



UNIVERSITAT DE
BARCELONA



Manual de SPSS

De un estudiante para los estudiantes

Capítulo 1, 2 y 3

Análisis descriptivo univariado

Operaciones básicas con matrices

Operaciones básicas con variables

Autores/as Daniel Bernal
Graduado en Sociología
Máster en Transformaciones
Sociales e Innovación.
Becario del Observatorio del
Estudiante de la Universidad de
Barcelona.

Edición Observatorio del Estudiante
Vicerrectorado de Estudiantes y
Política Lingüística

Fecha edición Enero 2017

Aquesta obra està subjecta a la llicència de:
Reconeixement-NoComercial-CompartirIgual



<http://creativecommons.org/licenses/by-nc-sa/3.0/es/deed.ca>



Índice.

0. Elementos previos al análisis	5
1. Análisis descriptivo univariado.....	6
1.1. Naturaleza de las variables y estadísticos	8
1.2. Gráficos.....	16
2. Operaciones básicas con matrices.....	17
2.1. Segmentar archivos	18
2.2. Seleccionar casos.....	18
3. Operaciones básicas con variables.....	21
3.1. Recodificar variables.....	21
3.2. Construcción de índices.....	27
3.3. Construcción de tipologías.....	34

Nota

Este manual pretende ser un recurso más brindado por el Observatorio del Estudiante de la Universidad de Barcelona. Ha sido pensado para el propio autoaprendizaje de los estudiantes o como soporte a las posibles clases que estén cursando o que cursaran a lo largo del Grado. Por ello, todas las explicaciones están estructuradas “paso a paso” tanto a nivel de interpretación, como de las instrucciones a seguir dentro de las diferentes interfaces que ofrece el programa SPSS. Los diversos capítulos del manual no recogen todas las posibilidades que ofrece SPSS, pero ofrecen los recursos necesarios para un uso en profundidad del programa. Además, en la medida de lo posible, se ha pretendido huir de un lenguaje demasiado técnico que dificulte el aprendizaje operativo de los estudiantes. Se recomienda realizar un seguimiento capítulo a capítulo, ya que existen ciertos aspectos que se dan por sabidos (o ya explicados en capítulos anteriores) a medida que se avanza a lo largo del manual.

0. Elementos previos al análisis.

Antes de analizar cualquier base de datos mediante las herramientas que ofrece el SPSS, es necesario tener en cuenta algunos elementos previos.

Vista de datos vista de variables.

La interface del SPSS se compone fundamentalmente de dos pestañas clave: *vista de datos* y *vista de variables*. La **vista de datos** contiene la matriz con cada uno de los casos (individuos, territorios, organizaciones, etc.) y sus respectivos valores (o categorías de respuesta) en las distintas variables. La **vista de variables** incluye la información relevante de cada una de las variables que han sido obtenidas. Los aspectos centrales de la misma son:

- El *nombre* por el cual se identifica la variable
- *Tipo* (pueden existir numéricas o de cadena). Las numéricas son aquellas que contienen valores numéricos, ya sean las variables cualitativas nominales, cualitativas ordinales o cuantitativas. Se trata de una variable de cadena cuando se escribe un texto (por ejemplo, que explique algunas de las respuestas que han sido seleccionadas en alguna pregunta del cuestionario).
- *Decimales*, se señalan cuantos si se da el caso que alguna variable los contenga. Un ejemplo podría ser el porcentaje de inmigrantes en diferentes territorios.
- *Etiqueta*, se puede incluir una descripción más detallada del tema de la variable.
- *Valores*, se incluyen las categorías de respuesta de las distintas variables o valores. No se incluyen las categorías de respuesta únicamente cuando se tratan de variables cuantitativas como por ejemplo podría ser la edad o el porcentaje de inmigrantes en territorios determinados.
- *Perdidos*, esta es una de las opciones más importantes de la vista de variables ya que su no correcta utilización puede condicionar los resultados obtenidos en los diferentes estadísticos. Es importante que antes de cualquier análisis se compruebe que los valores perdidos de una determinada variable hayan sido declarados como tal ya que pueden estar rompiendo la ordinalidad (más adelante será explicado). Las categorías de respuesta que se suelen declarar como valores perdidos son (aproximativo): No sabe, No contesta. Pueden existir otros muchos casos.

Datos muestrales y datos poblacionales.

Un elemento del que se debe ser consciente antes de realizar cualquier análisis es si los datos con los que se trabajan son de una población completa o se tratan de datos muestrales. Los datos muestrales se obtienen mediante técnicas estadísticas que permiten que un conjunto delimitado de individuos pueda ser representativo de una población completa con un margen de error determinado. En este caso, los diferentes test de hipótesis lo que permiten es contrastar si las diferencias que se observan para una muestra concreta pueden extrapolarse de forma representativa para todo el conjunto de la población que representa.

Las diversas tablas que serán utilizadas como ejemplos han sido extraídas de la EES (European Social Survey) para la submuestra del Estado español y la oleada de 2014¹. Por lo tanto, se trabajara con datos muestrales representativos de la población española.

¹ Se recomienda descargar esta base de datos para seguir y reproducir los diversos ejemplos que se irán exponiendo. Recuerda, la mejor forma de aprender es practicar. Disponible en: <http://www.europeansocialsurvey.org/data/country.html?c=spain>

1 Capítulo 1: Análisis descriptivo univariado.

El *Análisis descriptivo univariado* es la primera aproximación –y más básica- que se debe realizar para conocer el comportamiento de un conjunto de variables. Como su propio nombre indica permite describir el comportamiento de las variables (individualmente), es decir, como se distribuyen a lo largo de las diferentes categorías o valores que las componen.

Tabla 1. Member of a group discriminated against in this country

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Yes	127	6.6	6.6	6.6
	No	1791	93.0	93.4	100.0
	Total	1918	99.6	100.0	
Perdidos	Don't know	7	.4		
Total		1925	100.0		

El *output* más básico que ofrece cualquier Análisis Descriptivo Univariado es la *tabla de frecuencias*. Una distribución de frecuencias informa sobre los valores concretos que adopta una variable y sobre el número (y porcentaje) de veces que se repite cada uno de esos valores. La tabla 1 o tabla de frecuencias, por tanto, describe el comportamiento (distribución) de los individuos en relación a los valores o categorías de respuesta de una variable determinada. Concretamente esta variable está describiendo si los individuos pertenecen a un grupo discriminado en ese país (primera categoría de respuesta) o no (segunda categoría de respuesta). La *frecuencia* ofrece la información acerca de cuantas personas, en términos absolutos, se sitúan en cada categoría respecto del total. Los *porcentajes* describen la distribución porcentual de los individuos en cada categoría. El *porcentaje válido* describe la distribución porcentual de los individuos (igual que la anterior) pero recalcula los porcentajes, restando los casos que están dentro

de categorías que se han declarado como perdidas². El **porcentaje acumulado** ofrece información acerca del sumatorio de los porcentajes de las distintas categorías.

Si seguimos este esquema, podemos afirmar que 1791 personas de un total de 1925 (**frecuencia**) dicen no formar parte de un grupo discriminado. Esto representa que el 93% de los encuestados (**porcentaje**) no pertenece a un grupo discriminado. No obstante, cuando no se tiene en cuenta a aquellos que no saben si pertenecen o no a un grupo discriminado -ya que han sido declarados como casos perdidos (**porcentaje valido**)-, el colectivo de los que no pertenecen a grupos discriminados asciende hasta un 93,4%.

El comando dentro del SPSS para obtener esta tabla es el siguiente: **Analizar > Estadísticos descriptivos >> Frecuencias >> Trasladar a la lista Variables**, mediante el botón flecha, la variable o variables cuya distribución de frecuencias se desea obtener >> Aceptar

El procedimiento de **Frecuencias** permite, como hemos visto, obtener distribuciones de frecuencias en las diferentes categorías de respuesta de una variable, pero además contiene opciones para:

- Calcular algunos de los estadísticos descriptivos más utilizados (sobre tendencia central, posición, dispersión, asimetría, etc.).
- Construir algunos diagramas básicos (gráficos de barras, de sectores e histogramas).
- Controlar el formato de presentación de las distribuciones de frecuencias.

1.1 Naturaleza de las variables y estadísticos.

Los estadísticos que se han de seleccionar en cada caso dependerán de la propia naturaleza de la variable.

Naturaleza de las variables

A grandes rasgos se debe considerar que existen tres tipos de variables en relación a su naturaleza: cualitativas, cualitativas ordinales y cuantitativas (de intervalo o de razón).

² Hay múltiples razones para declarar que los casos de una categoría concreta se consideren como perdidos. La más habitual es aquellos que pertenecen a una categoría que recoge la opción: no sabe, no contesta..

- **Cualitativas:** son aquellas compuestas por diferentes categorías de respuesta en las que se distribuyen las distintas frecuencias, sin que estas sigan algún tipo de ordinalidad (de mayor a menor o de menor a mayor). Las categorías son valores diferentes por una cualidad, no por una cantidad. Ningún “valor” puede ser considerado mayor o menor que otro. La variable que ha sido escogida en el anterior apartado sería un ejemplo de una variable de esta naturaleza. Otros ejemplos serían: partido político al que votó; región en la que vive; sexo; estado civil; marca de coche que conduce.
- **Cualitativas ordinales:** se componen de categorías de respuesta que se ordenan o pueden ordenarse de mayor a menor o de menor a mayor. Un ejemplo de este tipo de variable sería: 1). Del 1 al 10 cuán feliz se siente (siendo 1 nada feliz y 10 máxima felicidad) o 2). ¿Qué clase social considera que es? (baja, media-baja, media, media-alta, alta). Por lo tanto, sus valores son “categorías” como en las variables cualitativas, pero cada valor es “mayor que” o “menor que” los demás valores, como en las variables cuantitativas.
- **Cuantitativas (de intervalo o de razón):** los valores de la variable son “números” = cada valor posible es menor o mayor que otro valor. El conjunto de valores forman una escala de intervalo. “Intervalo”: distancia entre valores. En las variables cuantitativas se puede calcular la distancia o intervalo entre cualquier par de valores de la variable. La diferencia fundamental con las variables cualitativas ordinales es, por tanto, que en este caso se puede medir la distancia entre los valores. Ejemplos: edad, ingresos, nota en un examen, número de años de educación, kilómetros de distancia entre trabajo y residencia.

Lo relevante de entender las diferencias entre la naturaleza de las variables y por tanto poder identificarlas, radica en que en función de estas diferencias se deben desarrollar **métodos estadísticos diferenciados**. Concretamente existe una progresión hacia estadísticos más complejos (completos), dependiendo de si se tratan de variables cualitativas o cuantitativas.

El comando dentro del SPSS para seleccionar los diferentes estadísticos descriptivos son los siguientes: Analizar > **Estadísticos descriptivos** > **Frecuencias** > Estadísticos > seleccionas aquellos que te interesan > continuar > aceptar (una vez ya has seleccionado la variable o el conjunto de ellas).

Variables cualitativas.

Las variables cualitativas son aquellas que admiten una menor cantidad de estadísticos dada su naturaleza nominal (no numérica ni ordinal). En este tipo de variables únicamente tiene sentido analizar *la moda (medida de tendencia central)*. La moda nos indica cual es el valor con mayor frecuencia, es decir, aquella categoría de respuesta que más se repite.

Tabla 2. Main activity, last 7 days.

		Frecuencia	Porcentaje	Porcentaje válido	acumulado
Válido	1 Paid work	877	45.6	45.6	45.6
	2 Education	195	10.1	10.1	55.8
	3 Unemployed, looking for job	167	8.7	8.7	64.5
	4 Unemployed, not looking for job	62	3.2	3.2	67.7
	5 Permanently sick or disabled	51	2.6	2.7	70.3
	6 Retired	326	16.9	17.0	87.3
	8 Housework, looking after children, others	226	11.7	11.8	99.1
	9 Other	18	.9	.9	100.0
	Total	1922	99.8	100.0	
Perdidos	99 No answer	3	.2		
Total		1925	100.0		

Tabla 3. Estadísticos.

N	Válido	1922
	Perdidos	3
Moda		1

La tabla 2 es *la tabla de frecuencias* (tabla2) de la variable seleccionada, en este caso, señala la *actividad principal de los encuestados en los últimos siete días*. Se trata de una *variable cualitativa nominal* en tanto que está compuesta por categorías de respuesta nominales que no pueden ser ordenadas de mayor a menor (que tu actividad principal en los últimos siete días sea el trabajo remunerado no puede ser situado por encima de que haya sido la educación, el desempleo u otras). Concretamente esta tabla permite afirmar que el trabajo remunerado es la actividad principal de los encuestados (representativos de la totalidad del Estado español dado que han sido seleccionados mediante un muestreo), ya que es la categoría con mayor frecuencia y por tanto con mayor representatividad porcentual (45,6%)³. Dado que es la categoría de respuesta con mayor frecuencia y porcentaje, la moda (categoría más seleccionada) de esta variable es esta misma categoría (categoría 1) como bien refleja *la tabla 3*.

Consejos

A la hora de interpretar (describir) una variable cualitativa no se debe caer en el error de describir uno a uno los porcentajes de todas las categorías. Se puede empezar por señalar si: 1). Hay una, dos o tres categorías que destacan porcentualmente. 2). Todas las categorías tienen porcentajes similares. 3). E incluso puedes sumar categorías a la hora de describirlas (Ej: Las actividades principales de un 55,4% de los españoles en los

Variables cualitativas ordinales.

Dado que las variables *cualitativas ordinales* poseen la cualidad de poder ser *ordenadas de menor a mayor o viceversa* –a pesar de que continúan siendo categorías nominales-, en éstas aumenta el número de estadísticos descriptivos que pueden ser utilizados para su descripción. Es el caso de **los percentiles (o cuartiles) y la mediana** que se añaden a la moda.

Cuartiles: Calcula los percentiles 25, 50 y 75, es decir, los valores por debajo de los cuales se encuentra el 25 %, el 50 % y el 75 % de los casos, respectivamente.

³ Recordemos que siempre es recomendable analizar la columna de porcentaje válido ya que está excluyendo los casos que hemos declarado como perdidos. Si alguna variable contiene un % muy relevante de casos declarados como perdidos se debe dejar constancia a la hora de describir el comportamiento de esa variable.

Percentiles: Permite solicitar percentiles concretos (valores que acumulan un determinado porcentaje de casos). Permite visualizar en que categorías (plural) de respuesta se acumulan los primeros 10%, 30%, 70%, 90% (cualquiera que selecciones) de los casos.

Mediana (tendencias central): Valor por debajo del cual se encuentran el 50 % de los casos (equivale al percentil 50).

Tabla 4. Subjective general health.

	Frecuencia	Porcentaje	Porcentaje válido	acumulado
Válido 1 Very good	361	18.8	18.8	18.8
2 Good	837	43.5	43.5	62.2
3 Fair	515	26.8	26.8	89.0
4 Bad	186	9.7	9.7	98.6
5 Very bad	26	1.4	1.4	100.0
Total	1925	100.0	100.0	

Tabla 5. Estadísticos.

N	Válido	1925
	Perdidos	0
Mediana		2.00
Moda		2
Percentiles	10	1.00
	20	2.00
	25	2.00
	30	2.00
	40	2.00
	50	2.00
	60	2.00

70	3.00
75	3.00
80	3.00
90	4.00

La variable que ha sido seleccionada para ejemplificar el análisis de los estadísticos descriptivos propios de las variables cualitativas ordinales ha sido la percepción subjetiva de la salud (tabla 4). Como se puede ver se trata de una variable cualitativa ordinal, en tanto que, se compone de categorías nominales (Muy buena, buena, justa, mala, muy mala) que pueden ser ordenadas (Muy buena, es mayor que, buena, y así sucesivamente) a pesar de que no se puede cuantificar la diferencia. Como muestra la *tabla 4* el 62,3% de los españoles se percibe con una salud apropiada (buena o muy buena), mientras que solo el 11.1 % dicen tener una salud insuficiente (mala o muy mala). La moda (tabla 4) indica que la categoría que predomina por encima de las demás categorías de respuesta es la de buena salud (43,5%) y la mediana (igual a 2), señala que el 50% de los casos se encuentran entre las categorías de respuestas 1 y 2 (50% o más, ya que la mediana solo indica en que categoría o valor se desborda el 50% de los casos). El análisis de los percentiles, permite afirmar de forma rápida que al menos el 80% de los casos no perciben tener una salud mala (Muy buena, buena o justa).

Es importante destacar que normalmente a las *variables cualitativas ordinales de más de 5 o más categorías* se les suelen tratar como variables cuantitativas en términos estadísticos, es decir, que si analizas una variable cualitativa ordinal de 5 o más categorías puedes analizarla utilizando también los estadísticos de las variables cuantitativas.

Variables cuantitativas.

El análisis descriptivo de las variables cuantitativas permite la utilización del mayor número de estadísticos descriptivos posibles. Además de poder utilizar aquellos propios

de las variables cualitativas y cualitativas nominales, también puede ser utilizada la **media**, dentro de los *estadísticos de tendencia central*, **la desviación típica y la varianza** dentro de los *estadísticos de dispersión*, así como **la asimetría y la curtosis**, dentro de los *estadísticos de distribución*.

Media o media aritmética: es la suma de todas las puntuaciones dividida por el número de puntuaciones. Te permite conocer el valor promedio de la variable.

Varianza: medida de dispersión que se obtiene dividiendo por $n-1$ la suma de los cuadrados de las diferencias entre cada puntuación y la media. El valor que se obtiene es difícilmente interpretable, es decir, no se interpreta. A pesar de ello siempre se pide este estadístico ya que la desviación típica (que sí que se interpreta) se deriva de la varianza.

Desviación típica (desviación estándar en SPSS): raíz cuadrada de la varianza. Mide el grado en que las puntuaciones de la variable se alejan de su media.

Mínimo: valor mínimo de la variable

Máximo: valor máximo de la variable.

Asimetría: es un indicador que pretende expresar si la distribución de la variable se comporta respetando la curva de normalidad. Si el valor de este indicador es 0, significa que la distribución de la variable se comporta siguiendo una distribución “normal”.

- *Asimetría a la derecha*: si el valor es más grande que 0, implica que la distribución de los casos se concentran en las primeras categorías de respuesta o valores.
- *Asimetría a la izquierda*: si el valor es menor que 0, implica que las personas están concentradas en los últimos valores o categorías de respuesta de la pregunta.

Curtosis: es un indicador que pretende expresar de forma numérica si los individuos se concentran en los valores centrales o si por el contrario están igualmente distribuidos a lo largo de todas las categorías de respuesta.

- *Mesocurtica*: cuando el valor del indicador es igual a 0,2663 se puede afirmar que se trata de una variable que se distribuye de forma mesocurtica, es decir, que respeta la distribución normal.

- *Platicurtica*: cuando el valor del indicador es menor que 0,2663 se puede afirmar que la variable se distribuye de forma mesocrática, es decir, hay aproximadamente los mismos casos en cada uno de los valores de la variable.
- *Leptocurtica*: si el valor del indicador es mayor que 0,2663 se puede afirmar que la variable se distribuye de forma leptocurtica, es decir, los casos se concentran en los valores centrales de la variable. La distribución de la curva, por tanto, será más puntiaguda.

La variable seleccionada para ejemplificar el análisis de los estadísticos de las variables cuantitativas, ha sido “peso de los encuestados”. La tabla de frecuencias no ha sido insertada en este documento ya que ocupaba aproximadamente cuatro páginas y aporta poca información para una interpretación apropiada del comportamiento general de la variable, debido a la gran variabilidad de los diferentes pesos de los encuestados. En su lugar este comportamiento puede ser sintetizado mediante los diferentes estadísticos.

En un primer momento, se puede destacar que el peso más frecuente entre los entrevistados es el de 70 kg (moda), así como que un 50% o más de los encuestados pesan entre 0 y 71 kg (Mediana). Más concretamente ese 50% (o más) se sitúa entre los 34 kg (Mínimo) y los 71 Kg. El peso promedio de los españoles (recordemos que estamos analizando la encuesta ESS en la oleada de 2014 para el Estado español) es de 72,42 Kilos (Media). No obstante, este peso promedio oscila entre 57,2 y 87,32 (Desviación estándar: he sumado y restado el valor de este indicador al valor de la media). Por otro lado, se vislumbra que el rango de pesos posibles oscila entre los 34 kg y los 140 kg (Mínimo y máximo).

Tabla 6. Estadísticos.

N	Válido	1875
	Perdidos	50
Media		72.423
Mediana		71.000
Moda		70.0
Desviación estándar		14.8803
Varianza		221.422
Asimetría		.589
Error estándar de asimetría		.057

Curtosis		.556
Error estándar de curtosis		.113
Mínimo		34.0
Máximo		140.0
Percentiles	25	61.000
	50	71.000
	75	82.000

1.2 Gráficos.

El procedimiento Frecuencias también ofrece la posibilidad de obtener algunos gráficos básicos que permiten analizar de forma visual la distribución de las variables. En concreto, permite obtener gráficos de *barras*, *sectores* e *histogramas*.

Analizar >>estadísticos descriptivos >> frecuencias >>Gráficos

Gráfico de barras: gráficos en los que a cada valor de la variable se le asigna una barra con una altura equivalente a su frecuencia absoluta o porcentual. La escala de la altura de las barras se ajusta automáticamente teniendo en cuenta la frecuencia más alta de las representadas.

Gráfico 1. Gráfico de barras

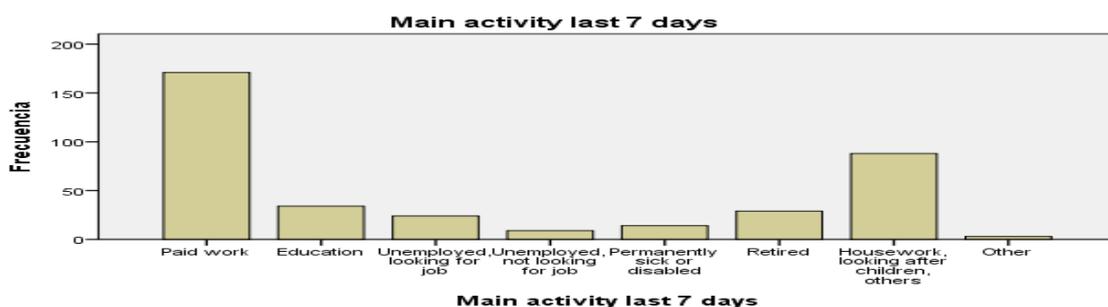
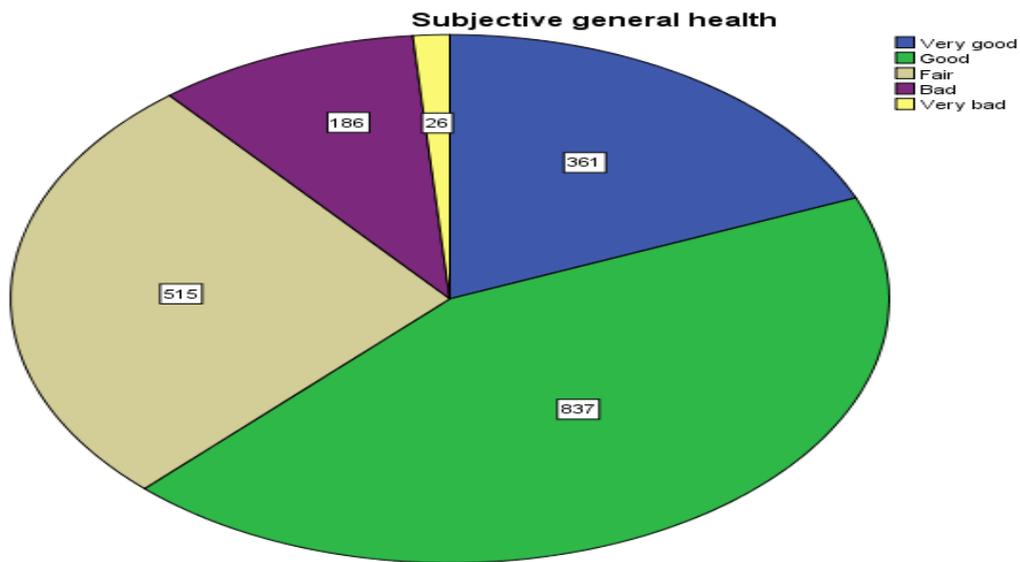


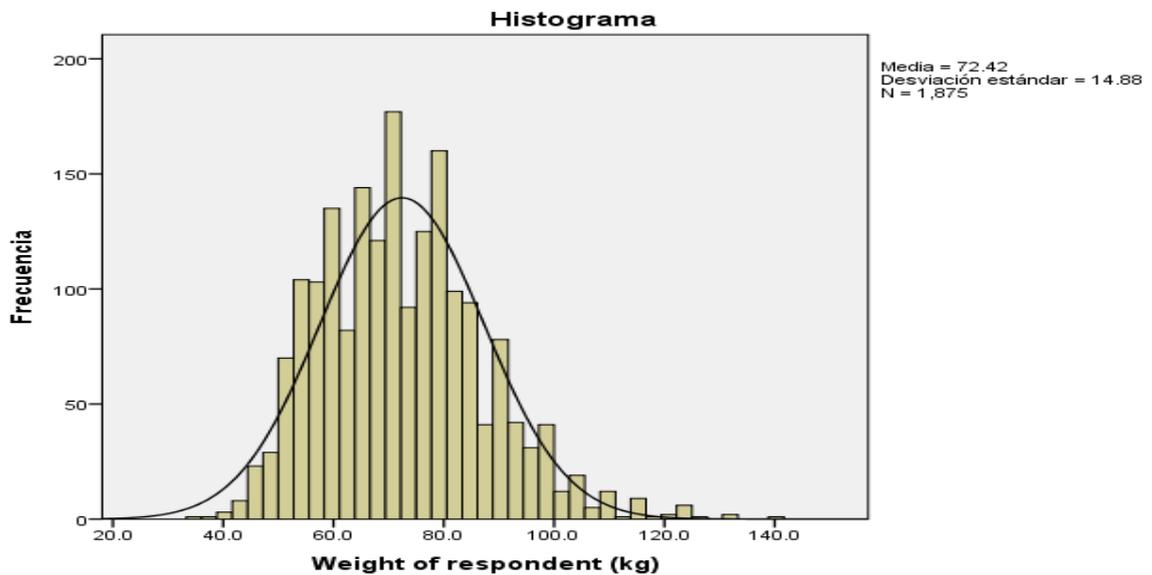
Gráfico de sectores: gráficos circulares en los que a cada valor de la variable se le asigna un sector de tamaño equivalente a su frecuencia absoluta o porcentual.

Gráfico 2. Gráfico de sectores



Histograma: similares a los gráficos de barras pero con las barras juntas, dando así la impresión de continuidad. Sólo pueden obtenerse con variables numéricas. Para construir el histograma el SPSS agrupa la variable en 21 intervalos (o menos, si la amplitud de la variable es menor que 21). La línea negra del gráfico representa como se debería comportar la distribución de la variable para respetar la distribución normal.

Gráfico 3. Histograma



2 Capítulo 2: Operaciones básicas con la matriz.

A la hora de analizar los datos un investigador puede estar interesado en describir cómo se comportan diferentes variables en diferentes grupos de individuos o en ver como esas variables se distribuyen únicamente en un conjunto determinado de casos que se definen por tener en común un número de características. Las diferentes operaciones básicas que el programa SPSS permite realizar hacen que estos objetivos sean posibles y que por tanto se pueda profundizar en la descripción uní-variada de los estadísticos.

2.1 Segmentar archivos.

La opción de segmentar archivos posibilita obtener cada uno de los resultados del análisis para cada una de las categorías de respuesta de la variable seleccionada. Debido a esto último es recomendable escoger variables cualitativas o cualitativas ordinales (de no muchas categorías) como variable segmentadora.

Para segmentar tu archivo debes seguir estas instrucciones: Datos >> Segmentar archivo >> Organizar los resultados por grupos >> Seleccionas la variable que segmentará tu archivo >> Aceptar.

Una vez se ha segmentado el archivo se debe proceder a realizar el análisis. Como el archivo ya ha sido segmentado obtendrás una tabla de frecuencia y de estadísticos para cada una de las categorías de respuesta de la variable seleccionada. Por tanto, *esta opción permite tener un análisis descriptivo detallado (los diferentes estadísticos asociados a su naturaleza) del comportamiento de una determinada variable para cada una de las categorías de respuesta de la variable que está segmentando tu archivo.*

Una vez que se ha finalizado el análisis se debe quitar la opción de segmentar archivo para que tu matriz vuelva a operar sus análisis con todo el conjunto de los individuos. Para ello debes seguir la siguiente instrucción: Datos >> Segmentar archivo >> Restablecer.

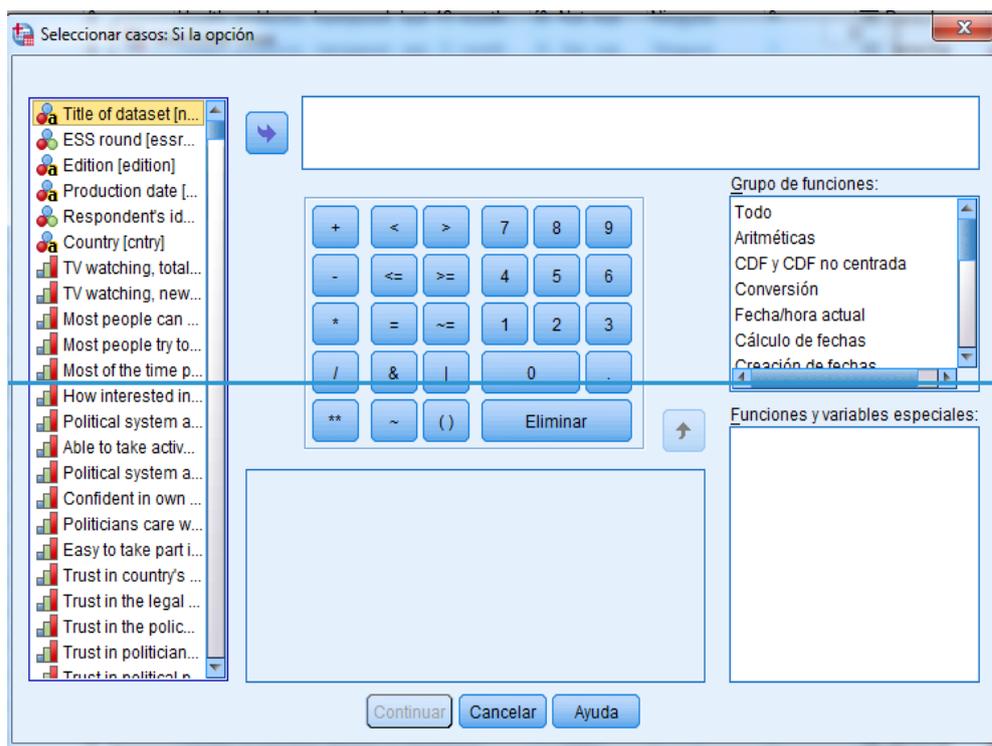
Al segmentar los archivos, los encuestados se distribuyen a lo largo de las distintas categorías de respuesta de la variable que están segmentando.

2.2 Seleccionar casos.

Tal vez, en un análisis se necesita lograr una mayor profundidad y se requiere describir como esas variables se comportan únicamente en un conjunto determinado de casos que se definen por tener en común un número de características. Por ejemplo, se quiere saber cómo se comporta la variable de salud subjetiva únicamente en los individuos mayores de 50 y de género femenino.

Para seleccionar un conjunto determinado de casos se deben seguir estas instrucciones: Datos >> Seleccionar casos >> Si satisface la condición. En este momento aparece una pantalla con una interface similar a una calculadora (imagen 1). Dado que los casos que se quieren seleccionar deben cumplir las siguientes condiciones: *Edad* ≥ 50 & *Genero* = *Femenino*, lo primero que se debe hacer es identificar esas variables en el recuadro “lista de variables”, que aparece en la parte izquierda de la interface. Una vez identificadas se debe pedir mediante la calculadora, que el programa te seleccione los casos que cumplen esa condición. ¿Cómo? 1). Se selecciona la variable y se clica a la flecha que está al lado del recuadro blanco para que se introduzca. 2). Una vez introducida se debe especificar que esa variable cumpla la condición que ha sido decidida. Por ejemplo = *Edad* ≥ 50 / *Genero* = *Femenino* (tienes que buscar con que número esta codificado el género femenino). En nuestro caso quedaría de la siguiente forma *age* >= 50 & *gndr* = 2. Dado que se están seleccionando dos condiciones simultaneas, es decir, personas que cumplen una determinada edad y un determinado género, el conector que debe unir estas dos condiciones es &. Una vez seleccionadas las condiciones que deben cumplir los casos damos clic en Continuar y Aceptar. El filtro ha sido activado. A partir de ahora y hasta que desactives el filtro, todos los análisis que se realicen serán para los casos que cumplan las condiciones anteriormente señaladas.

Imagen 1. Interface de selección condicional de casos



En el ejemplo únicamente han sido utilizados algunos de los comandos útiles de la calculadora. Ha continuación se explican con más detalle las diferentes operaciones y comandos que pueden ser útiles a la hora de seleccionar un grupo de casos.

1. Todas las instrucciones siempre deben ser introducidas mediante el teclado, ya que si son escritas pueden dar algún error.
2. A la hora de seleccionar casos la selección puede requerir características conjuntas de diferentes variables o la selección de casos que cumplan una u otra condición (*no necesariamente ambas*). Cuando se quiere seleccionar que los casos cumplan más de una característica (género femenino y edad mayor de 50), siempre se deben enlazar las dos variables con el siguiente símbolo **&**. Sin embargo, cuando se quieren seleccionar los casos que cumplen una u otra condición (*no necesariamente ambas*: o que sean mujeres o que sean mayores de 50) se debe utilizar el siguiente símbolo **|**.
3. Los diversos símbolos matemáticos siguen la lógica de significados convencional.
4. Cabe destacar la función Any. *ANY debe ser buscado en el recuadro de Funciones y variables especiales habiendo seleccionado, previamente, la opción*

de búsqueda en el recuadro del grupo de funciones. Fundamentalmente, ANY es una función que permite seleccionar más de una característica (categoría de respuesta o valor) de una variable determinada. Cuando se selecciona esta opción, aparece en la pantalla principal la siguiente información: ANY (?). El primer interrogante debe ser substituido por las variables cuyas categorías o valores se quieren filtrar los casos. El segundo interrogante debe ser substituido por las categorías de respuesta en las que se está interesado. **Si interesa más de una categoría**, añades una coma más y el número. **Por ejemplo**, si únicamente se quieren seleccionar los casos que tienen 30, 33 o 35 años, el resultado final que aparecería en el recuadro sería el siguiente Any (agea, 30, 33, 35). Primero el nombre de la variable (seleccionada con la flecha, y después las edades en las que se está interesado).

5. También cabe destacar la función Range. Esta función se encuentra ubicada en el mismo espacio que Any. Su utilidad es muy similar a la de Any (seleccionar un conjunto determinado de categorías de respuesta de una variable concreta). La diferencia fundamental es que en vez de seleccionar una única categoría de respuesta, seleccionas un conjunto de ellas (muy útil en variables cuantitativas) y más concretamente un rango de ellas. Por ejemplo, si se quiere seleccionar a los individuos que tienen una edad entre 25 y 40 años, el resultado final que aparecerá en el recuadro será: Range (agea, 25, 40). De esta forma el comando te selecciona cualquier edad comprendida entre el rango de 25 y 40 años (25 y 40 años, incluidos)

Recuadro 1: comandos calculadora

1. &=y (conjunción). *Para condiciones conjuntas.*
2. |= o. *Casos que cumples una u otra condición.*
3. Diversos símbolos matemáticos (\geq, \leq , etc). Siguen la lógica del significado común de esos símbolos.
4. Operaciones (+, -, *, /) * (multiplicación), / (División)
5. Any (Var, x, y, z, ...)
6. Range (Var, de x a y)

3 Capítulo 3: Operaciones básicas con variables.

Este capítulo se centra en las diversas operaciones básicas que se pueden realizar con una variable. Concretamente en la recodificación de variables, la construcción de índices y la elaboración de tipologías.

3.1 Recodificación de variables.

La función de recodificar variables puede ser útil para diferentes objetivos: 1). Agrupar categorías de respuesta de esa variable. 2). Cambiar el orden de la variable (de más a menos hacía de menos a más).

Agrupar categorías de respuesta.

Las razones que pueden llevar a querer agrupar las categorías de respuesta de una variable en un número menor de ellas son múltiples: 1). Pasar de una naturaleza determinada de una variable (cuantitativa) a otra diferente (cualitativa ordinal). Por ejemplo, recodificar la edad en 4 grandes grupos: 18-35, 36-55, 56-67, +68. 2). Que una de las categorías de respuesta tenga una frecuencia inferior al 5%. 3). Razones teóricas. Por ejemplo, asumir que los diferentes grupos de edad se corresponden a diferentes etapas de la vida laboral.

Para recodificar una variable se debe seguir las siguientes instrucciones:

Transformar >> recodificar en distintas variables (siempre se selecciona en diferentes variables, ya que si seleccionas la opción de recodificar en las mismas variables, estás perdiendo la variable anterior) >> Aparece la interface (imagen 2) de “recodificar variables” >> Seleccionas la variable que quieres recodificar en el recuadro de variables situado a la izquierda y la introduces al recuadro blanco mediante la flecha >> Una vez introducida, se debe indicar un nombre y una etiqueta en el recuadro de **variable de resultado**, y clicar cambiar >> clicas **valores antiguos y nuevos** >> Aparece la interface de **Valores antiguos y nuevos** (imagen 3). Si se observa la interface, se distinguen dos espacios bien definidos. Uno con el nombre de **valor antiguo** y otro con el de **valor nuevo**. El espacio de valor antiguo es para seleccionar los valores de la variable que deseas agrupar bajo una misma categoría de respuesta. El espacio de valor nuevo, es para seleccionar bajo que categorías de respuesta vas agrupar tales valores. Como se puede observar, hay diferentes formas de seleccionar los valores que quieres agrupar. [Continua]

Concretamente mediante la opción “*valor*” o “*rango*”. La opción valor es para seleccionar únicamente un único valor (categoría de respuesta de la variable), la opción rango, sirve para seleccionar un conjunto de valores o “**categorías de respuesta**” comprendidas entre el primer valor que seleccionas y el último. Una vez seleccionados los valores antiguos, debes atribuirles un nuevo valor o categoría de respuesta en la nueva variable, para ello se debe seleccionar ese nuevo valor o categoría de respuesta en el apartado de valor nuevo y clicar **añadir**. Una vez has hecho este proceso para todos los valores antiguos y nuevos, debes clicar “**todos los demás valores**” en el apartado de valores antiguos y clicar “**perdido del sistema**” en “valores nuevos” y clicas **añadir** >> **Completado este proceso, clicas en continuar (imagen 3) y Aceptar (imagen 2).**

Imagen 2. Interface “recodificar en distintas variables”

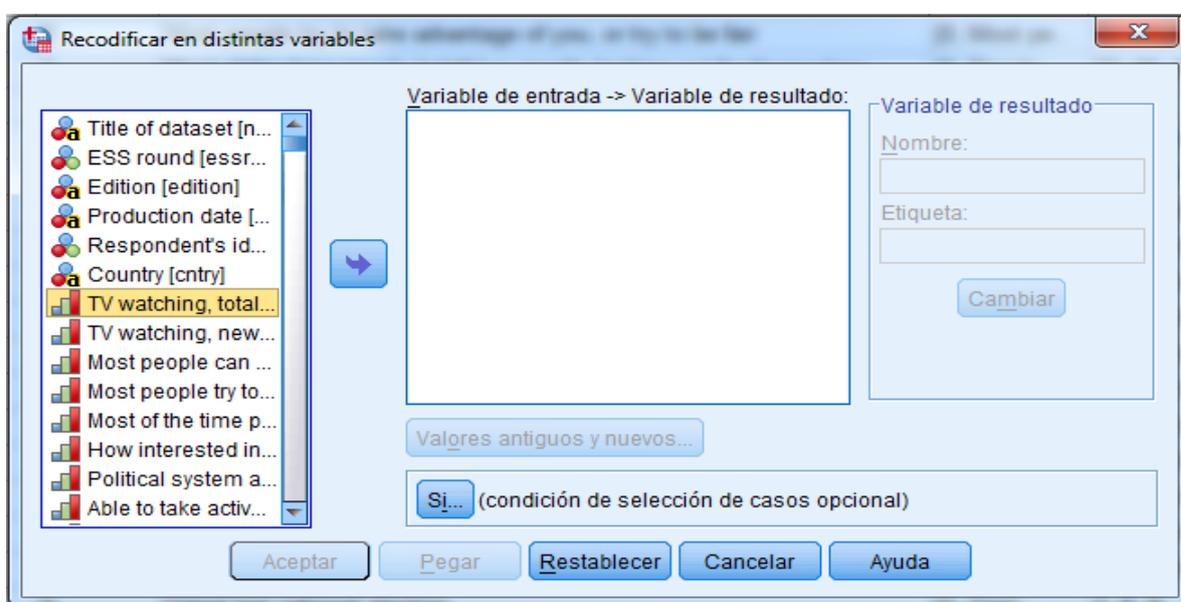
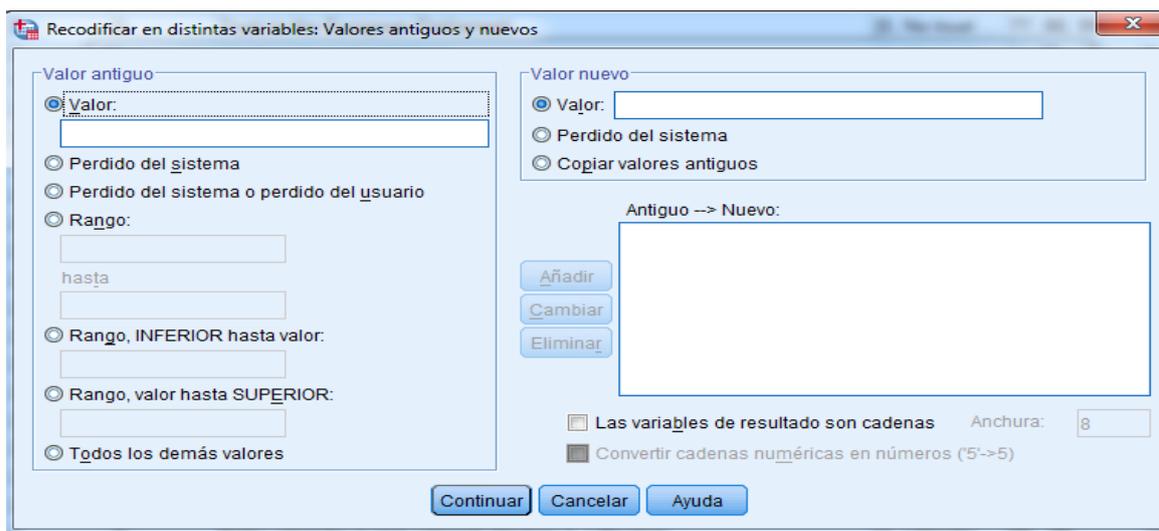


Imagen 3. Interface “valores antiguos y nuevos”

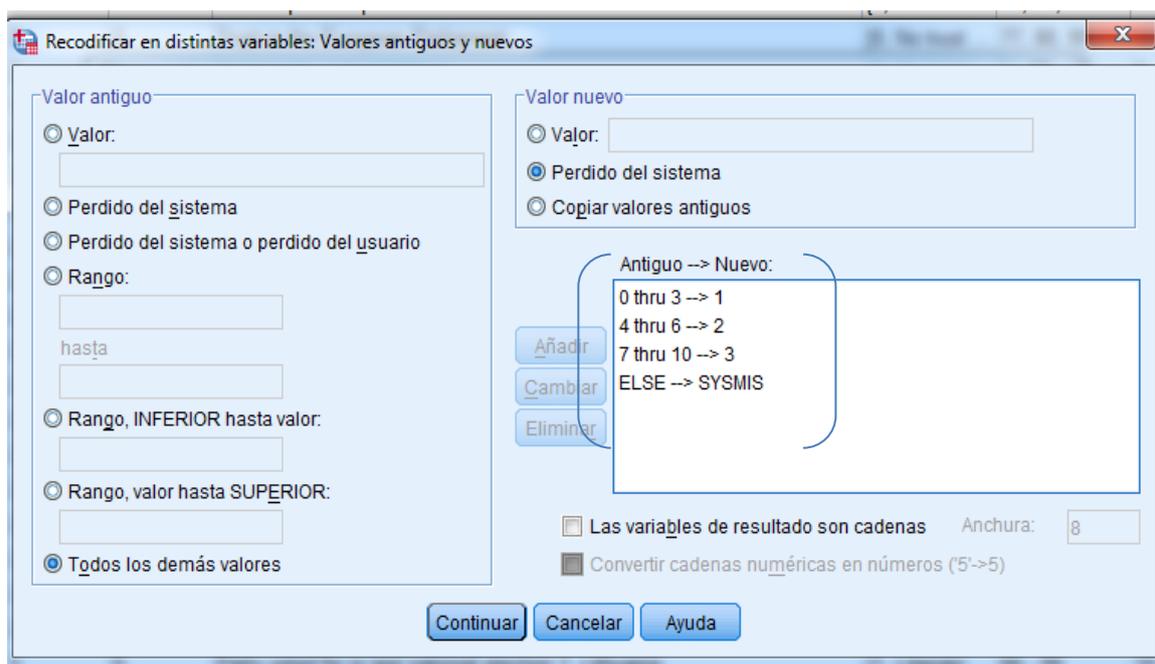


Una vez has completado el proceso del recuadro azul, ya tendrás la nueva variable recodificada que se situara al final del todo (ultima variable) en el apartado “*vista de variables*”. ¿Ya se ha finalizado el proceso? Casi. Una vez construida la variable se le deben poner las nuevas etiquetas.

Ejemplo:

La ESS (European Social Survey) recoge la variable de confianza en los políticos mediante una escala que va desde “0” (ninguna confianza) hasta 10 (completa confianza). Pero nos interesa resumir esa información que se distribuye a lo largo de 11 categorías de respuesta en únicamente 3 (alta confianza, media confianza y baja confianza). Este objetivo puede ser logrado, siguiendo los pasos que han sido explicados en el recuadro azul. En un primer momento se debe tomar la decisión de que categorías de respuesta o valores de la antigua variable serán agrupados dentro de cada una de las nuevas tres categorías de respuesta que compondrán la nueva variable. En este caso, se ha decidido respetar los parámetros de los valores de la antigua variable. Por tanto, “0, 1, 2, 3” pasaran a ser *baja confianza*, “4, 5, 6” *media* y “7, 8, 9, 10” *alta confianza*. Si se sigue el proceso que hemos señalado en el recuadro finalmente nos debe quedar la *imagen 3* completada con la siguiente información (*imagen 4*).

Imagen 4. Interface "valores antiguos y nuevos" completada



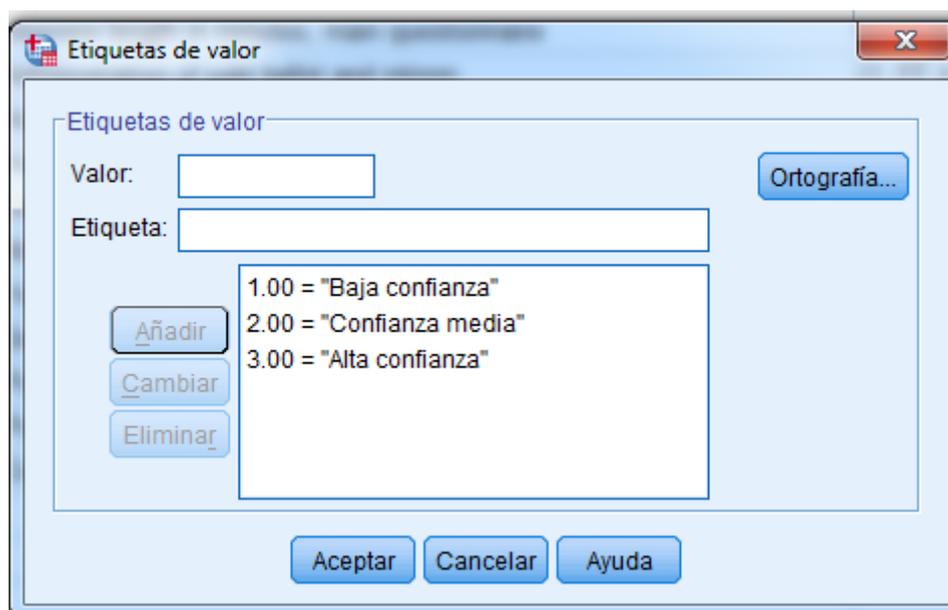
Una vez completada se debe clicar a continuar y aceptar. La variable habrá sido creada. Si se analiza la nueva variable sin que se le haya asignado una etiqueta a las nuevas categorías, el resultado sería el siguiente:

Tabla 7. Confianza en los políticos recodificada.

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1.00	1357	70.5	71.4	71.4
	2.00	473	24.6	24.9	96.3
	3.00	71	3.7	3.7	100.0
	Total	1901	98.8	100.0	
Perdidos	Sistema	24	1.2		
Total		1925	100.0		

Si lo que se quiere es que aparezcan las nuevas categorías de respuesta (alta confianza, confianza media y baja confianza). Se debe etiquetar esta nueva variable. Para ello, se debe ir a la interface de "vista de variables", identificar la nueva variable (recordar que siempre aparecerá al final), ir al apartado de valores y asignarle tales etiquetas al valor que le corresponde.

Imagen 5. Interface “etiquetas de valor”.



Una vez completado este proceso, si se vuelve a analizar la variable la tabla que aparecerá adoptará la siguiente forma.

Tabla 8. Confianza en los políticos recodificada.

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Baja confianza	1357	70.5	71.4	71.4
	Confianza media	473	24.6	24.9	96.3
	Alta confianza	71	3.7	3.7	100.0
	Total	1901	98.8	100.0	
Perdidos	Sistema	24	1.2		
Total		1925	100.0		

Cambiar el orden de la variable.

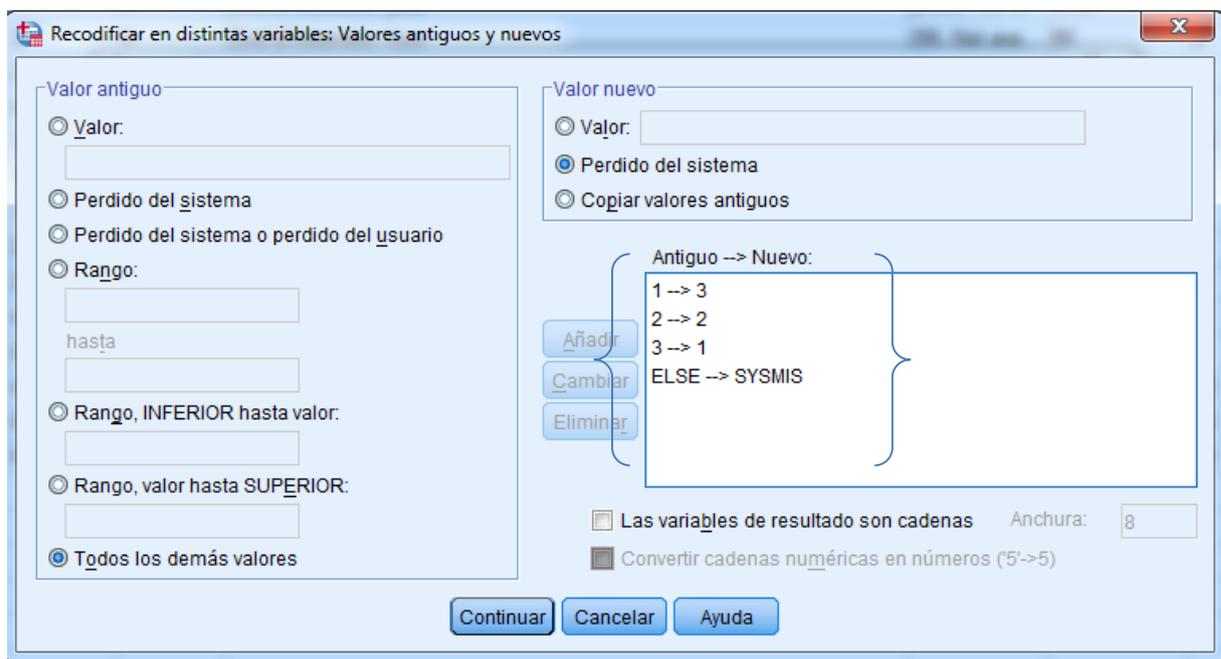
Otra de las opciones por las que puede interesar a un investigador recodificar una variable es cambiar el orden o sentido de la variable, es decir, si esta ordenada de menos a más, que pase a estar ordenada de más a menos. En el caso de nuestra variable ya recodificada, el orden, se establece de menos a más (de menos confianza a más confianza). Si interesa invertir el orden (de más confianza a menos) se debe proceder a recodificar la variable invirtiendo el orden de las categorías de respuesta, es decir, si el 1

es “baja confianza” en la nueva variable con el orden invertido, pasara a ser 3. Si el 3 es “alta confianza”, en la nueva variable de orden invertido pasara a ser 1. Y el 2 “confianza media” seguirá ocupando la posición intermedia de 2.

Ejemplo:

Si se sigue el caso expuesto en el párrafo anterior y se realiza de forma correcta se debería obtener una interface completada con los siguientes resultados (imagen6):

Imagen 6. Interface “recodificar en distintas variables”



De nuevo, se deberían etiquetar las categorías de respuesta de la nueva variable recordando que en esta nueva variable 1 “Alta confianza”, 2 “confianza media”, 3 “baja confianza”. Una forma de comprobar que se ha realizado de forma correcta, es ver si los porcentajes de alta confianza en la variable invertida y no invertida, se corresponden, a pesar de que no estén ubicadas en la misma posición de la escala ordinal. En este caso *es correcto*.

Tabla 9. Confianza en la política recodificada e invertida.

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Alta confianza	71	3.7	3.7	3.7
	Confianza media	473	24.6	24.9	28.6
	Baja confianza	1357	70.5	71.4	100.0
	Total	1901	98.8	100.0	
Perdidos	Sistema	24	1.2		
Total		1925	100.0		

3.2 Construcción de índices.

La mejor forma de responder a la pregunta ¿Qué es un índice? Es intentar dar respuesta, primeramente al interrogante ¿para qué sirven o como se constituyen los índices?

La realidad social es multidimensional y compleja. Para estudiarla una de las herramientas fundamentales son los conceptos. Los conceptos pretenden –de forma muy simplificada- sintetizar y describir la realidad y como ésta se comporta. Debido a su complejidad, en ciencias sociales, difícilmente se puede encontrar algún concepto que esté constituido por una única variable, sino más bien por conjuntos de ellas. Un ejemplo sería la *clase social*. Han quedado atrás las teorías que afirmaban que la clase social se definía únicamente en función de la relación que los individuos mantenían con los medios de producción (teoría de las clases sociales marxistas). Las nuevas formas de conceptualizar la clase social, han añadido a esta idea unidimensional aspectos como el conocimiento, el estatus, el poder... No obstante, a su vez el conocimiento, el estatus, etc., son elementos demasiado amplios para ser recogidos con una única y sola variable. Estos aspectos pueden ser entendidos como dimensiones del concepto de clase social, las cuales deben ser operacionalizadas mediante diferentes indicadores (variables). Por ejemplo, el conocimiento podría ser recogido mediante las variables del nivel de estudios alcanzado, el nivel educativo de los padres, el grado de preparación que requiere su puesto de trabajo, etc. Este ejemplo, *debe servir únicamente para reflejar que existen ciertos aspectos de la realidad social (la mayoría de ellos) que no pueden ser recogidos, caracterizados o definidos con una única variable.*

Una forma de agrupar variables para construir dimensiones o conceptos es mediante la creación de índices. Un índice se construye mediante la suma de variables cualitativas ordinales o cuantitativas.

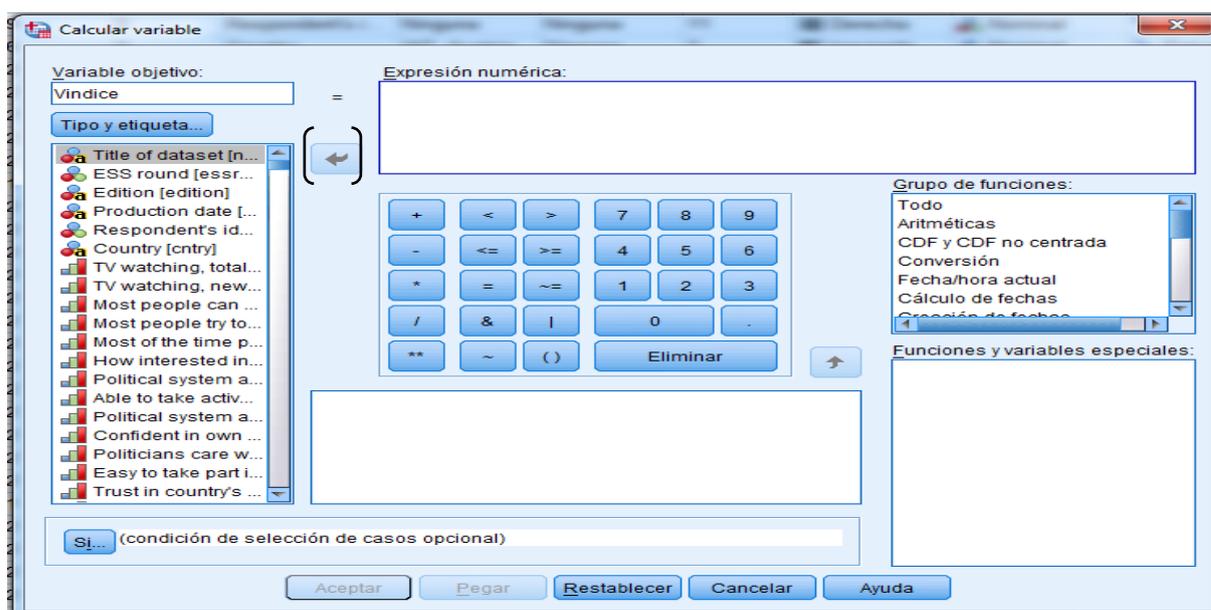
Para obtener un índice se deben seguir los siguientes pasos:

Transformar >> Calcular variable >> aparece una interface (imagen 7) muy similar a la de **la imagen 1**. Su funcionamiento es el mismo, por lo tanto para utilizar la calculadora para proceder en las diferentes operaciones que se han de realizar se recomienda repasar el **recuadro 1** (comandos calculadora) >> En el recuadro de *variable objetivo* se debe señalar el nombre de la nueva variable (índice) que será generada a través de la suma de 2 o más variables >> seleccionas la primera variable que quieres sumar en el recuadro que está situado en el apartado izquierdo de la pantalla y la desplazas con la flecha hacia el recuadro en blanco “expresión numérica” >> añades un símbolo de sumar (+) y buscas la siguiente variable que compondrá el índice que se está pretendiendo construir, para introducirla nuevamente en el apartado de expresión numérica (este proceso se ha de repetir hasta que se hayan introducido todas las variables que compondrán el índice) >> aceptar.

Elementos a tener en consideración para sumar variables.

1. Las variables que se van a sumar tienen que ser de la misma escala, es decir, no puedes sumar variables compuestas por 5 valores o categorías de respuesta y variables con 11 valores o categorías de respuesta. Para igualar las categorías de respuesta de una variable se puede proceder a recodificarlas y homogeneizar el tamaño de todas (siempre pasando las variables con más categorías al número de categorías de la variable que menos tiene de todas aquellas seleccionadas para construir el índice). Mirar el apartado de “agrupar categorías de respuesta de este capítulo”.
2. Las variables tienen que tener el mismo sentido. No puedes sumar variables cuya ordinalidad va de más a menos con variables que van de menos a más. Si este es tu caso, tienes que homogeneizar el sentido de tus variables. Mirar el apartado “cambiar el orden de la variable” de este capítulo.

Imagen 7. Interface “calcular variable”.



Ejemplo:

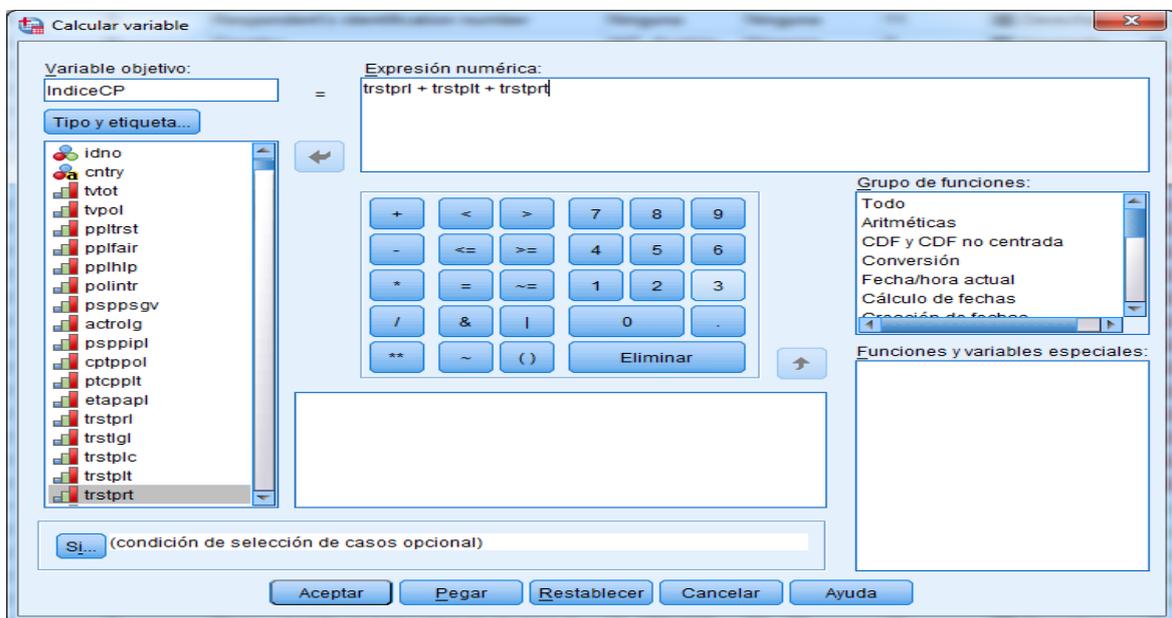
Si se propone estudiar la “*confianza en la política*” lo primero que se debe plantear un investigador es como se puede medir ese concepto. Es decir, que indicador (variable) puede reflejar la confianza que un individuo tiene en la política. Uno rápidamente puede pensar que un indicador válido para medir la “*confianza en la política*” es la confianza que los individuos dicen tener en los políticos, otro puede señalar que una mejor manera de medirlo es la confianza que se tiene en el parlamento. Tal vez otro señale que lo realmente relevante para medir la “*confianza en la política*” es la confianza que dicen manifestar los individuos en los partidos políticos. Lo cierto es que todos estos indicadores tienen la capacidad de medir diferentes aspectos representativos de la “*confianza en la política*” y que por tanto seleccionar uno solo de ellos, puede estar dejando escapar cierta información relevante. Es decir, que se esté escapando alguna dimensión que permita captar el concepto de “*confianza política*” en toda su complejidad.

¿No se podría medir la confianza política con todos esos indicadores simultáneamente y de esta forma recoger un mayor número de matices de este concepto multidimensional? Mediante la construcción de un índice se pueden agrupar todos esos indicadores en una única y nueva variable que logre capturar de forma más adecuada el concepto de “confianza en la política”. ¿Cómo? En un primer momento, se deben

identificar las variables con las que vas a construir el índice.⁴ En este caso, se ha decidido optar por la 1. Confianza en los políticos. 2. Confianza en el parlamento 3. Confianza en los partidos políticos. Lo primero es identificarlas en el SPSS (recordemos que estamos analizando la encuesta ESS en la oleada de 2014 para el Estado español). Una vez identificadas (*trstplt*= confianza políticos, *trstprl*= confianza en el parlamento, *trstprt*= confianza en los partidos políticos) se ha de comprobar que estas variables tengan el mismo rango y el mismo sentido. Para ello se va al apartado de valores dentro de la interface de “vista de variables”. Como se puede observar todas ellas tienen un rango de 11 categorías de respuesta (0-10) y además, todas están estructuradas en el mismo sentido 0 (nada de confianza) – 10 (plena confianza), es decir, de menos a más. Por lo tanto no se deben recodificar. Recuerda que si tus variables seleccionadas tienen rangos diferenciados o sentidos diferentes debes recodificarlas para homogeneizar su tamaño y su sentido.

Una vez se han comprobado estas dos condiciones previas, uno debe proceder a calcular el índice “calcular variables” (seguir los procedimientos y pasos marcados en el anterior recuadro azul).

Imagen 8. Interface calcular variables, índice “confianza política”.



⁴ ¿Cómo seleccionamos las variables que se deben utilizar para construir el índice? Uno siempre debe justificar por qué selecciona un conjunto de variables para construir un índice representativo de un concepto. Normalmente las variables que deben ser seleccionadas son establecidas por el marco teórico previo del que parte tu investigación. Recuerda, no hay análisis científico sin teoría, al igual que tampoco sino existe análisis empírico.

En el caso de que se haya hecho el proceso de forma correcta este debería ser el resultado que se debería obtener (imagen 8). Una vez realizado el sumatorio se debe clicar el botón de aceptar. El índice ya habrá sido creado y se ubicara como ultima variable dentro de la interface de “*vista de variables*”. Una vez se ha creado el índice, se procede a realizar un análisis de frecuencias para obtener el siguiente resultado (tabla 10).

Tabla 10. Índice confianza política

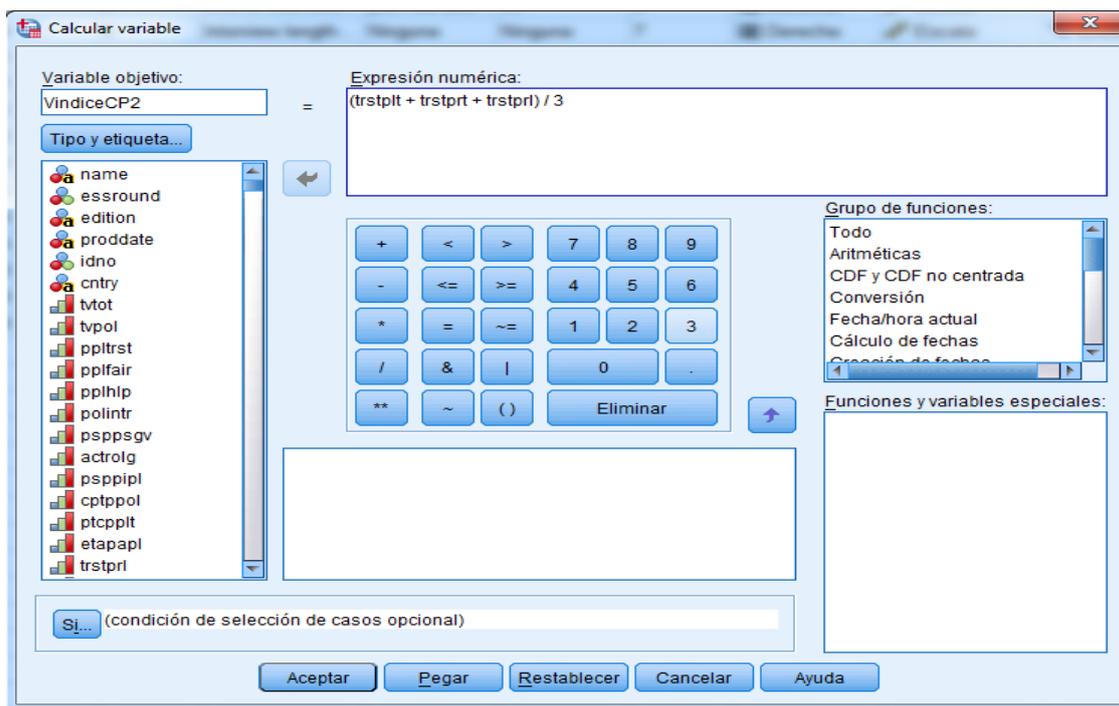
		Frecuència	Porcentaje	Porcentaje válido	Porcentaje acumulado
Valido	.00	253	13.1	14.1	14.1
	1.00	53	2.8	3.0	17.1
	2.00	88	4.6	4.9	22.0
	3.00	99	5.1	5.5	27.6
	4.00	103	5.4	5.8	33.3
	5.00	137	7.1	7.7	41.0
	6.00	91	4.7	5.1	46.1
	7.00	93	4.8	5.2	51.3
	8.00	85	4.4	4.8	56.0
	9.00	92	4.8	5.1	61.2
	10.00	88	4.6	4.9	66.1
	11.00	84	4.4	4.7	70.8
	12.00	62	3.2	3.5	74.2
	13.00	73	3.8	4.1	78.3
	14.00	57	3.0	3.2	81.5
	15.00	88	4.6	4.9	86.4
	16.00	56	2.9	3.1	89.5
	17.00	52	2.7	2.9	92.5
	18.00	37	1.9	2.1	94.5
	19.00	20	1.0	1.1	95.6
	20.00	20	1.0	1.1	96.8
	21.00	14	.7	.8	97.5

	22.00	7	.4	.4	97.9
	23.00	11	.6	.6	98.5
	24.00	11	.6	.6	99.2
	25.00	5	.3	.3	99.4
	26.00	4	.2	.2	99.7
	27.00	2	.1	.1	99.8
	28.00	1	.1	.1	99.8
	29.00	1	.1	.1	99.9
	30.00	2	.1	.1	100.0
	Total	1789	92.9	100.0	
Perdidos	Sistema	136	7.1		
Total		1925	100.0		

Como se observa el programa ha hecho un sumatorio de las categorías de respuesta de las distintas variables seleccionadas. Por lo tanto, el rango de esta nueva variable se sitúa entre 0 y 30 –y si hubiésemos sumado 4 variables se situaría entre 0 y 40-. *¿Cómo sitúa el programa a los individuos en cada uno de los valores?* Mediante un sumatorio. Si un individuo ha marcado que confía “3” en los políticos, “4” en los partidos políticos y “7” en el parlamento de su país, en el índice habrá obtenido una puntuación de “14” (3+4+7).

¿Este índice podría ser obtenido respetando el rango de las variables que lo conforma, es decir, aquel que iba de 0 (nada de confianza) – 10 (plena confianza)? Sí, se puede lograr mediante dos operaciones distintas (con resultados diferentes): 1. Recodificando esta nueva variable y agrupando las categorías de respuesta en una escala que haga equivalente el rango 0-30 al de 0-10 (mirar el apartado “agrupar categorías de respuesta” de este mismo capítulo). Es la operación menos recomendada ya que es más costosa, a pesar de que es más limpia. 2. Cuando se calcula el índice se divide la suma de variables por el número de variables que se está sumando. Es una operación menos costosa, pero, a pesar de que reduce el rango a 0 (valor mínimo)-10 (valor máximo), el número de categorías de respuesta seguirá como si el rango fuera de 0-30 (es decir, 31). Ver tabla 10.

Imagen 9. “Interface” calcular variables, índice “confianza política”.



Si se realiza el procedimiento correctamente debería aparecer la operación que es visible en la **imagen 9** $(trstplt + trstprt + trstprl) / 3$. Para comprobar que se ha realizado de forma apropiada se procede a visualizar la tabla de frecuencias correspondiente.

Tabla 11. Índice confianza política dividido entre el número de variables que lo componen.

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	.00	253	13.1	14.1	14.1
	.33	53	2.8	3.0	17.1
	.67	88	4.6	4.9	22.0
	1.00	99	5.1	5.5	27.6
	1.33	103	5.4	5.8	33.3
	1.67	137	7.1	7.7	41.0
	2.00	91	4.7	5.1	46.1
	2.33	93	4.8	5.2	51.3
	2.67	85	4.4	4.8	56.0
	3.00	92	4.8	5.1	61.2
	3.33	88	4.6	4.9	66.1
	3.67	84	4.4	4.7	70.8

4.00	62	3.2	3.5	74.2
4.33	73	3.8	4.1	78.3
4.67	57	3.0	3.2	81.5
5.00	88	4.6	4.9	86.4
5.33	56	2.9	3.1	89.5
5.67	52	2.7	2.9	92.5
6.00	37	1.9	2.1	94.5
6.33	20	1.0	1.1	95.6
6.67	20	1.0	1.1	96.8
7.00	14	.7	.8	97.5
7.33	7	.4	.4	97.9
7.67	11	.6	.6	98.5
8.00	11	.6	.6	99.2
8.33	5	.3	.3	99.4
8.67	4	.2	.2	99.7
9.00	2	.1	.1	99.8
9.33	1	.1	.1	99.8
9.67	1	.1	.1	99.9
10.00	2	.1	.1	100.0
Total	1789	92.9	100.0	
Perdidos	Sistema	136	7.1	
Total		1925	100.0	

Una vez se ha transformado el índice al formato deseado, se le deben poner las categorías de respuesta. 0 (Ninguna confianza)- 10 o 30 (Toda la confianza).

3.3 Construcción de tipologías.

La elaboración de tipologías consiste en la distinción, dentro de un conjunto de unidades determinadas (individuos, grupos de individuos, hechos sociales...), de grupos amplios que se pueden considerar como homogéneos y que son definidos en torno a un conjunto de características que comparten. Por lo tanto, la agrupación de esas unidades se

fundamenta en ciertas semejanzas definidas a partir de un sub-grupo de características compartidas.

Las tipologías se construyen tanto con variables cuantitativas como cualitativas ya que las variables cuantitativas se pueden cualificar. La sociología suele trabajar más con tipologías que con índices ya que posee un carácter más cualitativo. Además, la construcción de una tipología se sustenta en todo un proceso previo de teorización y conceptualización. Dada la mayor complejidad de este procedimiento se explicara directamente mediante un ejemplo.

Ejemplo:

Imagina que para una investigación que se está llevando a cabo se necesita analizar cómo se comportan diversos colectivos diferenciados de individuos. Concretamente interesa saber el comportamiento de *aquellos que ven programas políticos y confían en la política, aquellos que ven programas políticos y no confían en la política, aquellos que no ven programas políticos y confían en la política y aquellos que no ven programas políticos y no confían en la política.*

Además, nuestro marco teórico indica que aquellos a los que se pueden considerar que realmente ven programas políticos son los que los consumen más de 1,5 horas al día. Por otro lado, también señala que para capturar la confianza en la política se necesita tener en consideración la opinión sobre los políticos y aquella que vierten sobre la institución política central del país (el parlamento). Para ello se cuenta con tres variables: 1. Tvpol: el número de horas que se ven programas políticos al día. 2. Trstprl: la confianza en el parlamento. 3. Trstplt: confianza en los políticos

Si se sintetiza la información que ha sido expresada en los dos anteriores párrafos se obtiene el cuadro siguiente:

Dimensión	Indicadores	Categorías de respuesta		
D1. Confianza en la política	II. Confianza en los políticos	0	No	confió
		1		
		2		
		3		
		4		
		5		
		6		
		7		
		8		
		9		
				10

		I2. Confianza en el parlamento.	0	No	confió
			1		
			2		
			3		
			4		
			5		
			6		
			7		
			8		
			9		
			10	Confió	plenamente
D2. Visualización programas políticos	I1. Visualización de programas políticos diariamente (en horas)		0	No	los veo
			1	Menos	de 0,5h
			2	Entre	0,5h y 1h.
			3	Entre	1h y 1,5h
			4	Entre	1,5h y 2h
			5	Entre	2h y 2,5h
			6	Entre	2,5h y 3h
			7	Mas	de 3h

1. Ven programas políticos y confían en la política.
2. Ven programas políticos y no confían en la política
3. No ven programas políticos y confían en la política.
4. No ven programas políticos y no confían en la política.

La dimensión uno “confianza en la política” se construye mediante dos indicadores. Para ello se necesita previamente construir una nueva variable (índice) que las agrupe (mirar el apartado de construcción de índices de este mismo capítulo). Ahora bien, el marco teórico del que se parte no indica a partir de qué valor se puede considerar que un individuo tiene confianza en la política y cual no. Dado que se carece de un criterio teórico se opta por uno numérico: 0 a 4 (no confió), 5 (ni confió ni desconfió) y de 6 a 10 (confió). El valor central “5” se declarara como caso perdido, ya que se asume que al ser el valor situado en el centro de la escala los individuos que los seleccionan no se posicionan ni en la confianza ni en la desconfianza. Así pues, en la nueva variable (tipología) estarán presentes los que no confían (0-4) y los que confían (6-10).

La dimensión 2 “visualización de programas políticos” se construye mediante un único indicador. En este caso el marco teórico sí que indica que únicamente considera a los que ven programas políticos en la televisión más de 1,5h diarias, como “visualizadores de programas políticos”. Una vez realizado el proceso de construcción del índice de confianza política y su reagrupación en “confían” y “no confían”, se ha realizado la

agrupación de aquellos que ven programas políticos más de 1,5h diarias (que pasan a ser los visualizadores de programas políticos) y los que no.

Ya estructurada la información de esta forma, se debe proceder a “construir” una **tabla de cruces** de las diferentes categorías de respuesta de las variables que componen la tipología. En este caso confianza en la política y visualización de programas políticos.

Tabla 12. Cruce de las diferentes categorías de respuesta.

		Confianza en la política	
		Confía (2)	No confía (1)
Visualización de programas políticos	Si (2)	1	2
	No (1)	3	4

A partir de esta tabla se puede reflejar las características que deben cumplir cada una de las categorías de respuesta de la tipología (al ser un cruce de dos variables con dos categorías de respuesta, se podría hacer de forma automática. No obstante, cuando se trata de una tipología construida con tres variables y/o variables de más de dos categorías de respuesta, este proceso no es tan intuitivo y la tabla ayuda a ordenar la información)

Por tanto, la nueva variable se compondrá de 4 categorías de respuesta que cumplirán las siguientes características.

1. Ven programas políticos y confían en la política.
2. Ven programas políticos y no confían en la política
3. No ven programas políticos y confían en la política.
4. No ven programas políticos y no confían en la política

Una vez se ha realizado este proceso previo e imprescindible se procede a ejecutarlo mediante el SPSS.

Transformar >> Calcular Variable >> Aparece la interface de la imagen 10 >> En el recuadro de variable objetivo señalas un nombre (tipología por ejemplo) >> en Expresión Numérica debes poner el número de tu primera categoría de respuesta (es decir “1”) >> clicas al botón “si” (condición de selección de casos opcional) >> se abre la interface de la imagen 11 >> seleccionas el botón “incluir si el caso satisface la condición” >> Una vez llegados a este punto se debe proceder a indicar cuales condiciones deben cumplir los casos que irán en la categoría de respuesta 1 de la tipología. Recordemos: ver programas políticos (2) y confianza en la política (2). Es decir, los individuos que cumplen la condición de ver programas políticos (categoría de respuesta 2 de esa variable) y confiar en la política (categoría de respuesta 2 de esa variable) simultáneamente (interface imagen 11). >> continuar. El proceso se vuelve a repetir siguiendo los mismos pasos pero esta vez para las condiciones de las categorías de respuesta 2. Ver programas políticos (2) y no confiar en la política (1) y así sucesivamente hasta completar todas las categorías de respuesta de la tipología. El nombre de la variable objetivo siempre debe ser la misma.

Imagen 10. Interface calcular variable.

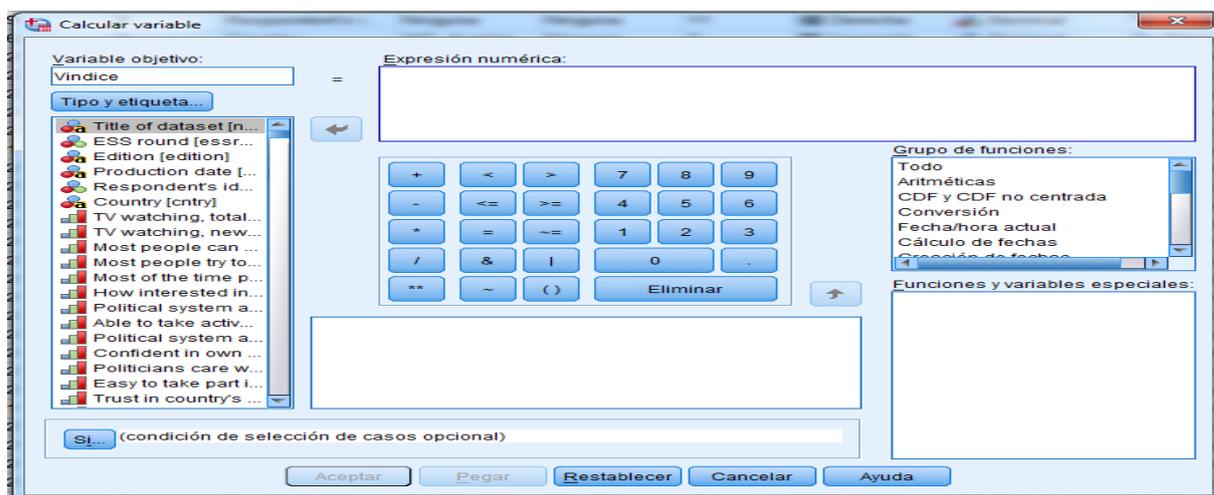
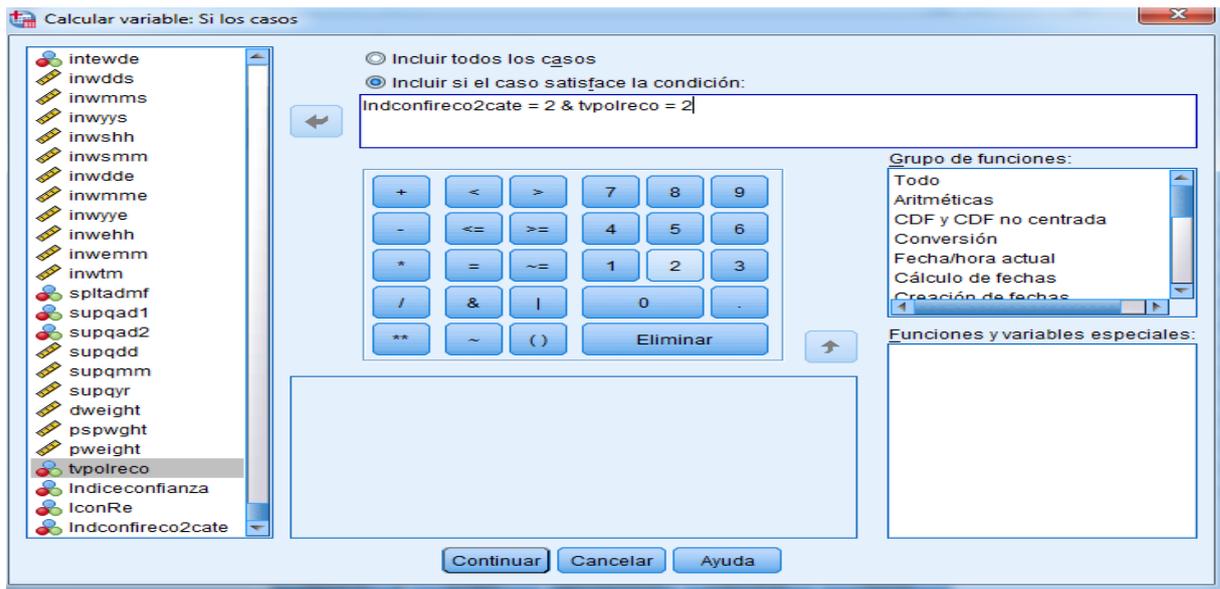


Imagen 11. Interface botón “si...”



Una vez se ha realizado el proceso se expone el resultado de la tabla de contingencia de la nueva variable tipológica para observar sus resultados

Imagen 12. Tipología.

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Ven programas políticos y confían en la política	40	2.1	2.6	2.6
	Ven programas políticos y no confían en la política	195	10.1	12.9	15.5
	No ven programas políticos y confían en la política.	161	8.4	10.6	26.2
	No ven programas políticos y no confían en la política	1117	58.0	73.8	100.0
	Total	1513	78.6	100.0	
Perdidos	Sistema	412	21.4		
Total		1925	100.0		