Ejercicio práctico:

Regresión logística binaria con PSPP

Texto elaborado por el Equipo Docente

para la asignatura

Análisis Multivariante

Luis Camarero Rioja

Alejandro Almazán

Beatriz Mañas Ramírez

Ejercicio práctico de Regresión logística binaria con PSPP por Luis Camarero Rioja, Alejandro Almazán Llorente, Beatriz Mañas Ramírez se distribuye bajo una Licencia Creative Commons Atribución-CompartirIgual 4.0 Internacional.
Basada en una obra en http://www.uned.es/socioestadistica/Multivariante/Plan_trabajo.htm.

Ejercicio práctico de Regresión logística binaria con PSPP

Se tratará aquí el procedimiento práctico de análisis de regresión logística binaria con el paquete estadístico PSPP[1]. Para el ejercicio utilizaremos los datos del barómetro del CIS nº 2941, que contiene diversas preguntas relacionadas con la participación en acciones sociales y políticas. En concreto nos centraremos en la pregunta del cuestionario, que trata sobre si el entrevistado asiste a las manifestaciones. Tomaremos esa variable como dependiente y trataremos de observar la probabilidad de encontrar un individuo que asista a una manifestación, dependiendo de los valores observados en otras variables del cuestionario. Con el programa estadístico construiremos un modelo que relacione la variable dependiente con otras variables del cuestionario: Sexo, Edad, Estudios, Situación laboral, escala de Ideología, etc.

Obtención del software y los datos, y lectura del barómetro

Enumeraremos a continuación los primeros pasos a seguir para obtener los datos, el software PSPP y leer los datos en nuestro ordenador.

El **primer paso** consiste en obtener los microdatos del estudio 2941 en la web del Centro de Investigaciones Sociológicas (CIS).
El **segundo paso** será abrir el archivo comprimido del barómetro: tomamos el archivo ZIP llamado "*fichero_datos*", lo descomprimimos y obtenemos los 7 archivos que contiene:

codigo2941.pdf (Contiene los códigos de cada categoría de la variable y la columna que ocupa la variable en la matriz de datos).
cues2941.pdf (Es una copia del cuestionario que se utilizó para hacer las entrevistas de la encuesta).
DA2941 (es un archivo de texto que contiene la matriz de datos con las respuestas de los entrevistados).
ES2941 (Es un archivo de texto que contiene la sintaxis para poder leer la matriz de datos con el software PSPP).
FT2941.pdf (Es un archivo que contiene los datos de la ficha técnica de la investigación, con información sobre la muestra, el error muestral, etc.).
Sas2941 (Es un archivo que contiene la matriz de datos para ser utilizada con el paquete estadístico SAS)
tarjetas2941.pdf (Contiene las tarjetas que se utilizan durante la administración del cuestionario a los entrevistados, cuando se hace necesario mostrar gráficamente las opciones de las respuestas).

El **tercer paso** es renombrar el archivo DA2941 poniendo su extensión de archivo de texto, de forma que se llame DA2941.txt

Antes de continuar recomendamos colocar dicho archivo en una memoria USB, directamente en la raíz, sin utilizar carpetas. (Este procedimiento se trató en la asignatura de primer curso y formó parte de la prueba de evaluación continua). Para repasar el procedimiento de **obtención**

[1] En este ejercicio se ha utilizado la versión de PSPP "psppire.exe 0.8.1-g876da2" , disponible en: http://pspp.awardspace.com/

del software y la apertura de un archivo de encuesta del CIS puede acceder a las instrucciones detalladas que se encuentran en las siguientes direcciones:

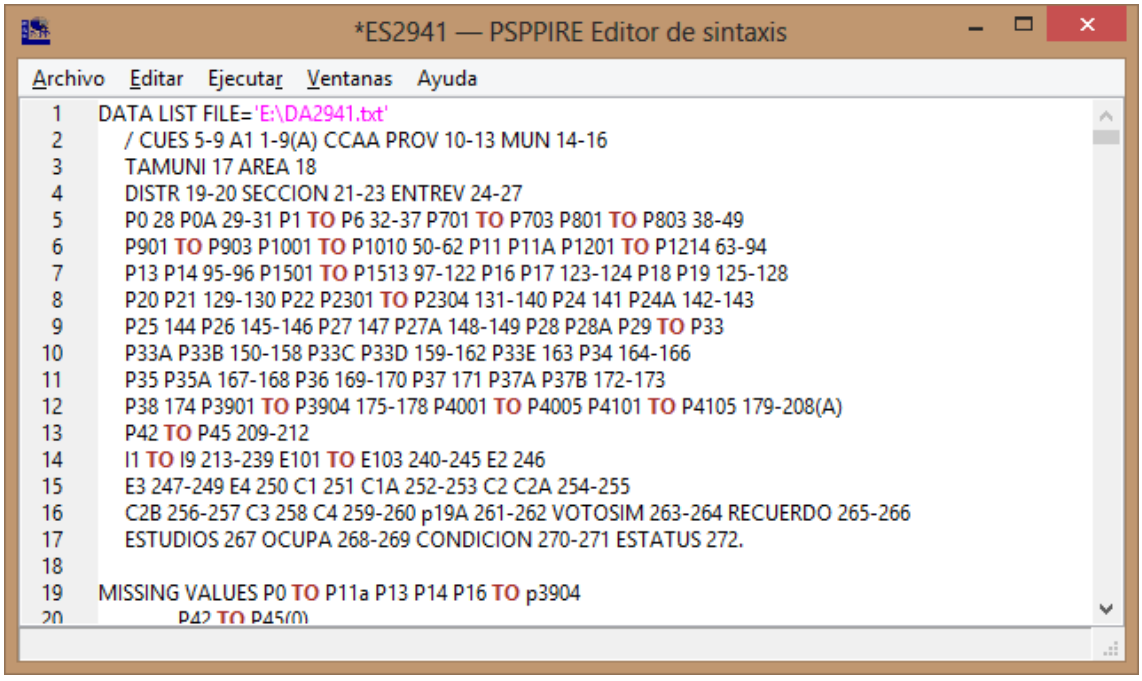
http://www.uned.es/socioestadistica/Practicas_%20ejercicios_guia/Practica_2(2013).pdf

http://www.uned.es/socioestadistica/Practicas_%20ejercicios_guia/Practica3.pdf

El **cuarto paso** es abrir la encuesta con el paquete estadístico PSPP, para lo cual comenzamos por ejecutar el programa PSPP.

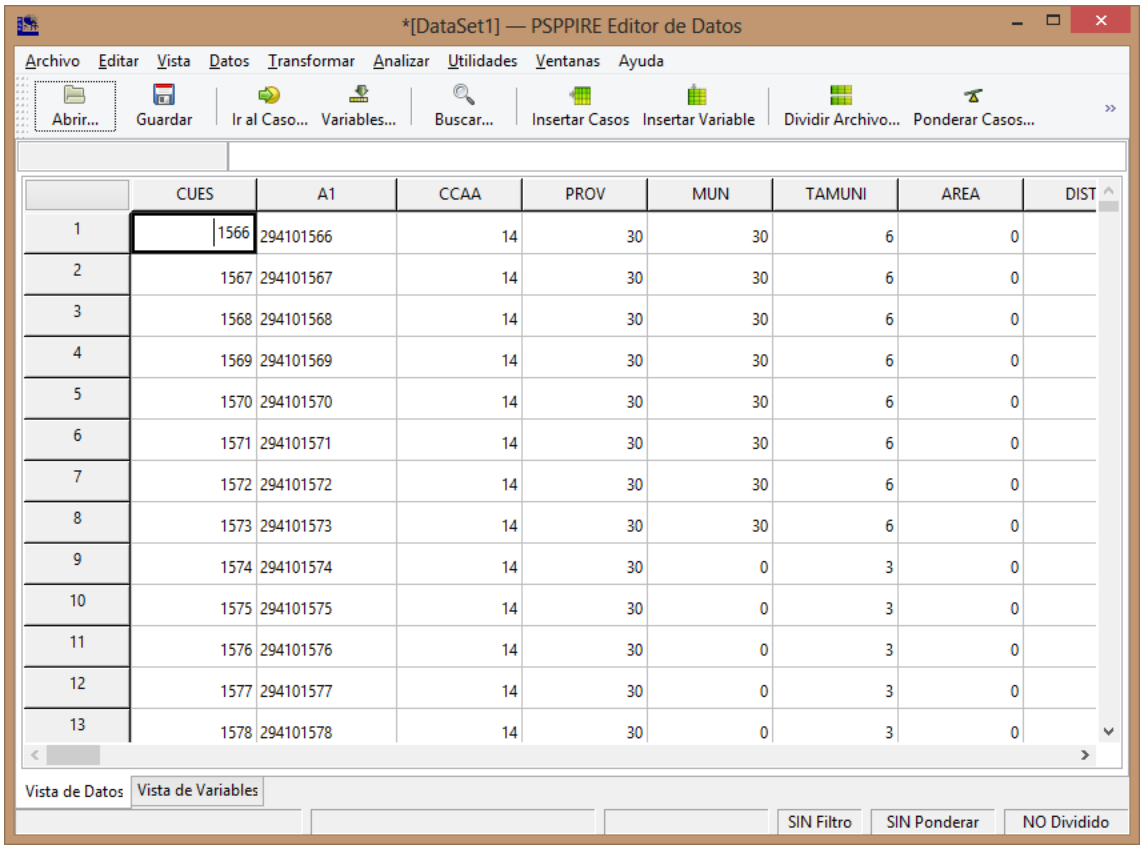
En el menú superior pinchamos en la opción *Archivo-> Abrir* y se abrirá una ventana para localizar el archivo de la sintaxis del barómetro. Buscamos el archivo ES2941 (marcando la opción "todos los archivos" en la pestaña tipo de archivo) y seleccionamos Abrir. En la ventana de sintaxis nos aparece el texto que contiene el archivo, que como hemos dicho se refiere a la clave para leer los datos, especificando en qué fila y en que columna de la matriz de datos se encuentran las variables y las respuestas de los entrevistados, así como las etiquetas y los códigos de las variables.

A continuación rectificamos la primera línea de la sintaxis que hemos abierto con el PSPP para leer el archivo con los datos e informando de la unidad que nuestro sistema operativo ha asignado al USB al enchufarlo al ordenador. Suele ser una letra, como D, E, F, G, etc. En el ordenador que he utilizado para hacer este ejercicio, al pinchar la memoria USB se le ha asignado la letra E. por tanto hay que rectificar la primera línea de la sintaxis para que aparezca de la siguiente manera:



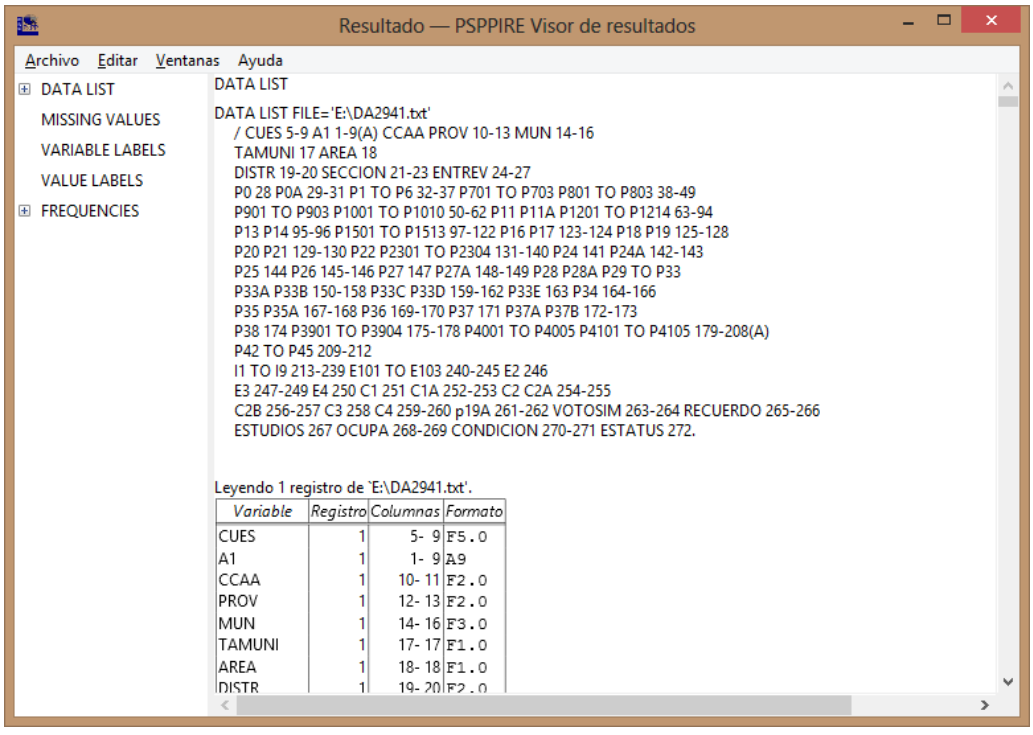
Recuerde que no hay que añadir espacios en blanco, sino simplemente insertar dentro de las comillas y antes del nombre del archivo E:\

Podemos entonces ejecutar la sintaxis para que se abra la matriz que contiene los datos de las respuestas de los entrevistados a cada pregunta del cuestionario, seleccionando del menú superior de la ventana del PSPP el comando RUN siguiendo la secuencia *Ejecutar->Todos*.
Se nos abrirá una ventana con la matriz de datos.



	CUES	A1	CCAA	PROV	MUN	TAMUNI	AREA	DIST
1	1566	294101566	14	30	30	6	0	
2	1567	294101567	14	30	30	6	0	
3	1568	294101568	14	30	30	6	0	
4	1569	294101569	14	30	30	6	0	
5	1570	294101570	14	30	30	6	0	
6	1571	294101571	14	30	30	6	0	
7	1572	294101572	14	30	30	6	0	
8	1573	294101573	14	30	30	6	0	
9	1574	294101574	14	30	0	3	0	
10	1575	294101575	14	30	0	3	0	
11	1576	294101576	14	30	0	3	0	
12	1577	294101577	14	30	0	3	0	
13	1578	294101578	14	30	0	3	0	

Y otra ventana con los resultados:




```
DATA LIST

DATA LIST FILE='E:\DA2941.txt'
     / CUES 5-9 A1 1-9(A) CCAA PROV 10-13 MUN 14-16
     TAMUNI 17 AREA 18
     DISTR 19-20 SECCION 21-23 ENTREV 24-27
     P0 28 P0A 29-31 P1 TO P6 32-37 P701 TO P703 P801 TO P803 38-49
     P901 TO P903 P1001 TO P1010 50-62 P11 P11A P1201 TO P1214 63-94
     P13 P14 95-96 P1501 TO P1513 97-122 P16 P17 123-124 P18 P19 125-128
     P20 P21 129-130 P22 P2301 TO P2304 131-140 P24 141 P24A 142-143
     P25 144 P26 145-146 P27 147 P27A 148-149 P28 P28A P29 TO P33
     P33A P33B 150-158 P33C P33D 159-162 P33E 163 P34 164-166
     P35 P35A 167-168 P36 169-170 P37 171 P37A P37B 172-173
     P38 174 P3901 TO P3904 175-178 P4001 TO P4005 P4101 TO P4105 179-208(A)
     P42 TO P45 209-212
     I1 TO I9 213-239 E101 TO E103 240-245 E2 246
     E3 247-249 E4 250 C1 251 C1A 252-253 C2 C2A 254-255
     C2B 256-257 C3 258 C4 259-260 p19A 261-262 VOTOSIM 263-264 RECUERDO 265-266
     ESTUDIOS 267 OCUPA 268-269 CONDICION 270-271 ESTATUS 272.


Leyendo 1 registro de `E:\DA2941.txt'.
```

Variable	*Registro*	*Columnas*	*Formato*
CUES	1	5- 9	F5.0
A1	1	1- 9	A9
CCAA	1	10- 11	F2.0
PROV	1	12- 13	F2.0
MUN	1	14- 16	F3.0
TAMUNI	1	17- 17	F1.0
AREA	1	18- 18	F1.0
DISTR	1	19- 20	F2.0

Tendremos entonces tres ventanas abiertas en el programa PSPP: una con la sintaxis, llamada *Editor de sintaxis*, otra con los datos llamada *Editor de datos* y una tercera con los resultados de nuestras acciones en el programa que se llama *Visor de resultados*.

Fijémonos en la ventana de *Editor de datos*, que en la parte inferior izquierda tiene dos pestañas, *Vista de Datos* con la matriz de datos abiertos y *Vista de Variables* con las variables de la encuesta.
Sobre esta ventana de las variables trabajaremos más adelante para hacer las necesarias modificaciones de las variables originales transformándolas para adaptarlas a nuestro ejercicio de regresión logística binaria.
Antes de continuar es recomendable guardar el archivo que hemos abierto y que contiene todos los datos y sus variables, en el nuevo formato proporcionado por el programa PSPP, para que cuando queramos abrirlo de nuevo, no haya que repetir todo ese proceso y podamos abrirlo en lo sucesivo haciendo doble clic sobre el archivo. Llamamos a nuestro archivo, por ejemplo 2941original.sav o REGRE_2941.sav (Evite poner acentos en el nombre del archivo o la carpeta donde quiera guardarlo para evitar que el programa dé error)
Cada vez que realicemos una sesión con nuestro trabajo, guardaremos los contenidos de las tres ventanas. La de datos queda escrita en el archivo REGRE_1941.sav y nos servirá para abrirla en otra sesión. La sintaxis la guardamos en un archivo Sintaxis_1.sps y los resultados en el archivo Resultados_1.spv Nótese que cada tipo de archivo tiene una extensión específica. Al guardarlos, el programa les asigna la extensión automáticamente. (Asegúrese de que su navegador de archivos de Windows muestre las extensiones de los archivos para identificarlos mejor).
Una vez que nos hemos familiarizado con el software y que tenemos los archivos abiertos en el PSPP, podemos plantearnos el ejercicio que nos ocupa.

Selección y preparación de las variables

Selección y preparación de la variable dependiente

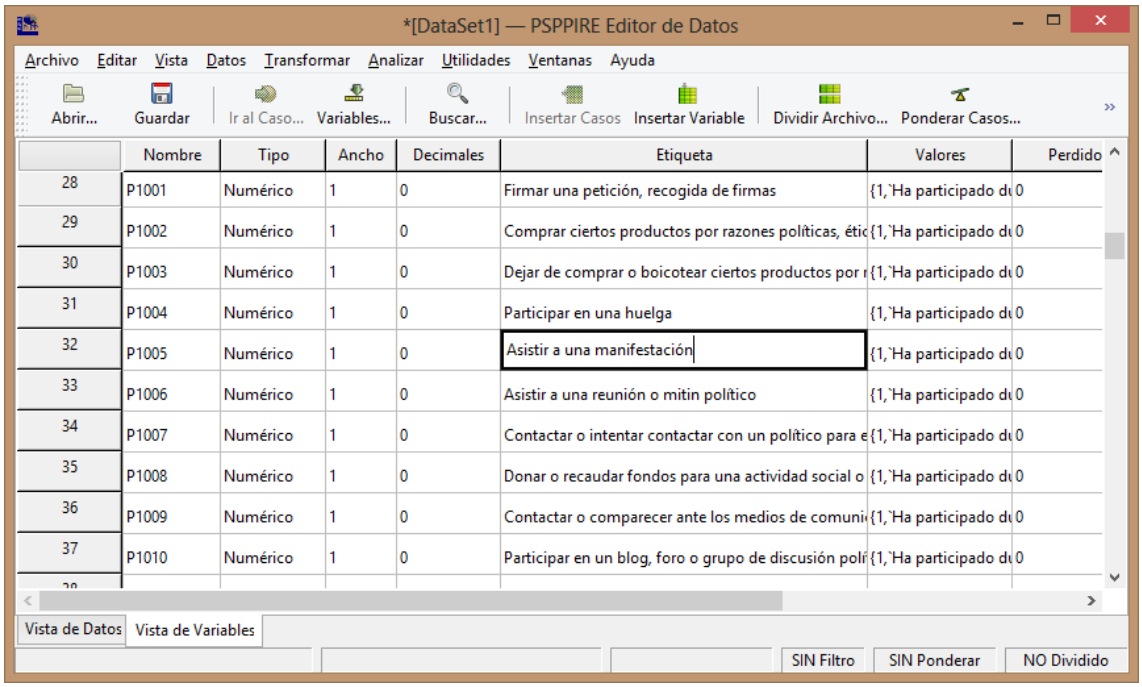
Como hemos anunciado más arriba, tendremos que seleccionar las variables del barómetro que emplearemos en nuestro análisis y transformarlas para su tratamiento en el modelo de regresión logística binaria. Para tener una visión general de las variables de la encuesta conviene abrir el archivo **cues2941.pdf** que obtuvimos de la página del CIS e imprimirlo.

En el cuestionario localizaremos fácilmente la variable que vamos a emplear como variable dependiente. Recordemos que en el modelo logístico binario la variable dependiente debe ser dicotómica, es decir, sólo debe tomar dos valores, 0 y 1.

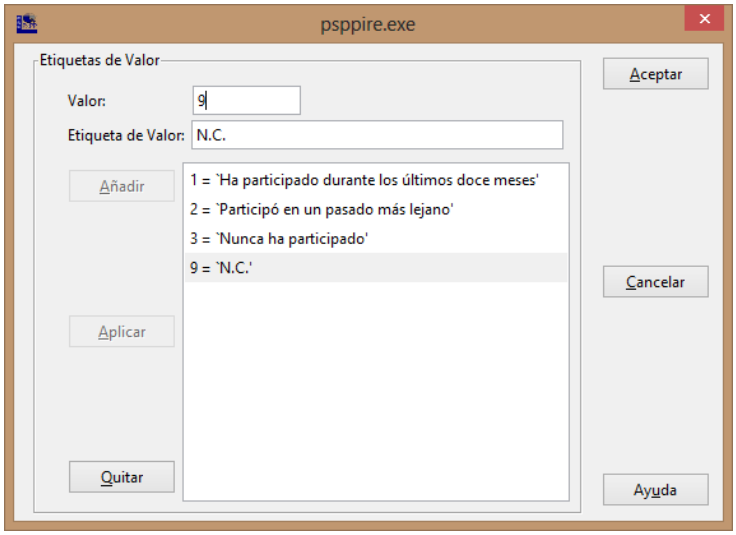
En el cuestionario las preguntas están numeradas y en la segunda página, bajo el rótulo “P10” encontramos una serie de preguntas enunciadas conjuntamente de la siguiente manera: “*Existen diversas formas de participación en acciones sociales y políticas que la gente puede llevar a cabo. Por favor indíqueme para cada una de ellas si Ud.*”:

1. Ha participado en los últimos doce meses
2. Participó en un pasado más lejano
3. Nunca ha participado

En realidad bajo el rótulo P10 se hacen 10 preguntas P101, P102, P103, P104, P105, P106, P107, P108, P109 y P110 (como podemos ver en el listado de variables de la pestaña *Vista de Variables* en la imagen que aparece bajo estas líneas) relacionadas con la participación y referidas a: firmar una petición o recogida de firmas, comprar ciertos productos por razones políticas o éticas, dejar de comprar o boicotear ciertos productos, participar en una huelga, asistir a una manifestación, asistir a una reunión o mitin, contactar con un político para expresarle opiniones, donar o recaudar fondos para una actividad, contactar o comparecer ante los medios de comunicación para expresar opiniones o participar en un blog, foro o grupo de discusión política en internet.

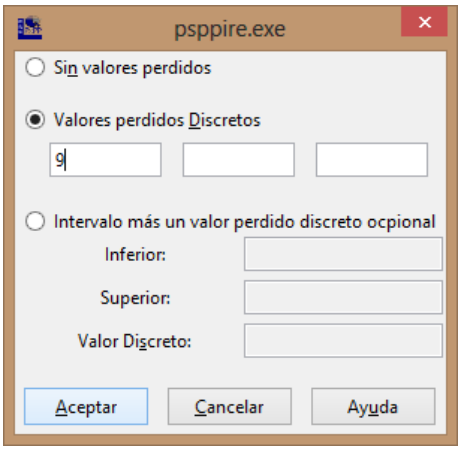


Para nuestro ejercicio hemos seleccionado como variable dependiente la pregunta concreta P105: “Asistir a una manifestación”. Será la variable dependiente de nuestro modelo y está codificada con los siguientes valores: (1), (2), (3), (9). Valores que podemos ver pinchando en la correspondiente casilla de *Valores* de esta Vista de Variables:



Debemos recodificar la variable de forma que sólo tenga dos categorías, entonces establecemos una categoría que recoja a los que han ido a una manifestación en los últimos 12 meses con el valor (1), y los casos con valores (2) y (3) los recodificamos con el valor (0). Además tenemos algunos casos codificados como “no contesta” con el valor 9, y que vamos a eliminar de los cálculos, para lo cual debemos considerarlos “casos perdidos”, estableciendo el valor perdido desde la pestaña de variables, seleccionando la casilla de perdidos y añadiendo el código 9.

Comenzamos por eliminar los casos marcados con el código 9 desde la ventana de variables. Pinchamos en la casilla correspondiente de la columna *Perdidos* para la variable P105, ponemos el valor 9 en una de las casillas de Valores perdidos discretos y aceptamos:

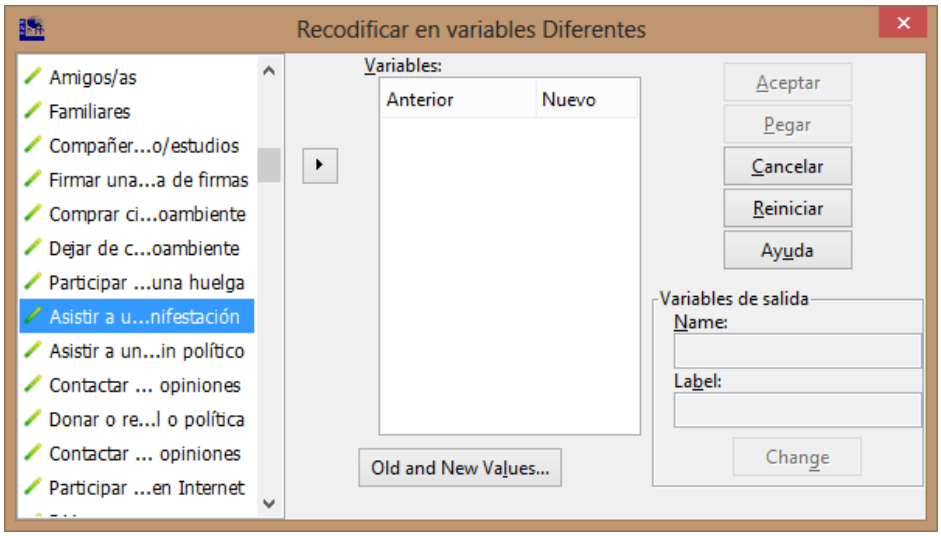


Desde ese momento no se tendrán en cuenta para los cálculos las respuestas de las personas que no contestaron a esa pregunta.

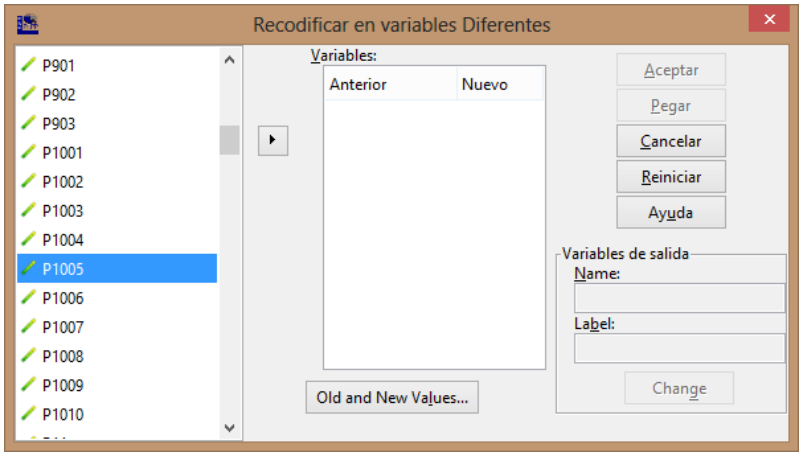
Para recodificar los valores de la variable, podemos hacerlo desde la matriz de datos, seleccionando la opción desde el menú superior:

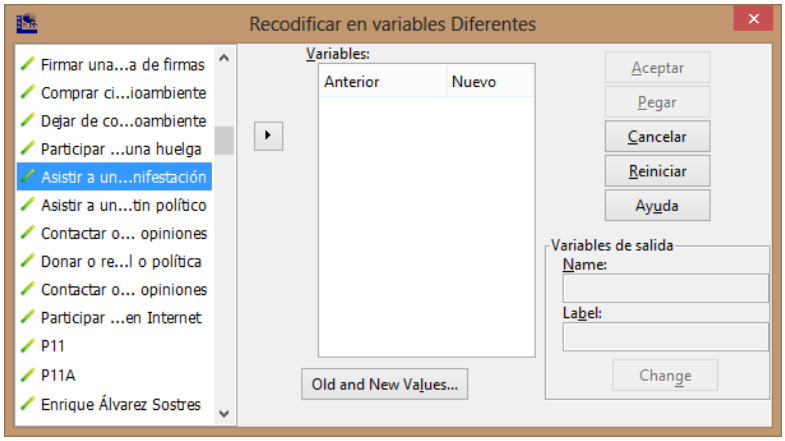
Transformar-> recodificar en variables diferentes

Se nos abrirá una nueva ventana:



Pinchando con el botón derecho del ratón en la lista de variables que aparecen podemos elegir entre ver los nombres de las variables o sus códigos, como se ve en las siguientes imágenes:

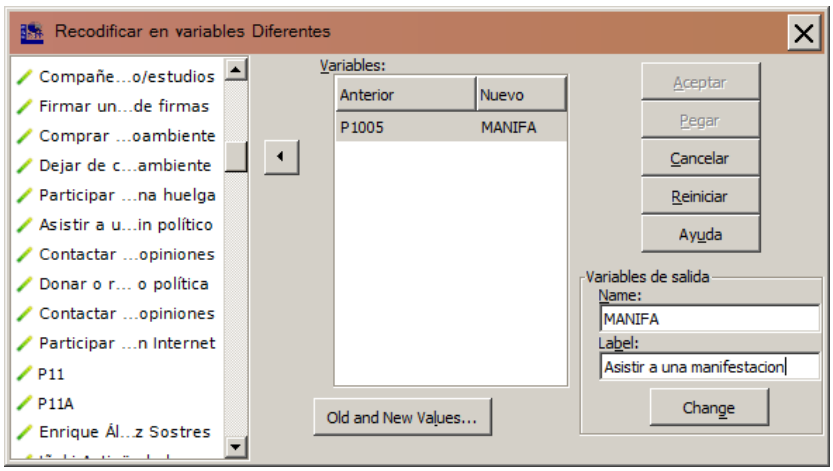




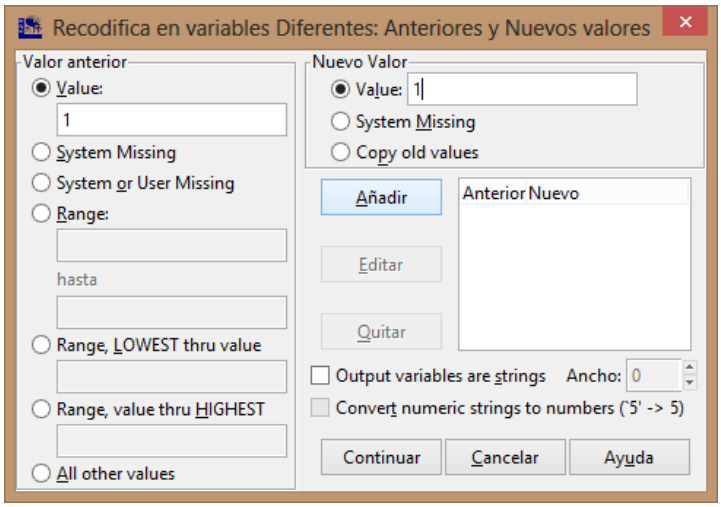
Como norma general utilizaremos siempre la opción (recodificar en variables diferentes), creando una variable nueva, para conservar así la variable original por si quisiéramos volver sobre ella en otro momento, pues si lo hacemos en la misma variable, se sobre-escribe y perdemos la información de la variable original.

Seleccionamos la variable P1005 y la pasamos a la casilla de la derecha con el botón

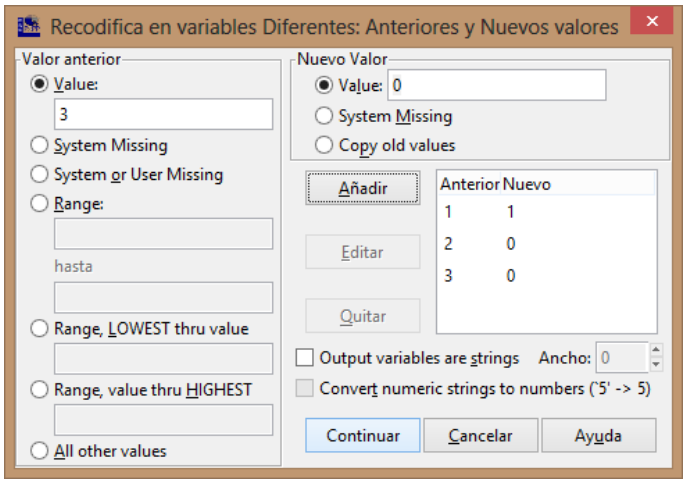
Marcamos la P1005 en la nueva lista y asignamos un nombre para la variable de salida: por ejemplo MANIFA y una etiqueta, por ejemplo: Asistir a una manifestación



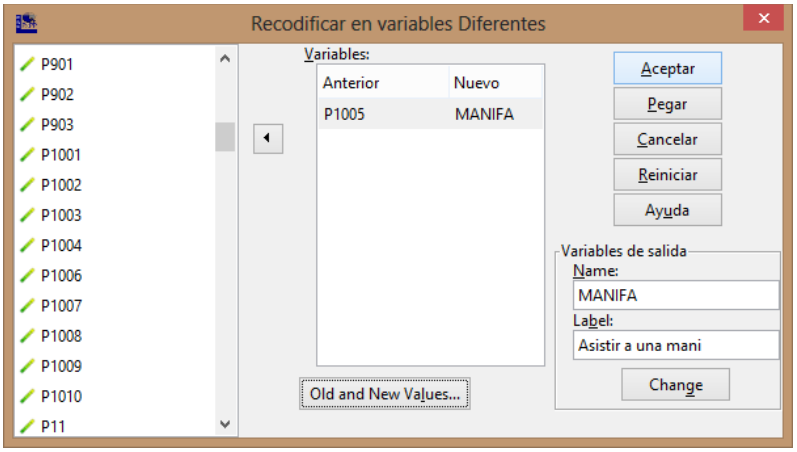
Pinchamos a continuación en el botón *Change* para que asuma el nuevo nombre. Después asignamos los nuevos valores a la variable pinchando en *Old and New Value,* con lo que se nos abrirá otra ventana:



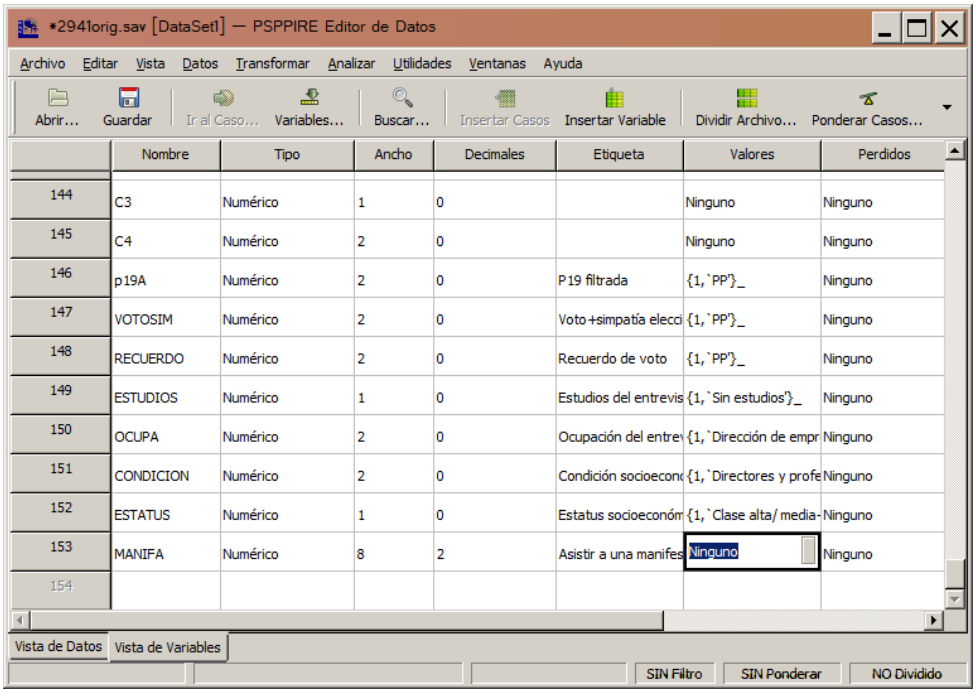
El valor 1 era haber asistido a una manifestación en los últimos 12 meses, y será también el valor 1 en nuestra nueva variable. Escribimos el valor 1 como *valor anterior*, y 1 como *nuevo valor* y pinchamos Anadir. En esta misma ventana, sucesivamente, ponemos el valor 2 y el 3 con el nuevo valor 0 para que quede así:



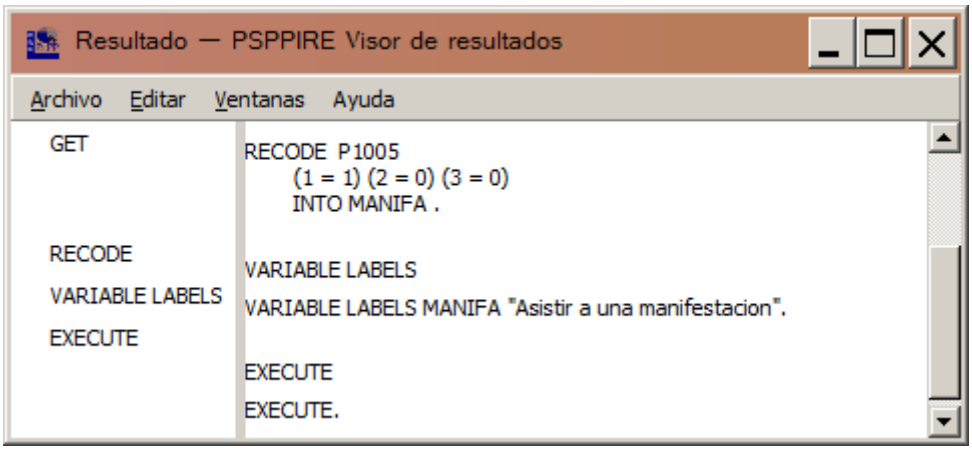
Pinchamos en la casilla *Continuar* y regresaremos a la ventana de recodificación para *Aceptar*:



Si vamos a la matriz de datos, en la pestaña de variables veremos que se ha añadido la nueva variable al final del todo:



NOTA IMPORTANTE: Antes de continuar vamos a ver lo que aparece en la ventana de resultados tras la transformación realizada:

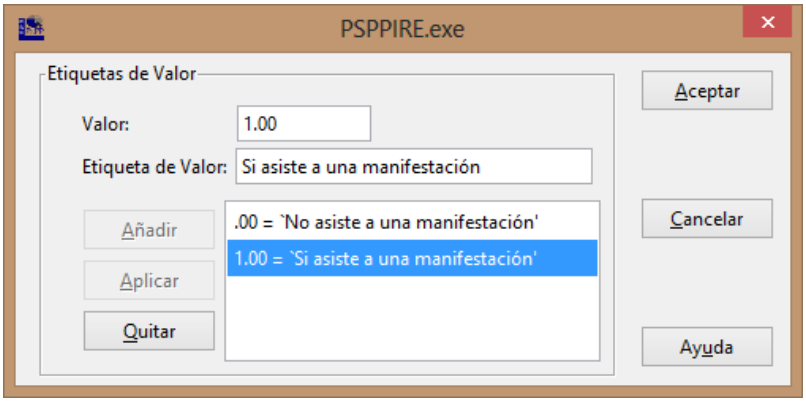


Podemos ver en la ventana de resultados que al ejecutar el comando de recodificar en una variable diferente mediante los menús del programa, en realidad el PSPP ha ejecutado el código que puede verse (Observe que los comandos siempre terminan con un punto “.”).

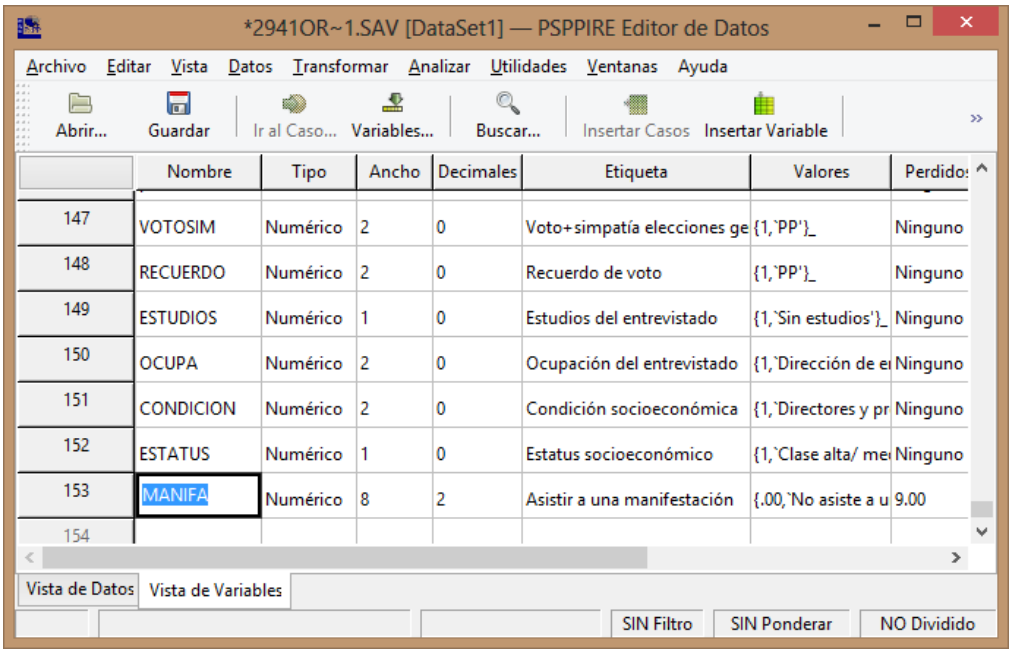
```
RECODE P1005 (1=1) (2=0) (3=0) INTO MANIFA.
VARIABLE LABELS MANIFA “Asistir a una manifestacion”.
```

Hemos asignado la etiqueta “Asistir a una manifestación” a la variable MANIFA y los valores 0 y 1, pero hemos de añadir también *etiquetas de valor* para que al leer los datos sepamos qué es

0 y qué es 1. Lo hacemos desde la Vista de variables pinchando en la casilla Valores de la variable MANIFA, para añadir los dos valores sucesivamente:



En la vista de variable veremos cómo ha quedado:



Para añadir etiquetas a los valores (0) y (1) de nuestra nueva variable MANIFA también podemos optar por utilizar los menús de la GUI que PSPP nos ofrece, como hemos hecho más arriba, o bien teclear los comandos en una ventana de sintaxis.

En realidad podríamos haber escrito ese código directamente en una ventana de sintaxis, ejecutarlo, y habría hecho lo mismo. Este código de texto podemos guardarlo o bien copiarlo de un documento y ejecutarlo después. También podemos almacenar varias instrucciones escritas y ejecutarlas todas más tarde. Con el tiempo nos acostumbraremos a escribir el código en lugar de emplear los menús y el ratón y comprobaremos que es más sencillo.

La alternativa a los menús y el ratón es teclear el comando:

```
VALUE LABELS /MANIFA 0 "No asiste a una manifestación" 1 "Si asiste a una manifestación".
```

De la misma manera la consideración del valor 9 original asignado a los que no contestan y que hemos considerado caso perdido, también podría haberse introducido tecleando sintaxis:

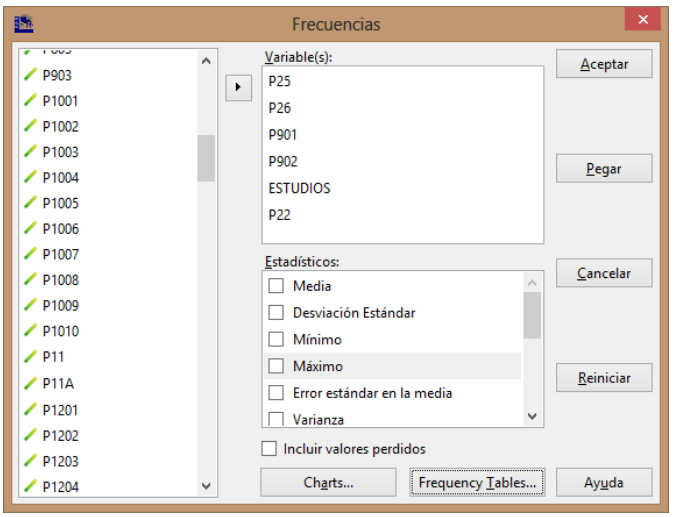
```
MISSING VALUES MANIFA (9).
```

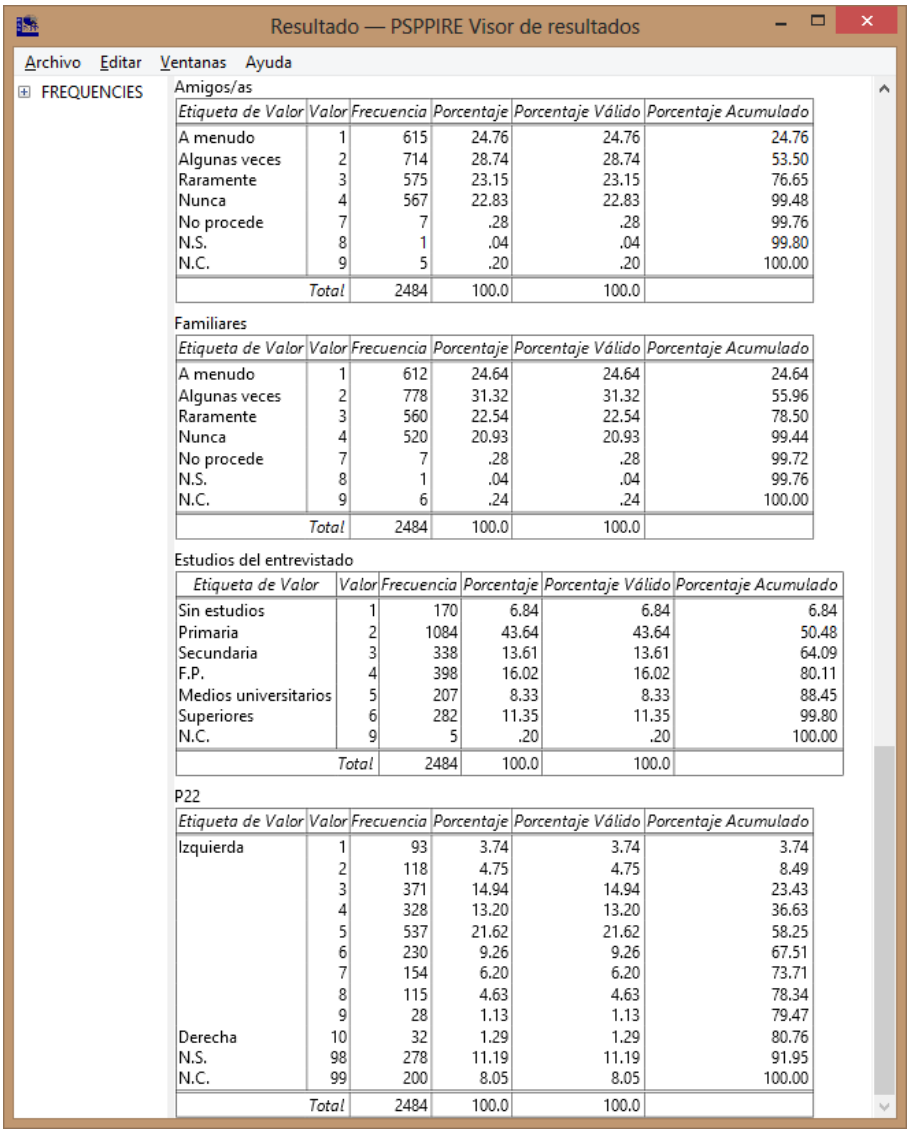
Tenemos ya preparada la variable dependiente de nuestro modelo con los valores 1 ("Si asiste a una manifestación" para los que han participado en los últimos 12 meses) y 0 ("No asiste a una manifestación" para los que participaron en el pasado o bien nunca ha participado).

Selección y preparación de las variables independientes

Pasemos ahora a preparar el resto de las variables que vamos a utilizar en nuestro análisis. Las variables independientes que elegimos del cuestionario serán el sexo, la edad, el nivel de estudio, hablar de política con los amigos y familiares.

Para saber si hay que tener en cuenta los "no sabe", los "No contesta" o los "No procede" en cada una de las variables a utilizar, bastará con observar las frecuencias de la variable. Para ello podemos seleccionar en el menú superior *Analizar->estadística descriptiva->Frecuencias* incluir las variables que vamos a emplear y observar luego las tablas de frecuencias:





Resultado — PSPPIRE Visor de resultados

Archivo Editar Ventanas Ayuda

FREQUENCIES

Amigos/as

Etiqueta de Valor	Valor	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
A menudo	1	615	24.76	24.76	24.76
Algunas veces	2	714	28.74	28.74	53.50
Raramente	3	575	23.15	23.15	76.65
Nunca	4	567	22.83	22.83	99.48
No procede	7	7	.28	.28	99.76
N.S.	8	1	.04	.04	99.80
N.C.	9	5	.20	.20	100.00
	Total	2484	100.0	100.0	

Familiares

Etiqueta de Valor	Valor	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
A menudo	1	612	24.64	24.64	24.64
Algunas veces	2	778	31.32	31.32	55.96
Raramente	3	560	22.54	22.54	78.50
Nunca	4	520	20.93	20.93	99.44
No procede	7	7	.28	.28	99.72
N.S.	8	1	.04	.04	99.76
N.C.	9	6	.24	.24	100.00
	Total	2484	100.0	100.0	

Estudios del entrevistado

Etiqueta de Valor	Valor	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
Sin estudios	1	170	6.84	6.84	6.84
Primaria	2	1084	43.64	43.64	50.48
Secundaria	3	338	13.61	13.61	64.09
F.P.	4	398	16.02	16.02	80.11
Medios universitarios	5	207	8.33	8.33	88.45
Superiores	6	282	11.35	11.35	99.80
N.C.	9	5	.20	.20	100.00
	Total	2484	100.0	100.0	

P22

Etiqueta de Valor	Valor	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
Izquierda	1	93	3.74	3.74	3.74
	2	118	4.75	4.75	8.49
	3	371	14.94	14.94	23.43
	4	328	13.20	13.20	36.63
	5	537	21.62	21.62	58.25
	6	230	9.26	9.26	67.51
	7	154	6.20	6.20	73.71
	8	115	4.63	4.63	78.34
	9	28	1.13	1.13	79.47
Derecha	10	32	1.29	1.29	80.76
N.S.	98	278	11.19	11.19	91.95
N.C.	99	200	8.05	8.05	100.00
	Total	2484	100.0	100.0	

Las variable sexo y edad no tienen ningún caso de ausencia de respuesta, pero como vemos en la tabla las variables de Hablar de política con los amigos, Hablar de política con familiares, Estudios y Escala política Izquierda-derecha si cuentan con casos que hemos de considerar perdidos, para lo cual podemos seguir el procedimiento que vimos con la variable P1005 al principio.

```
MISSING VALUES P22 (98,99).
MISSING VALUES P901 P902 (7,8,9).
```

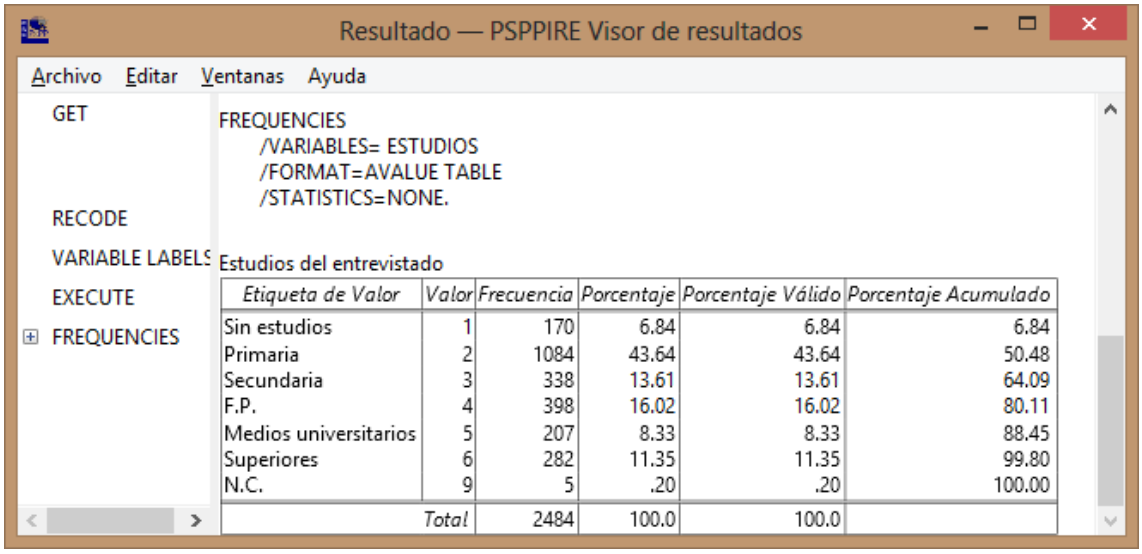
Si bien este paso podría hacerse incluyendo los valores perdidos en el comando de recodificación, como haremos en adelante.

La variable sexo (P25), podemos dejarla como está: 1 Hombre y 2 Mujer. Y será tratada como variable categórica, no numérica. De esa forma, cuando leamos los resultados del análisis la categoría de referencia será la última (2 mujer) y el Odd ratio nos dirá cuántas veces será más

probable que un hombre vaya a una manifestación, en relación a una mujer. (Téngase en cuenta que podría codificarse al revés, en cuyo caso la referencia sería la categoría de los hombres y el Odd ratio nos diría el número de veces que las mujeres van a la manifestación en relación a los hombres).

La variable edad (P26), puede dejarse con los valores que vienen en la encuesta (años cumplidos del entrevistado, en cuyo caso será una variable de intervalo). En este caso el valor del Odd ratio que resulte se referirá al incremento de posibilidades de ir a una manifestación por cada año cumplido. (La podríamos recodificar agrupando las edades de los entrevistados en grupos o categorías de, por ejemplo, 5 años. En cuyo caso el Odd ratio se referiría a la razón entre las probabilidades para una categoría respecto a la siguiente en cada tramo de 5 años).

El nivel de estudios de la encuesta se corresponde con las preguntas P27 y P27a, pero utilizaremos la variable sintética codificada por el CIS y que aparece entre las últimas variables de la lista, que podemos ver en la Vista de Variables de nuestra ventana de datos, bajo el nombre de ESTUDIOS, con la etiqueta "Estudios del entrevistado", con 6 categorías:



Etiqueta de Valor	*Valor*	*Frecuencia*	*Porcentaje*	*Porcentaje Válido*	*Porcentaje Acumulado*
Sin estudios	1	170	6.84	6.84	6.84
Primaria	2	1084	43.64	43.64	50.48
Secundaria	3	338	13.61	13.61	64.09
F.P.	4	398	16.02	16.02	80.11
Medios universitarios	5	207	8.33	8.33	88.45
Superiores	6	282	11.35	11.35	99.80
N.C.	9	5	.20	.20	100.00
	Total	2484	100.0	100.0	

Recodificaremos esta variable en otra que se llame ESTU2 con 4 categorías: estudios superiores, medios, otros, sin estudios (y eliminaremos el valor 9 de los que no contestan).

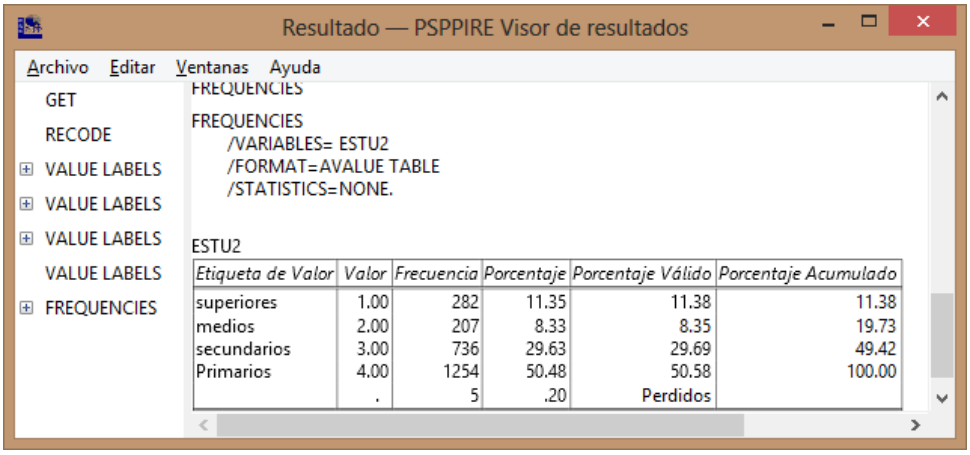
Para mayor comodidad en la lectura de los resultados del análisis, además de reducir el número de categorías a 4, invertiremos los valores de las categorías recodificándolos valores antiguos agrupando e invirtiendo el orden:

Valores antiguos	Valores actuales
Sin Estudios (1) Primaria (2) Secundaria (3) F.P. (4) Medios universitarios (5) Superiores (6)	Sin estudios y Primaria (4) Secundaria y F.P. (3) Medios universitarios (2) Superiores (1)

De esa forma la interpretación de los coeficientes resultantes de los cálculos se interpretarán en referencia a la última categoría (Sin estudios).

```
RECODE ESTUDIOS (1 =4) (2 = 4) (3 = 3) (4 = 3) (5=2) (6=1) (9 = SYSMIS) INTO ESTU2.
VALUE LABELS /ESTU2 1 “superiores” 2 “medios” 3 “secundarios” 4 “Primarios”.
```

Obtenemos las frecuencias de la nueva variable:



Etiqueta de Valor	Valor	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
superiores	1.00	282	11.35	11.38	11.38
medios	2.00	207	8.33	8.35	19.73
secundarios	3.00	736	29.63	29.69	49.42
Primarios	4.00	1254	50.48	50.58	100.00
	.	5	.20	Perdidos	

Obsérvese que se han anulado los casos perdidos (SYSMIS perdidos por el sistema) en el comando de recodificación.

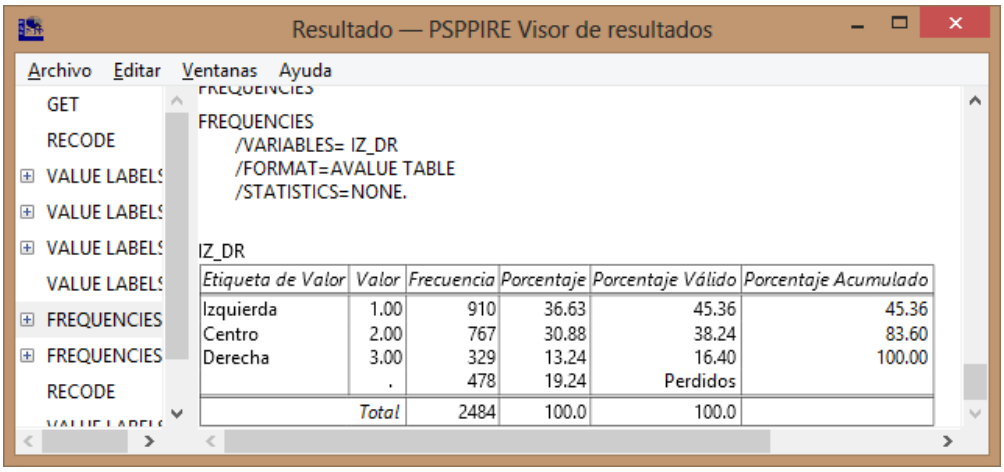
La variable P22 de escala izquierda-derecha presenta una distribución de frecuencias muy nutrida en el centro y con pocos casos en los extremos, por tanto bien podemos agrupar la variable en 3 categorías:

- 1,2,3 y 4 Izquierda (1)
- 5 y 6 Centro (2)
- 7,8,9 y 10 Derecha (3)

```
RECODE P22 (1 =1) (2 = 1) (3 = 1) (4 = 1) (5 = 2) (6 = 2) (7 =3) (8 = 3) (9 = 3) (10 = 3) (98,99 =
SYSMIS) INTO IZ_DR.

VALUE LABELS /IZ_DR 1 “Izquierda” 2 “Centro” 3 “Derecha”.
```

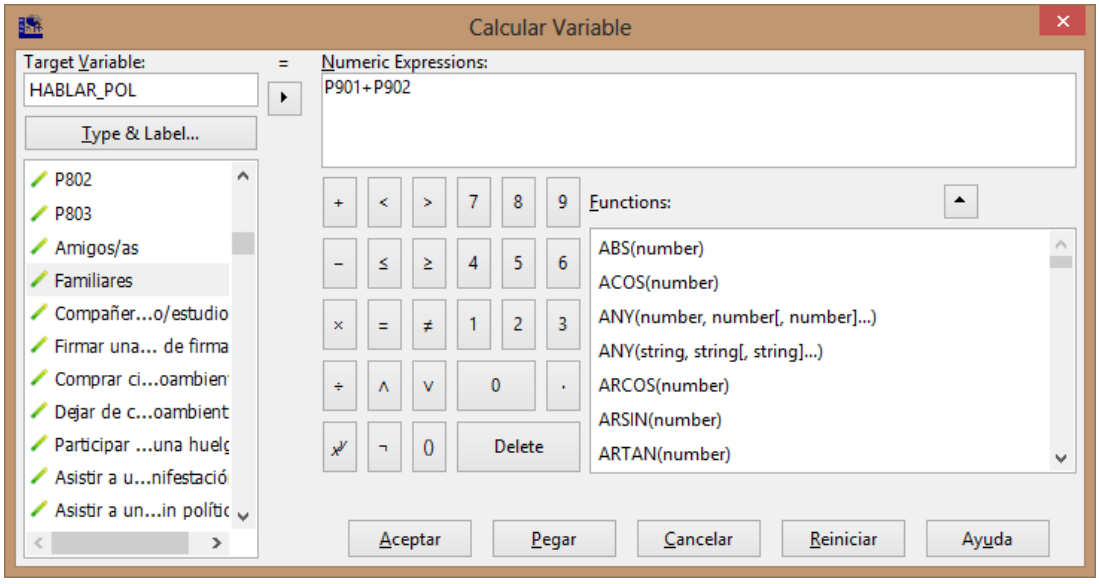
Obtenemos las frecuencias de la nueva variable:



En cuanto a la variable P901 habar de política con los amigos y P902 hablar de política con familiares, las dejamos como están en 4 categorías.
Para mejor interpretación de los valores numéricos y evitar incluir valores perdidos sumados a las respuestas en otra variable los anularemos antes de generar la nueva variable con el comando (Si no lo hemos hecho al principio):

```
MISSING VALUES P901 P902 (7,8,9).
```

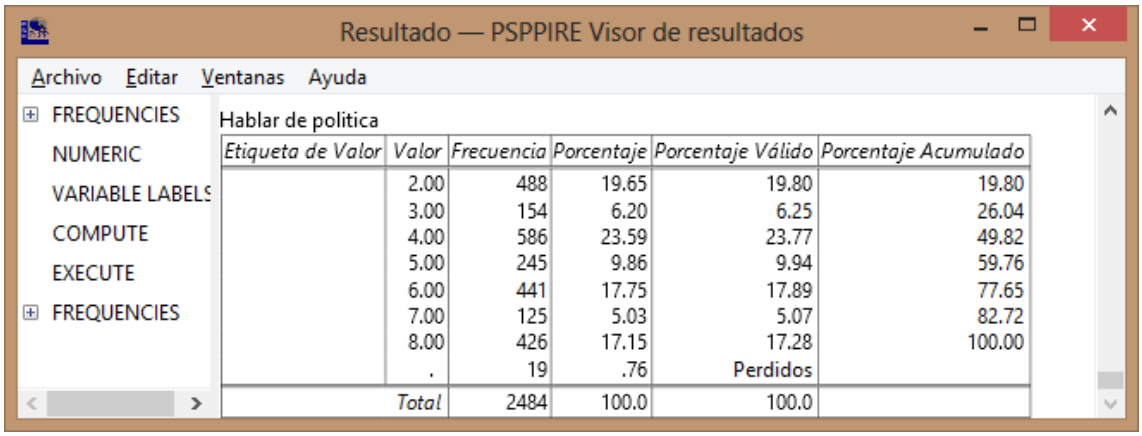
Podemos además construir una variable a partir de estas dos, sumando los valores de ambas: Obtendríamos una variable "HABLAR DE POLÍTICA con amigos o familiares" con 8 valores. Veamos el procedimiento de construcción de esta nueva variable.
En el menú superior seleccionamos *Transformar->Calcular* y nos sale una ventana en la que ponemos el nombre de la nueva variable HABLA_POL en la casilla *Target Variable* y su etiqueta en la opción *Type&Label*. A continuación pasamos a la ventana del cálculo la variable P901, añadimos el signo de suma (+) del menú inferior y pasamos la variable P902 :



La sintaxis correspondiente sería:

```
MISSING VALUES P901 P902 (7,8,9).
NUMERIC HABLA_POL.
VARIABLE LABEL HABLA_POL "Hablar de politica".
COMPUTE HABLA_POL = P901+P902.
EXECUTE.
```

Cuando aceptamos hemos construido una nueva variable suma de las dos anteriores. Hemos obtenido una escala desde hablar mucho (2), hasta no hablar (8):



Hablar de politica

Etiqueta de Valor	Valor	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
	2.00	488	19.65	19.80	19.80
	3.00	154	6.20	6.25	26.04
	4.00	586	23.59	23.77	49.82
	5.00	245	9.86	9.94	59.76
	6.00	441	17.75	17.89	77.65
	7.00	125	5.03	5.07	82.72
	8.00	426	17.15	17.28	100.00
	.	19	.76	Perdidos	
	Total	2484	100.0	100.0	

Podríamos añadir etiquetas tales como muchísimo, mucho, bastante, con frecuencia, alguna vez, pocas veces, raramente y nunca. Pero puede bastar con los valores de la escala numérica del 1 al 8, siendo 1 el máximo y 8 el mínimo[2]. Con esta codificación, la **categoría de referencia** para la interpretación de los valores de los Odds ratio será la categoría con valor **8**, es decir la categoría que agrupa a los que **no hablan nunca de política** ni con familiares ni con amigos.

Como hemos señalado al hablar de cada variable, el tratamiento de las variables de intervalo y las categóricas es diferente. Y al introducir el comando que calcula la regresión logística tendremos que diferenciarlas.

Podemos hacer un breve esquema del modelo que va a relacionar una variable dependiente dicotómica con otras variables categóricas o de intervalo. Dicho modelo nos servirá para conocer las relaciones entre las diferentes variables, con expresión del comportamiento de las diferentes categorías de esas variables.

[2] Téngase en cuenta que no se trata tanto de analizar la influencia de las conversaciones políticas en la asistencia a las manifestaciones, sino de un ejercicio de manejo del programa PSPP para hacer análisis de regresión logística. Lo mismo puede decirse de la nueva categorización de los niveles de estudio y la escala ideológica. Deben tomarse como un simple ejercicio de uso del programa.

Como se ha señalado en el apartado teórico no pretenderemos hacer un modelo predictivo, sino indagar sobre la relación entre las variables. Además, dadas las condiciones de los datos obtenidos de una encuesta del CIS, que no fueron recogidas con el propósito del tipo de predicciones que suele hacerse con la regresión logística, y el objetivo de este sencillo ejercicio que pretende sobre todo mostrar la práctica del manejo del software que utilizaremos en esta asignatura.

Modelo:

Una vez modificadas las variables que vamos a emplear en el modelo logístico, podemos pasar a ejecutar el cálculo:

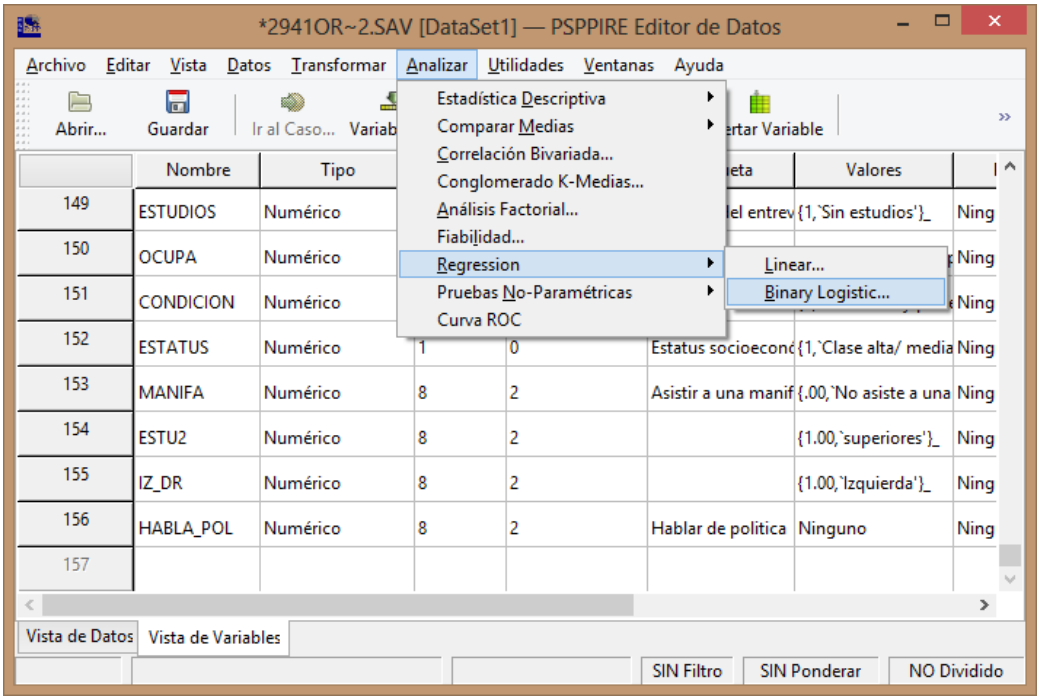
Variable dependiente:

- Asistir a una manifestación (Dicotómica)

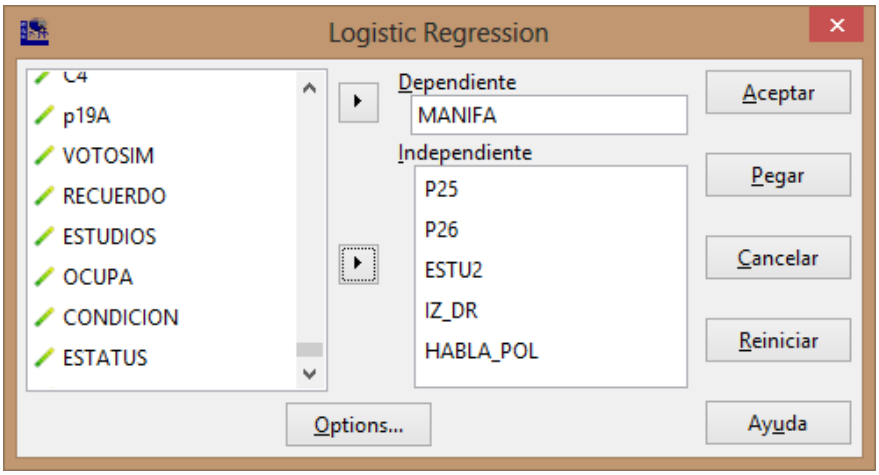
Variables independientes:

- Sexo (Categórica
- Edad (Intervalo)
- Estudios (Categórica)
- Hablar de política (Categórica)
- Escala I-D (Categórica)

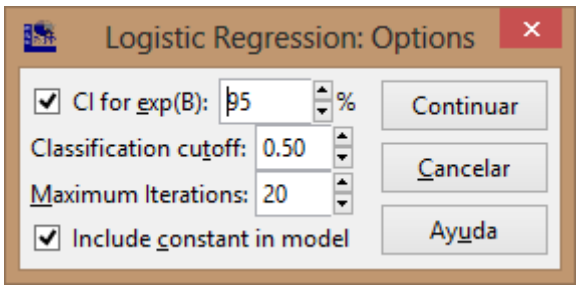
El comando LOGISTIC REGRESSION se encuentra en el menú superior del *Editor de Datos*, y para seleccionarlo debemos seguir la secuencia *Analizar->Regression->Binary Logistic*



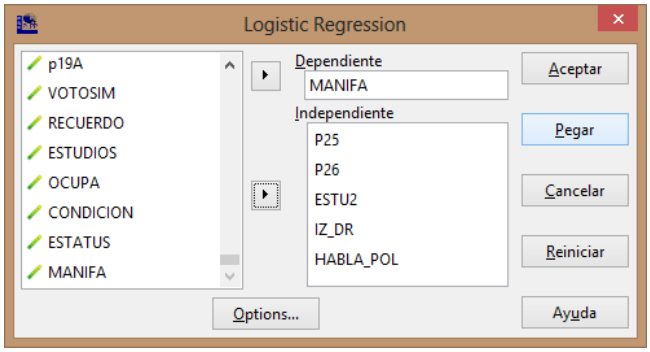
Se nos abrirá una ventana para especificar los parámetros del comando:
Pondremos en las casillas correspondientes la variable dependiente MANIFA y las variables independientes P25, P26, ESTU2, IZ_DR y HABLA_POL



Seleccionaremos en esta ventana el botón *Options* y se nos abre una nueva ventana en la que marcaremos la opción CI for exp (B) y dejaremos las otras opciones marcadas por defecto como están.



Una vez hecho esto pulsamos el botón Continuar, con lo que volvemos a la anterior ventana, pero **NO** pulsaremos directamente en ***Aceptar***, sino que seleccionaremos la opción ***Pegar***



La razón de esto es que el software PSPP no tiene implementada una opción esencial para nuestro cálculo, que es la que especifica qué variables son de intervalo y qué variables son categóricas, pues tienen un tratamiento diferente.

Una vez que hemos hecho Pegar, podemos cerrar esa ventana. El resultado de Pegar es que la instrucción no se ha ejecutado, sino que se ha escrito en una nueva ventana de sintaxis, para que podamos modificarla. En la ventana de sintaxis aparece el siguiente texto:

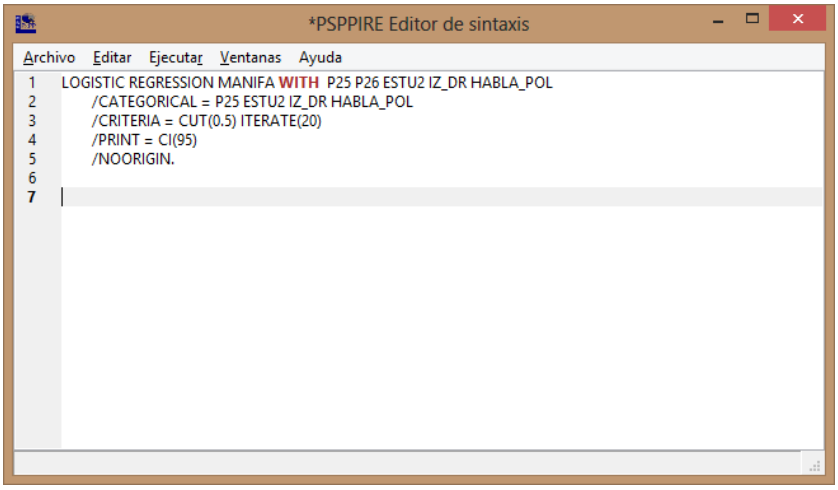
```
LOGISTIC REGRESSION MANIFA WITH  P25 P26 ESTU2 IZ_DR HABLA_POL
        /CRITERIA = CUT(0.5) ITERATE(20)
        /PRINT = CI(95)
        /NOORIGIN.
```

Debemos especificar las variables categóricas escribiendo como segunda línea:

```
        /CATEGORICAL = P25 HABLA_POL IZ_DR ESTU2
```

```
LOGISTIC REGRESSION MANIFA WITH  P25 P26 ESTU2 IZ_DR HABLA_POL
        /CATEGORICAL = P25 ESTU2 IZ_DR HABLA_POL
        /CRITERIA = CUT(0.5) ITERATE(20)
        /PRINT = CI(95)
        /NOORIGIN.
```

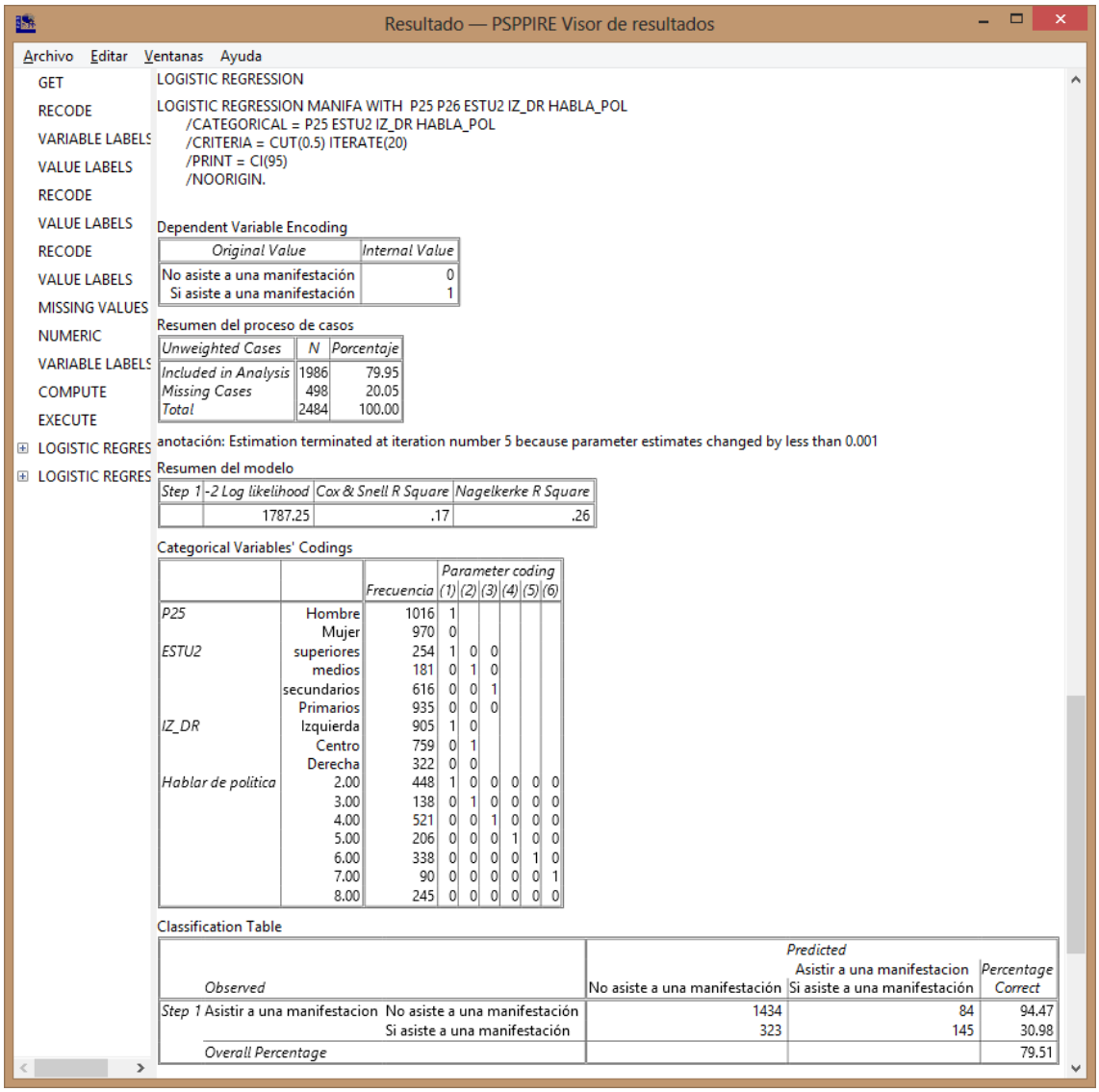
El texto de la sintaxis debe quedar como se muestra en la siguiente ventana:



(Tenga en cuenta que no debe añadir ninguna letra ni signo de puntuación más).

Una vez corregida la sintaxis seleccionaremos en el menú superior la opción *Ejecutar->Todos*

El programa PSPP ejecutará el comando y mostrará inmediatamente el resultado en la ventana de resultados como se muestra más abajo (he partido la ventana en dos partes dado que los nuevos ordenadores tienen la pantalla apaisada y sólo es visible una parte de la ventana, pues hay que hacer Scroll para desplazarse a la parte inferior):



Como puede verse en la ventana de resultados en primer lugar nos informan del comando que se ha ejecutado.

A continuación figura una tabla con las categorías de la variable dependiente.

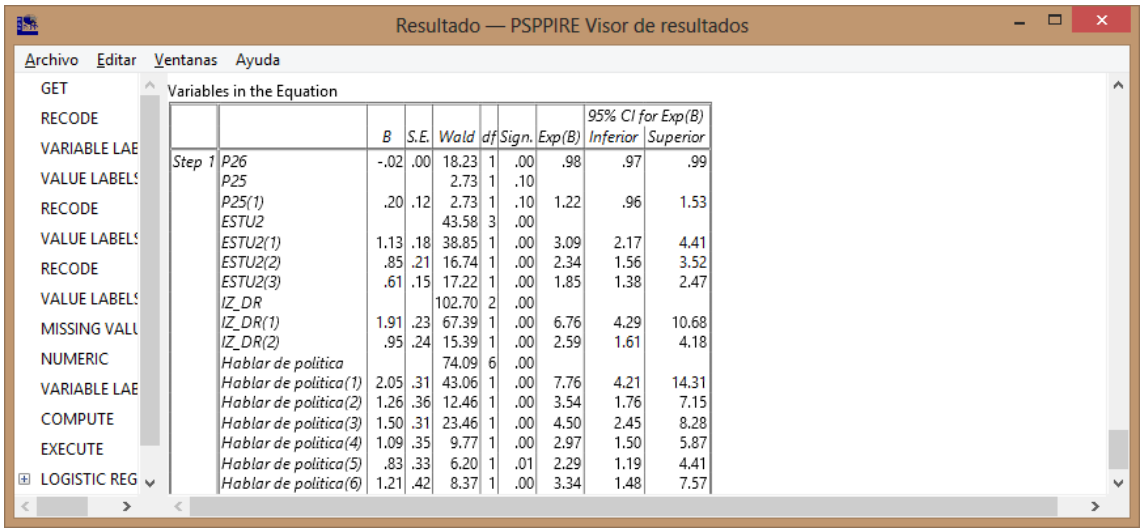
Más abajo nos informa del resumen de casos que han intervenido en el proceso

Una línea de texto nos informa de que aunque habíamos marcado el número de iteraciones en 20, el programa sólo ha realizado 5 y ha suspendido ahí el análisis al no apreciarse modificaciones superiores a 0,001

A continuación otra tabla muestra los estadísticos calculados.

Otra tabla más nos muestra las variables categóricas del proceso y su codificación

La Tabla de clasificación muestra la clasificación de los casos y el número de casos que ha conseguido predecir el modelo en número de casos y en porcentaje



Variables in the Equation

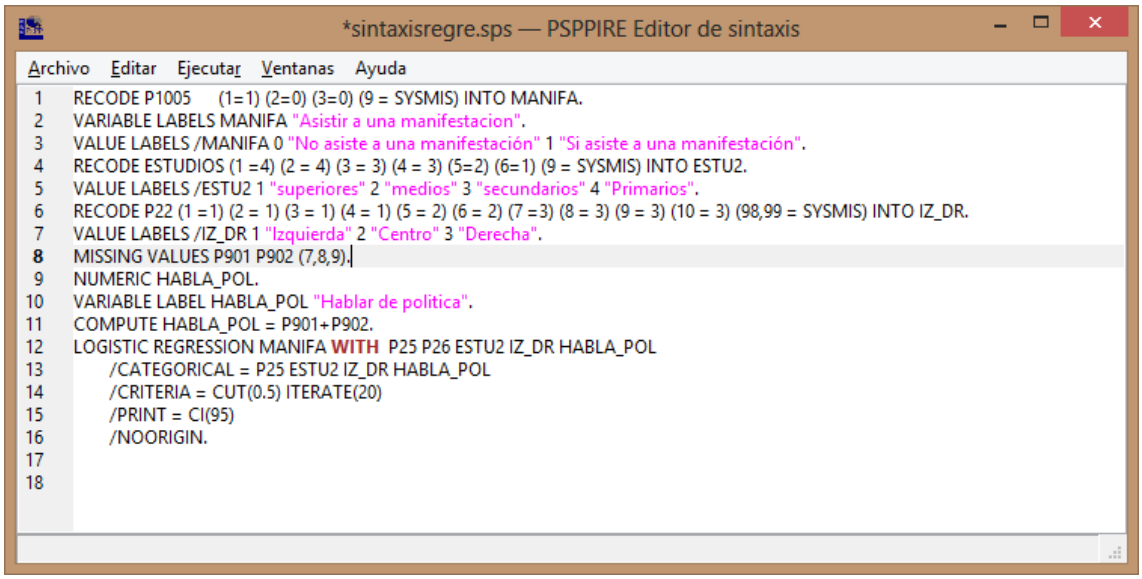
		B	S.E.	Wald	df	Sign.	Exp(B)	95% CI for Exp(B) Inferior	Superior
Step 1	P26	-.02	.00	18.23	1	.00	.98	.97	.99
	P25			2.73	1	.10			
	P25(1)	.20	.12	2.73	1	.10	1.22	.96	1.53
	ESTU2			43.58	3	.00			
	ESTU2(1)	1.13	.18	38.85	1	.00	3.09	2.17	4.41
	ESTU2(2)	.85	.21	16.74	1	.00	2.34	1.56	3.52
	ESTU2(3)	.61	.15	17.22	1	.00	1.85	1.38	2.47
	IZ_DR			102.70	2	.00			
	IZ_DR(1)	1.91	.23	67.39	1	.00	6.76	4.29	10.68
	IZ_DR(2)	.95	.24	15.39	1	.00	2.59	1.61	4.18
	Hablar de politica			74.09	6	.00			
	Hablar de politica(1)	2.05	.31	43.06	1	.00	7.76	4.21	14.31
	Hablar de politica(2)	1.26	.36	12.46	1	.00	3.54	1.76	7.15
	Hablar de politica(3)	1.50	.31	23.46	1	.00	4.50	2.45	8.28
	Hablar de politica(4)	1.09	.35	9.77	1	.00	2.97	1.50	5.87
	Hablar de politica(5)	.83	.33	6.20	1	.01	2.29	1.19	4.41
	Hablar de politica(6)	1.21	.42	8.37	1	.00	3.34	1.48	7.57

Finalmente, la última tabla de *Variables en la ecuación* nos muestra los exponentes de la ecuación de regresión logística binaria correspondientes a cada variable (caso de la variable de intervalo P25 que es de intervalo) o a las categorías (resto de las variables categóricas). En el caso de la edad, B= -0,02 con signo negativo, indica que la probabilidad de que una persona asista a una manifestación disminuye 0,02 por cada año cumplido (ya que tiene signo negativo) Mientras que el valor B= 1,91 de la categoría "Izquierda" de la escala ideológica en la variable IZ_DR significa que la probabilidad de que alguien de esa categoría vaya a una manifestación en comparación con la categoría de referencia, que es la derecha, se incrementa un 1,9 veces.

En la Columna Exp(B) se presenta el valor del Odd Ratio, que para la edad es 0,98 cuántas veces es más probable que un individuo de determinada edad vaya a una manifestación respecto a otro de una edad un año mayor, y que al ser menor que la unidad habla de una disminución. Para la Variable de la escala ideológica Exp(B)=6,79 significa que es casi siete veces más probable que una persona de izquierdas vaya a una manifestación en comparación con una de derechas. Dichos valores vienen dados en forma de intervalo con un nivel de confianza del 95% visible en la parte derecha de la tabla y con un nivel de significación que figura en la columna titulada "Sign.". Como puede verse, el nivel de significación es bueno para todas las variables (y sus categorías) excepto para el sexo P25.

RESUMEN DE SINTAXIS DE LOS COMANDOS:

```
RECODE P1005 (1=1) (2=0) (3=0) (9 = SYSMIS) INTO MANIFA.
VARIABLE LABELS MANIFA "Asistir a una manifestacion".
VALUE LABELS /MANIFA 0 "No asiste a una manifestación" 1 "Si asiste a una manifestación".
RECODE ESTUDIOS (1 =4) (2 = 4) (3 = 3) (4 = 3) (5=2) (6=1) (9 = SYSMIS) INTO ESTU2.
VALUE LABELS /ESTU2 1 "superiores" 2 "medios" 3 "secundarios" 4 "Primarios".
RECODE P22 (1 =1) (2 = 1) (3 = 1) (4 = 1) (5 = 2) (6 = 2) (7 =3) (8 = 3) (9 = 3) (10 = 3) (98,99 =
SYSMIS) INTO IZ_DR.
VALUE LABELS /IZ_DR 1 "Izquierda" 2 "Centro" 3 "Derecha".
MISSING VALUES P901 P902 (7,8,9).
NUMERIC HABLA_POL.
VARIABLE LABEL HABLA_POL "Hablar de politica".
COMPUTE HABLA_POL = P901+P902.
LOGISTIC REGRESSION MANIFA WITH  P25 P26 ESTU2 IZ_DR HABLA_POL
```

```
/CATEGORICAL = P25 ESTU2 IZ_DR HABLA_POL
/CRITERIA = CUT(0.5) ITERATE(20)
/PRINT = CI(95)
/NOORIGIN.
```

Copiados en una ventana nueva y una vez cargado el archivo de datos del cuestionario del barómetro 2941 puede ejecutarse directamente todo el proceso descrito anteriormente desde el primer paso: