25	51.25
	45	3	3.75	3.75	55.00
	47	3	3.75	3.75	58.75
	48	3	3.75	3.75	62.50
	49	2	2.50	2.50	65.00
	50	1	1.25	1.25	66.25
	51	1	1.25	1.25	67.50
	52	1	1.25	1.25	68.75
	53	6	7.50	7.50	76.25
	54	3	3.75	3.75	80.00
	55	1	1.25	1.25	81.25
	56	2	2.50	2.50	83.75
	57	2	2.50	2.50	86.25
	58	1	1.25	1.25	87.50
	59	1	1.25	1.25	88.75
	60	4	5.00	5.00	93.75
	61	1	1.25	1.25	95.00
	63	3	3.75	3.75	98.75
	66	1	1.25	1.25	100.00
	<i>Total</i>	80	100.0	100.0	

Edad en años cumplidos

<i>N</i>	Válido	80
	Perdidos	0
<i>Media</i>		43.41
<i>Intervalo</i>		46.00
<i>Mínimo</i>		20.00
<i>Máximo</i>		66.00

En primer lugar tenemos la tabla de frecuencias y en segundo lugar los estadísticos solicitados. Son 80 casos, no hay casos sin valores o perdidos. El valor mínimo son 20 años, el máximo 66 y por tanto el recorrido –intervalo- 46. También hemos obtenido que la edad media es de 43,41 años.

Parte Tercera: Elaboración de una tabla de frecuencias con datos agrupados

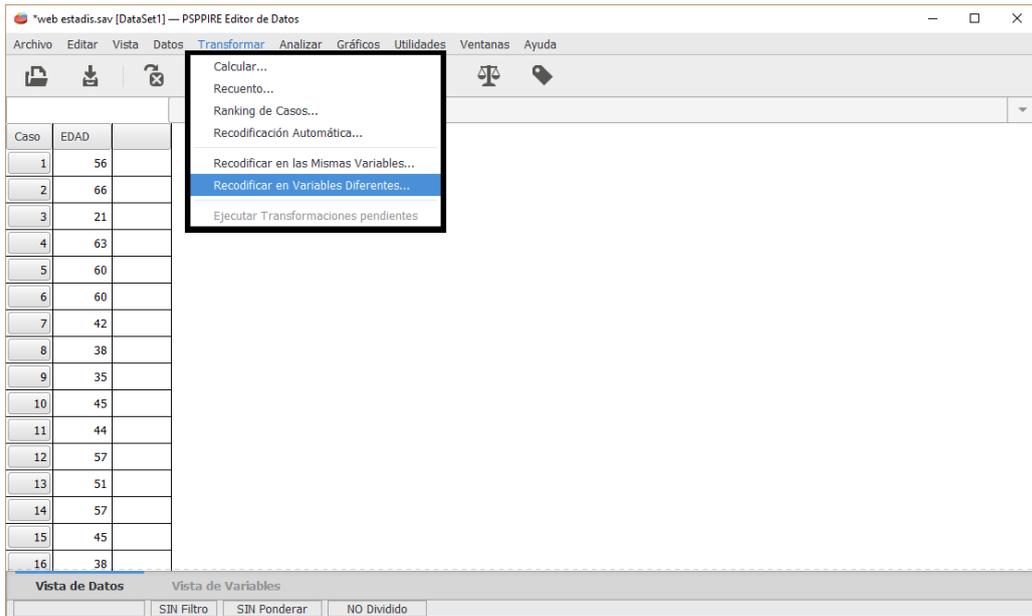
La tabla de frecuencias anterior al no estar agrupados los datos no resulta de mucha utilidad analítica. Vamos a construir una nueva tabla con las edades agrupadas en intervalos decenales. En concreto en 5 categorías:

- 20 a 29 años
- 30 a 39 años
- 40 a 49 años
- 50 a 59 años
- 60 a 66 años

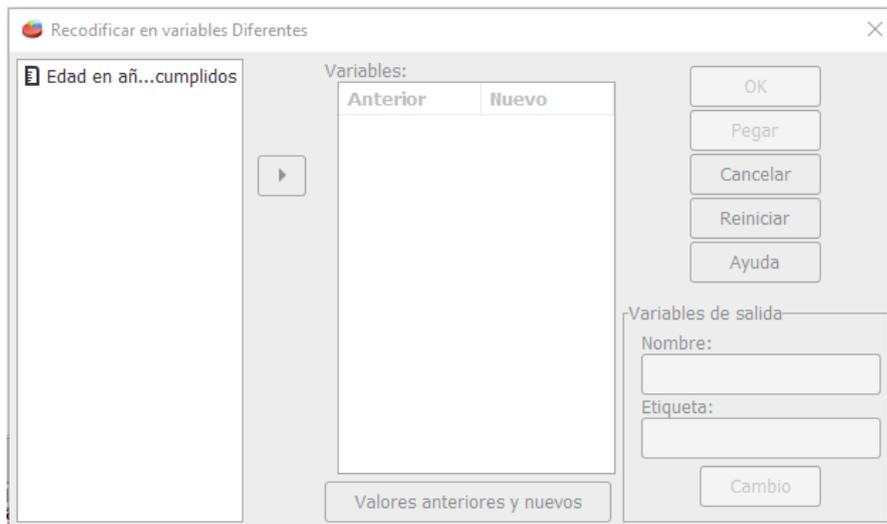
Para agrupar los datos vamos a recodificar, por ejemplo quienes tienen de 20 a 29 años, les daremos el valor 25 –que es su marca de clase-, quienes tienen de 30 a 39 años les daremos el valor 35, y así sucesivamente.

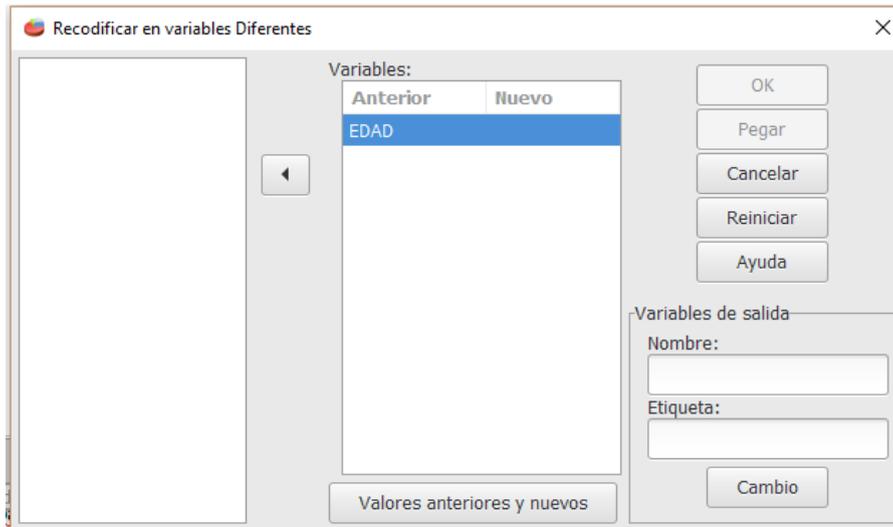
Esta la operación la haremos en una nueva variable –que denominaremos Edad2- y conservaremos la variable original.

En el menú principal, seleccionaremos “Transformar” y en el desplegable que se abre seleccionamos “Recodificar en variables diferentes”.

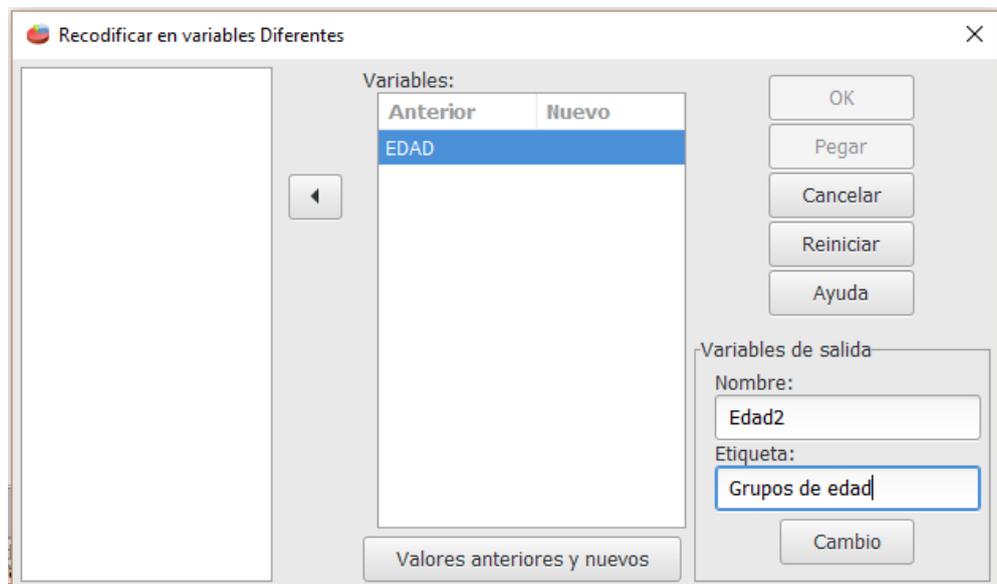


Se abrirá una ventana en la que seleccionaremos la variable EDAD y pinchando en la flecha  la pasamos a la parte derecha de la ventana

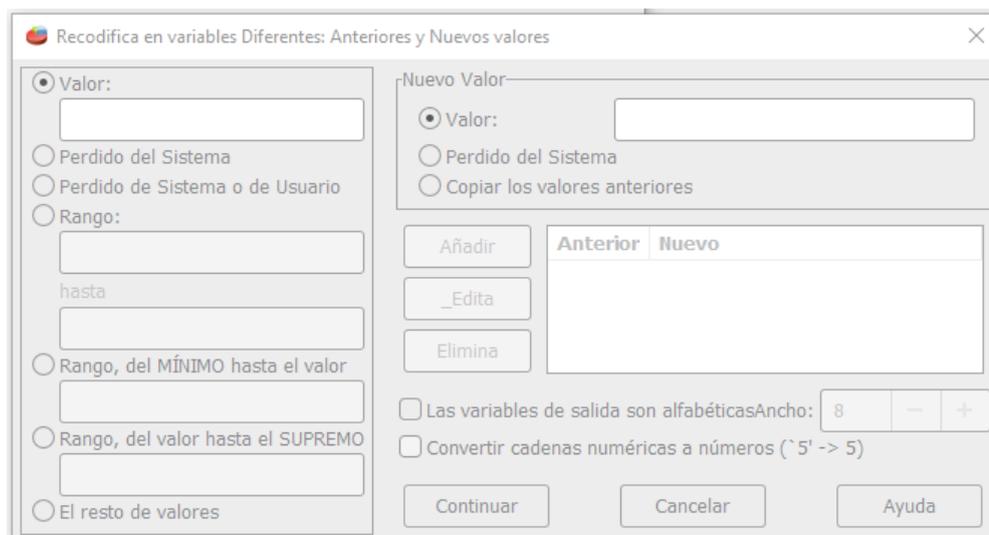




Pondremos nombre y etiqueta a la nueva variable “Variables de salida”. Usamos un nombre breve sin espacios y signos especiales, por ejemplo “Edad2”, para la variable, y el nombre que queramos, por ejemplo “Grupos de edad”, para la etiqueta; en este caso si se permiten los espacios en blanco. Seleccionamos la variable (se pondrá azul) y escribimos los nombres. Y pulsamos en el botón “**cambio**”.

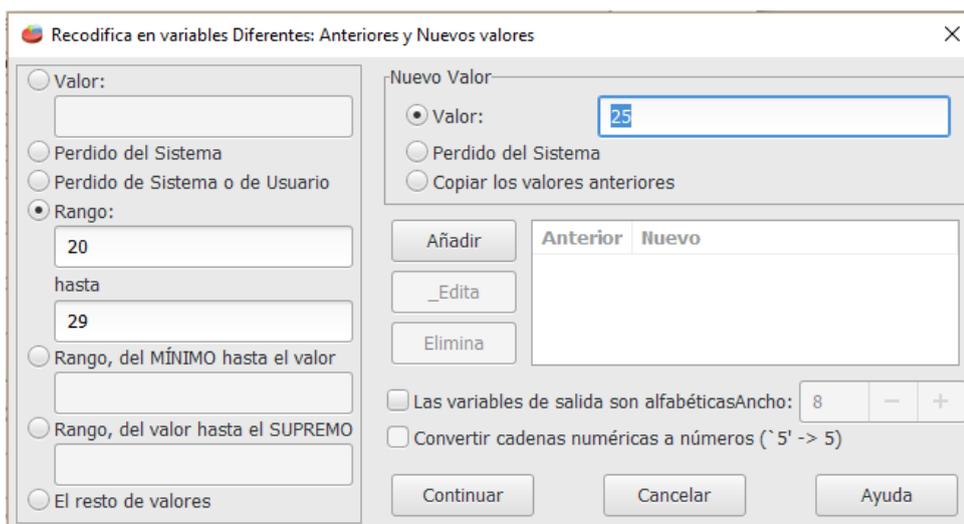


A continuación en la misma pestaña pinchamos en “**Valores anteriores y nuevos**”; se abrirá una nueva pestaña para introducir los valores concretos que tomará la nueva variable. Como sabemos los valores que tomará cada intervalo, seleccionamos intervalo:



Sabemos que el valor menor es 20 y el máximo 66. Haremos 4 intervalos iguales de 10 años y un último de 6.

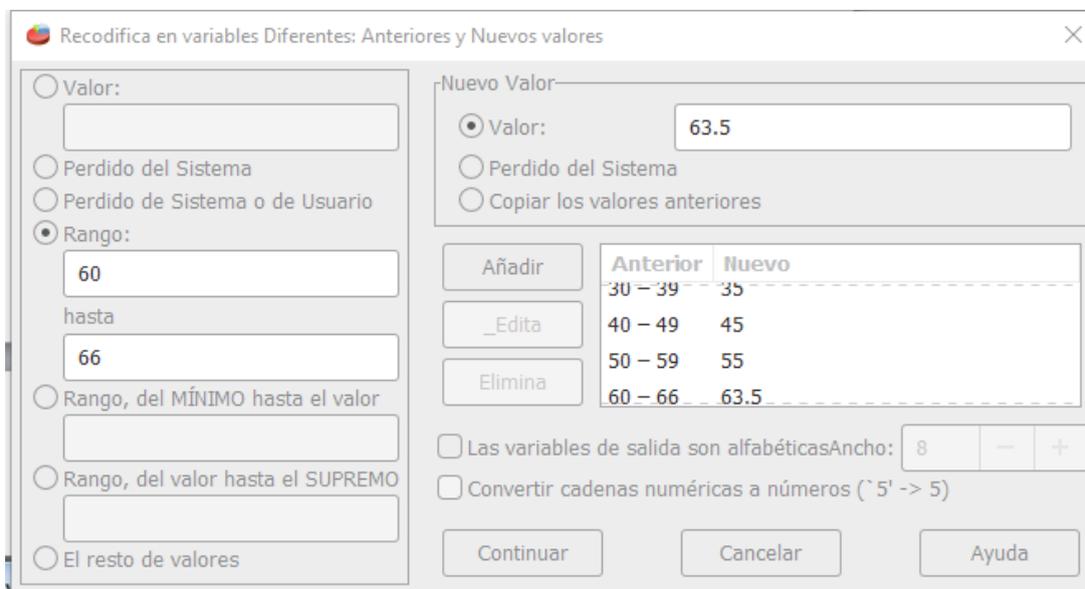
En la imagen de abajo se muestra cómo se introducen los límites inferior y superior del primer intervalo como “**rango**”, y como “**valor**” se asigna, en este caso, la **marca de clase** o punto central. El primer intervalo incluye las edades entre 20 y 30 años, pero como se trata de años cumplidos, escribiremos como límite inferior 20 y superior 29, y como marca de clase el punto medio: 25. Pulsamos en “Anadir”, para grabar la categoría. Sin salir de la ventana introducimos los demás rangos de edad con su marca de clase como valor.



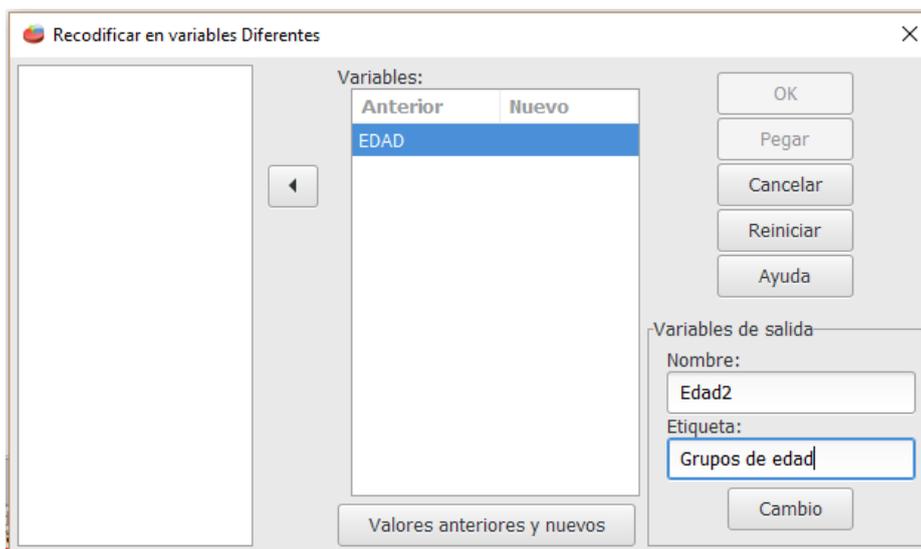
En el visor de la derecha se mostrarán los límites del intervalo y la marca de clase. El programa asignará el valor 25 a todos los casos que presenten de 20 a 29 años cumplidos.

Sin cerrar la ventana repetiremos la operación para los otros intervalos, y una vez introducidos todos los intervalos con sus marcas de clase como valores, pincharemos en “Continuar”.

Tras terminar con la operación observe que la marca de clase del último intervalo es 63,5 –el valor se introduce con punto en vez de con coma-. Como sabemos el intervalo va desde 60 años hasta 66, como es edad cumplida, el límite superior llega hasta 67.



Tras pulsar “Continuar” aparecerá de nuevo la ventana anterior.



Y ahora simplemente pinchamos en OK para que se lleve a cabo la recodificación de la variable EDAD en otra nueva variable que hemos llamado Edad2, con etiqueta Grupos de edad, en 5 intervalos de edad.

(No olvide que en cualquier momento puede guardar el archivo que está modificando para poder abrirlo más tarde y continuar donde lo dejó)

En la ventana principal de los datos, nos aparece una nueva variable, Edad2, con 80 casos, pero que sólo toma 5 valores diferentes, los de las marcas de clase de cada categoría. Debe tenerse

en cuenta que esta variable creada tiene menos variabilidad que la original y aunque es útil para representar la tabla y resumir los datos, no será muy útil para calcular estadísticos como la media o la varianza, pues en lugar de los 80 datos individuales, la variable sólo toma los 5 valores de las marcas de clase.

The screenshot shows the PSPPIRE Editor de Datos interface. The main window displays a data table with the following content:

Caso	EDAD	Edad2
1	56	55.00
2	66	63.50
3	21	25.00
4	63	63.50
5	60	63.50
6	60	63.50
7	42	45.00
8	38	35.00
9	35	35.00
10	45	45.00
11	44	45.00
12	57	55.00
13	51	55.00
14	57	55.00
15	45	45.00
16	38	35.00

At the bottom of the window, there are tabs for 'Vista de Datos' (selected) and 'Vista de Variables'. Below the tabs are buttons for 'SIN Filtro', 'SIN Ponderar', and 'NO Dividido'.

Si seleccionamos la pestaña de las variables, podemos cambiar los atributos de la nueva variable. En la cuarta columna se pueden establecer los decimales. Pinchemos en la casilla “Decimales” para la fila de “Edad2” y cambiemos el número de decimales a 1.

The screenshot shows the PSPPIRE Editor de Datos interface with the 'Vista de Variables' tab selected. The table below shows the attributes for the variables 'EDAD' and 'Edad2':

Variabl	Nombre	Tipo	Ancho	Decimale	Etiqueta	Etiquetas de Valor	Valores perdidos	Columna	Alineación	Medida	Rol
1	EDAD	Numérico	8	0	Edad en años cumplidos	Ninguno	Ninguno	8	Derecha	Escala	Entrada
2	Edad2	Numérico	8	0	Grupos de edad	Ninguno	Ninguno	8	Derecha	Escala	Entrada
3											

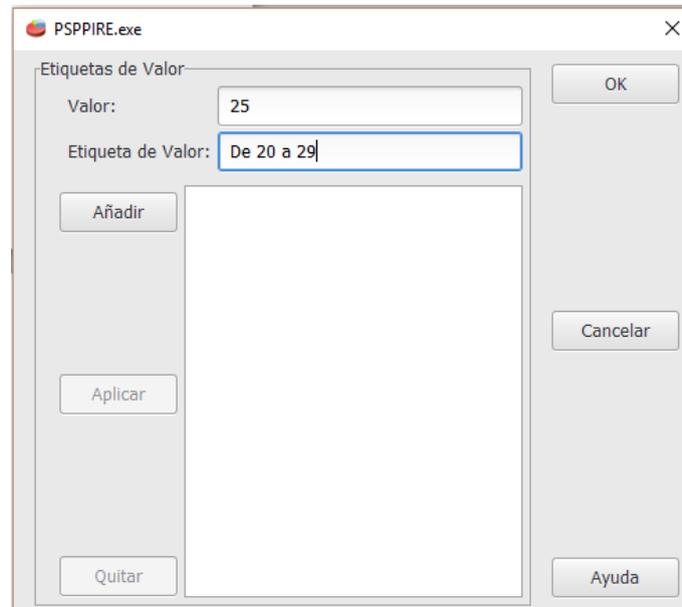
At the bottom of the window, there are tabs for 'Vista de Datos' and 'Vista de Variables' (selected). Below the tabs are buttons for 'SIN Filtro', 'SIN Ponderar', and 'NO Dividido'.

A continuación pondremos **etiquetas para los valores** de la variable, para que puedan

verse luego en la tabla que pretendemos obtener en la ventana de resultados.

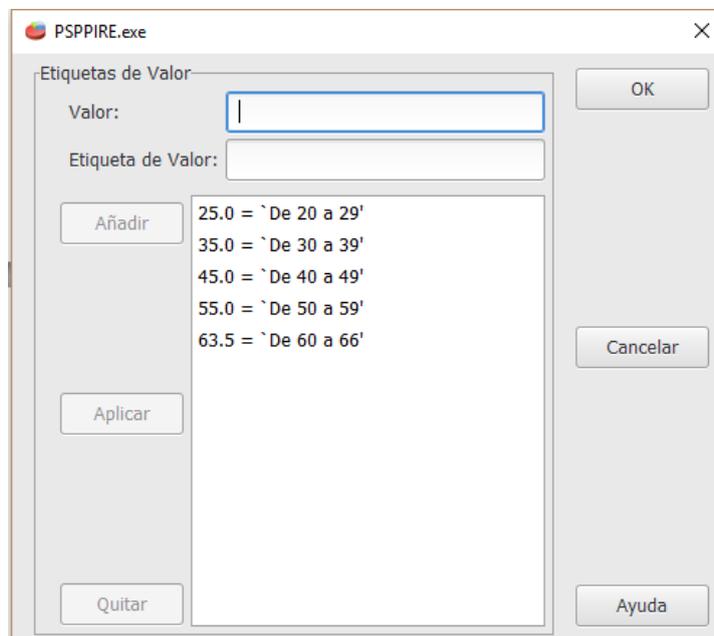
Pinchamos con el ratón la casilla correspondiente a la variable Edad2 y la columna de los valores (Etiquetas de Valor):

Dentro de la casilla aparece un pequeño botón gris cuadrado, pinchamos en él y se abrirá una nueva ventana:



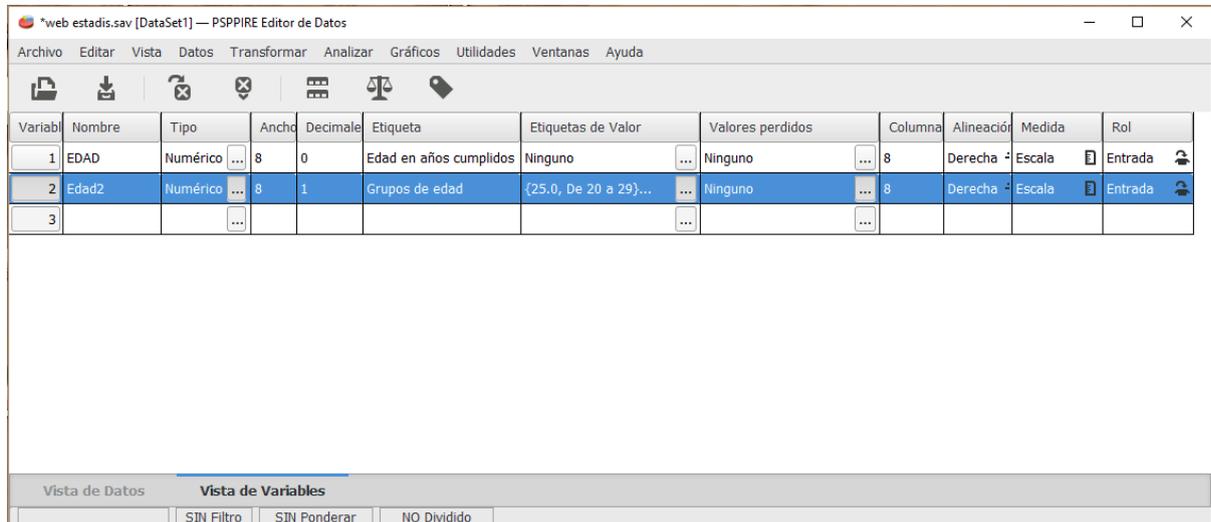
En la casilla Valor introduciremos los valores de cada **marca de clase** o punto medio del intervalo y también la etiqueta del valor en la casilla correspondiente, y las añadimos una a una pinchando en **“Añadir”** (para poner el valor de la última categoría, que es 63.5 utilice el signo del punto en lugar de la coma) .

Para la primera categoría, el valor es su marca de clase, 25 y su etiqueta “De 20 a 29”, y aceptamos con Añadir. Repetimos la operación sin cerrar la ventana para cada intervalo (**para poner el valor de la última categoría que es 63.5 utilice el signo del punto en lugar de la coma**)

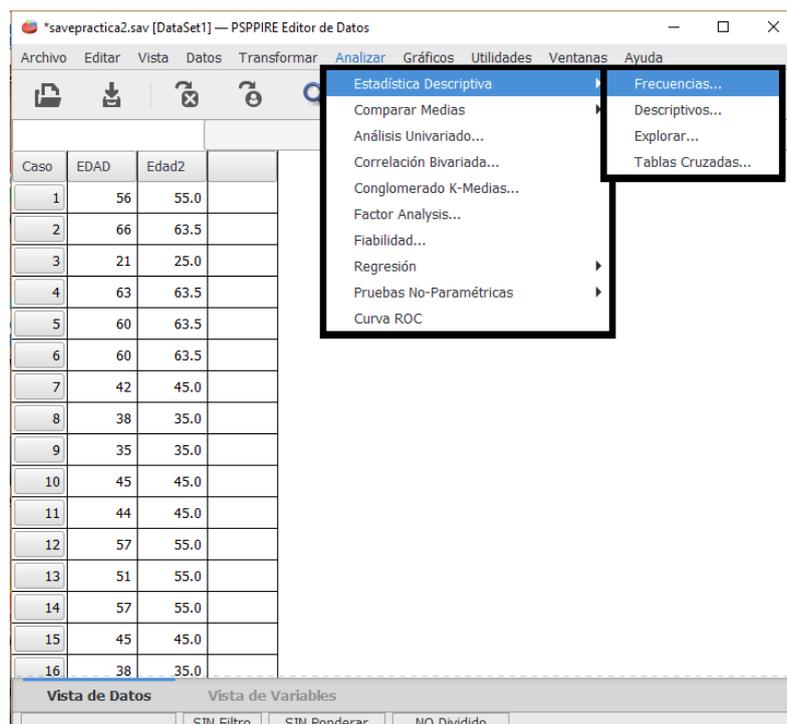


Cuando hemos introducido los datos de todos los intervalos aceptamos con OK.

En la vista de variables podemos comprobar que las características de la variable son correctas:

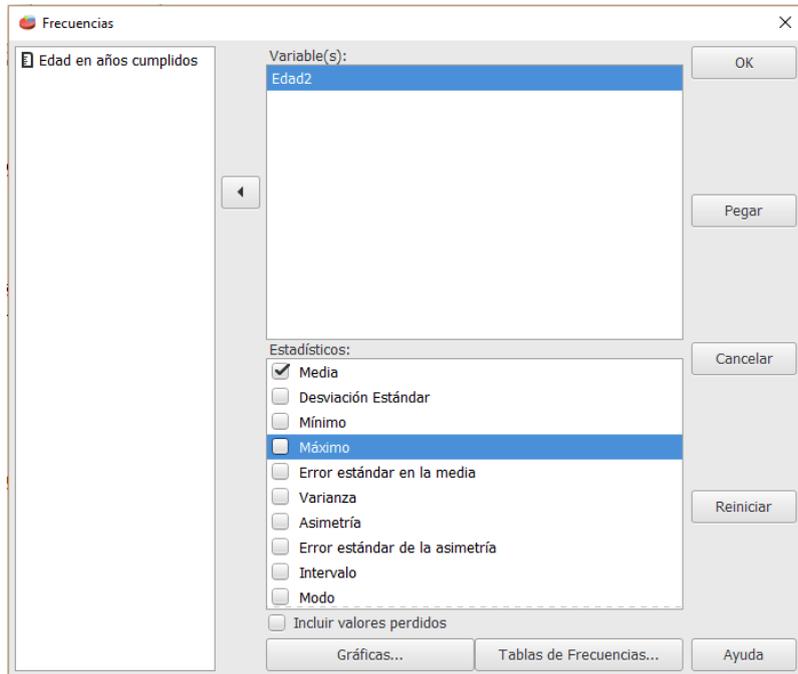


Volvemos a la ventana principal de Vista de datos y puede entonces solicitar la confección de una tabla, y nuevos análisis para la variable creada: Solicitamos el cálculo de la tabla de frecuencias, desde el menú principal, pinchando en Analizar-> Descriptivas-> y seleccionamos el comando "Frecuencias".



Se abrirá la ventana del comando de Frecuencias y seleccionaremos la variable Edad2

pinchando en  y la variable pasa a la lista de variables de la derecha. Abajo marcamos únicamente el estadístico Media. A continuación ejecutamos con OK.



En la ventana de resultados tendremos la siguiente tabla:

```
FREQUENCIES
/VARIABLES= Edad2
/FORMAT=AVALUE TABLE
/STATISTICS=MEAN.
```

Grupos de edad

<i>Etiqueta de Valor</i>	<i>Valor</i>	<i>Frecuencia</i>	<i>Porcentaje</i>	<i>Porcentaje Válido</i>	<i>Porcentaje Acumulado</i>
De 20 a 29	25.0	14	17.50	17.50	17.50
De 30 a 39	35.0	17	21.25	21.25	38.75
De 40 a 49	45.0	21	26.25	26.25	65.00
De 50 a 59	55.0	19	23.75	23.75	88.75
De 60 a 66	63.5	9	11.25	11.25	100.00
<i>Total</i>		80	100.0	100.0	

Grupos de edad

<i>N</i>	<i>Válido</i>	80
	<i>Perdidos</i>	0
<i>Media</i>		43.83

En la tabla podemos ver los intervalos de edad, las marcas de clase, las frecuencias absolutas, las frecuencias relativas y las frecuencias relativas acumuladas. Otra tabla nos presenta información sobre el número de casos y la media solicitada.

Observamos que hay ligeras diferencias en el cálculo de la media con datos agrupados y sin agrupar. Cuando se agrupan los datos se pierde precisión, debido a que se toman los mismos

valores para cada individuo perteneciente al mismo intervalo, eliminando la variabilidad. Por este motivo, para conocer estos estadísticos será siempre preferible utilizar la variable original, y utilizaremos esta recodificación sólo en los casos en que nos interese representar la información en una tabla.