

Práctica 5

Elaboración de un índice normalizado

La práctica tiene un objetivo combinado, por una parte quiere mostrar las posibilidades de generación de nuevas variables a través de la información que proporciona la respuesta a distintos enunciados (items) y por otra ilustrar el uso de distribuciones de probabilidad para confeccionar índices estandarizados.

La realización de la práctica permite desarrollar la capacidad para construir medidas sintéticas mediante el uso de comandos de sintaxis así como introducirse en el uso de las distribuciones de probabilidad como modelos.

Primera parte: Elaboración del índice

Comenzaremos cargando, como en otras ocasiones, el fichero CIS2927.SAV. Una vez abierto el fichero nos fijamos en la pregunta 9. (Recuerde que el cuestionario está en el fichero cues2927.pdf.

P.9 A continuación voy a citarle una lista de temas que preocupan hoy a la gente. ¿Cree Ud. que cada uno de estos problemas, durante el año 2012 mejorará o empeorará?					
	(NO LEER)				
	Mejorará	Seguirá igual	Empeorará	NS	NC
- El paro	1	2	3	8	9 (50)
- La inseguridad ciudadana	1	2	3	8	9 (51)
- El terrorismo.....	1	2	3	8	9 (52)
- Las drogas	1	2	3	8	9 (53)
- El medio ambiente	1	2	3	8	9 (54)
- La asistencia sanitaria	1	2	3	8	9 (55)
- La vivienda	1	2	3	8	9 (56)
- La inmigración	1	2	3	8	9 (57)
- La educación	1	2	3	8	9 (58)
- Las pensiones	1	2	3	8	9 (59)

La pregunta 9 consta de un total de 10 items sobre los que se pregunta la percepción sobre su evolución. La percepción sobre el futuro no es una cuestión simple, sino compleja en la medida en que el "futuro" tiene múltiples dimensiones. En la medida de lo posible recoger la amplitud de dimensiones –condiciones materiales (las pensiones), condiciones contextuales (la inseguridad ciudadana)- así como el diferente peso que dichas dimensiones tienen para los

entrevistados –por ejemplo, la vivienda preocupa más o menos en función del estatus- permite una lectura con mayor profundidad.

A partir de estos 10 items se va a construir un índice –variable- que recoja de forma compuesta la opinión de cada entrevistado sobre las diferentes cuestiones. Se construirá un indicador aditivo es decir que sea simplemente la suma de las 10 respuestas y cuyo valor mínimo lo alcance una persona que piensa que las 10 cuestiones empeorarán, mientras que el valor máximo lo alcance aquella persona que considere que todo mejorará. Se trata de un índice de optimismo –confianza en las condiciones sociales-.

El primer item de la pregunta 9 se refiere a “El paro” que es la variable P901. Las respuestas posibles son:

1-Mejorará

2-Seguirá igual

3-Empeorará

8-NS

9-NC

Par que el valor máximo sea optimismo y el mínimo pesimismo podemos convertir los valores de la siguiente manera:

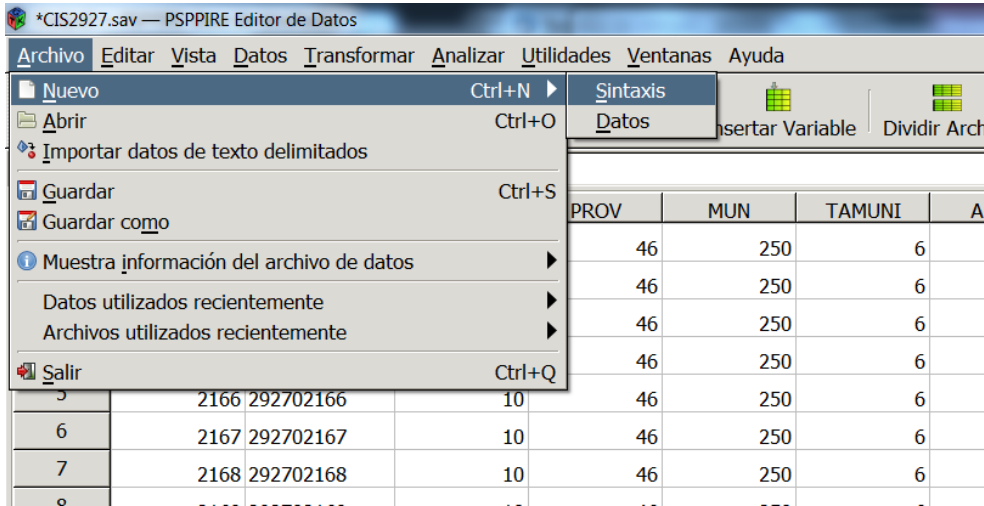
Valor Original		Valor Transformado
1	Mejorará	3
2	Seguirá igual	2
3	Empeorará	1
8	NS	0
9	NC	0

Esta transformación la repetiremos para cada uno de los 10 items desde “El paro” hasta “Las pensiones”. Este conjunto de operaciones lo haremos mediante el comando RECODE, bien en un fichero de sintaxis –opción recomendada- o bien utilizando menús (**Transformar: Recodificar en variables Diferentes**).

Una vez transformadas las variables simplemente sumaremos –por eso se denomina índice agregado o aditivo- los valores de las respuestas, utilizaremos el comando COMPUTE. De esta forma quien haya considerado que cada una de las 10 dimensiones mejorará obtendrá un valor de 30 (10x3). Mientras que quien considere que todo empeorará obtendría un valor de 10 (10x1). Como puede apreciarse si examinamos las frecuencias de cada uno de los items, no todos los entrevistados han dado su opinión sobre los 10 items. Cuando han respondido que no sabían o no han contestado les hemos dado un valor de 0 a dichos items. Por lo tanto, el indicador puede alcanzar valores mínimos de 0, cuando una persona no ha respondido a ninguna de las cuestiones. (Por ejemplo quien tenga un valor de 1, quiere decir que no ha respondido a 9 cuestiones y considera que una de ellas empeorará.)

En el caso en que no hay respondido a ninguna de las preguntas, vamos a sacar dicho cuestionario del análisis. A continuación exponemos el fragmento de sintaxis utilizado para realizar dicho conjunto de transformaciones.

Vamos a abrir un nuevo fichero de sintaxis (**Archivo: Nuevo: Sintaxis**)



Y en la ventana que se abre escribimos la siguiente secuencia:

```
*****COMANDOS PARA GENERACIÓN DE INDICADOR CON LA PREGUNTA 9.

RECODE P901 TO P910
(1 = 3) (2 = 2) (3 = 1) (8,9 =0)
INTO PA901 TO PA910.

COMPUTE INDI=PA901+PA902+PA903+PA904+PA905+PA906+PA907+PA908+PA909+PA910.

MISSING VALUE INDI (0).

VARIABLE LABEL INDI 'INDICE DE CONFIANZA EN LAS CONDICIONES SOCIALES'.

EXECUTE.
```

En total hay 6 comandos (fíjese que hay en total 6 puntos “.” que indican el final de los comandos).

El primer comando ya lo conocemos “*”. Comienza con un asterisco y termina con un punto y se utiliza para introducir comentarios.

El segundo comando RECODE ha sido utilizado en prácticas anteriores. Ahora lo utilizamos de forma en que recodifica en una nueva variable. Para una sola variable indicamos después del nombre del comando (RECODE) el nombre de la variable. En este caso es P901 que es el primer ítem de la pregunta 9, la percepción sobre la evolución de “El paro”, a continuación indicamos los cambios de valores originales a los valores nuevos mediante paréntesis. Así el valor 1 se transformará en 3, etc... La partícula **INTO** que viene a continuación lo que indica es la variable en la que se van a guardar las transformaciones, que en este caso la denominados PA901

RECODE P901 (1 = 3) (2 = 2) (3 = 1) (8,9 =0) INTO PA901.

Con este comando hemos generado una nueva variable que es copia y que ha cambiado los valores respecto a la original.

En el fichero de sintaxis hemos escrito el comando que hace esta operación de forma automática para las 10 variables mediante el uso de la partícula **TO**.

RECODE P901 **TO** P910 (1 = 3) (2 = 2) (3 = 1) (8,9 =0) INTO PA901 **TO** PA910.

De esta forma para cada uno de los ítems hemos generado una nueva variable:

	Variable Original	Variable Transformada
El paro	P901	PA901
La inseguridad ciudadana	P902	PA902
El terrorismo	P903	PA903
Las drogas	P904	PA904
El medio ambiente	P905	PA905
La asistencia sanitaria	P906	PA906
La vivienda	P907	PA907
La inmigración	P908	PA908
La educación	P909	PA909
Las pensiones	P910	PA910

Con el comando COMPUTE generamos una nueva variable que va a sumar los valores de cada una de las 10 variables transformadas. El uso de este comando resulta sencillo, después de COMPUTE se indica el nombre de la variable nueva –en este caso la denominamos INDI- y después del igual la expresión algebraica que la compone, en este caso la suma (+) de las diferentes variables.

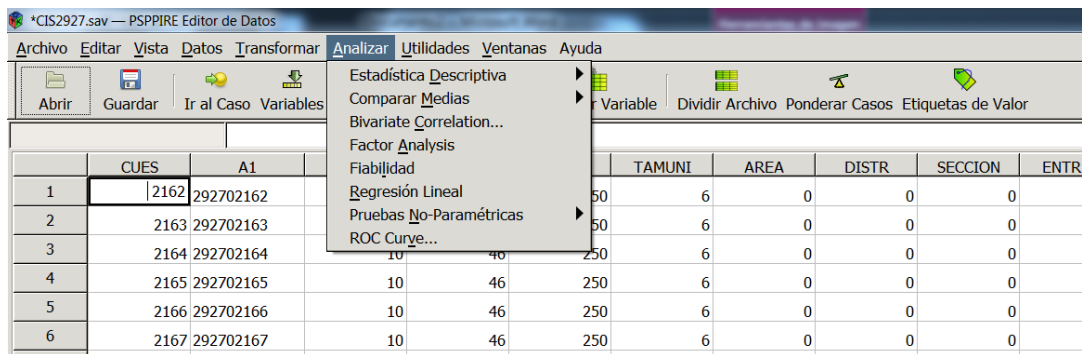
Después del comando COMPUTE encontramos el comando MISSING VALUES, que nos indica que los cuestionarios que tengan un valor “0” en la variable INDI sean sacados del análisis.

Recordemos que quien tiene un valor “0” es porque no ha contestado a ninguna de las 10 cuestiones de la pregunta.

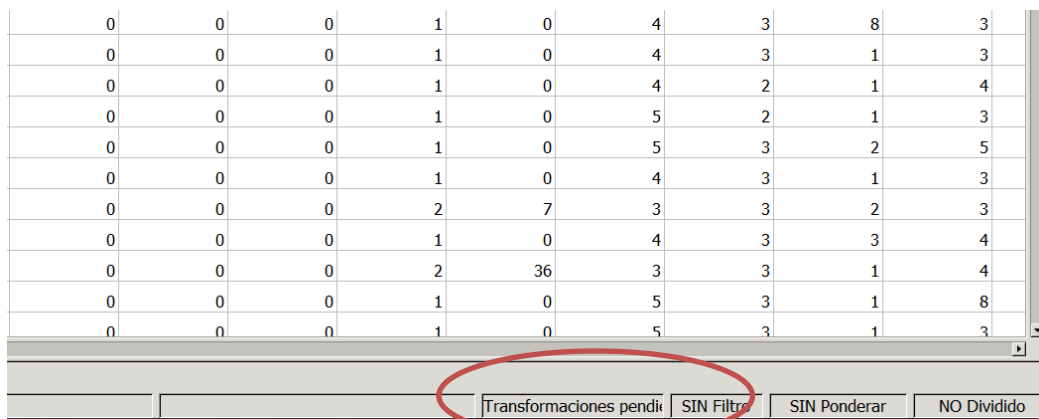
El comando de declaración de valores perdidos es seguido por el comando que da nombre a la variable INDI. En este caso se rotula como “ÍNDICE DE CONFIANZA EN LAS CONDICIONES SOCIALES”.

El último de los comandos –el sexto- es EXECUTE, dicho comando indica a PSPP que ejecute todas las órdenes anteriores. Los comandos de transformación no se ejecutan directamente sino que se acumulan como órdenes que se ejecutarán posteriormente.

-Al solicitar un resultado mediante cualquiera de las instrucciones del apartado **Analizar**, las órdenes acumuladas se ejecutan.



-Si en el fichero de sintaxis no se introduce la orden EXECUTE, después de ejecutar distintas instrucciones de transformación nos encontraremos en la barra inferior la expresión “Transformaciones pendientes”:



Para activar dichas transformaciones podemos hacerlo manualmente mediante: **Transformar: Ejecutar Transformaciones pendientes** (que es el equivalente al comando EXECUTE).

*CIS2927.sav — PSPPIRE Editor de Datos

Archivo Editar Vista Datos Transformar Analizar Utilidades Ventanas Ayuda

Abrir Guardar Ir al C... Variable Dividir Archivo Pond

Calcular
Rango de casos
Recodificar en las Mismas Variables
Recodificar en variables Diferentes
Ejecutar Transformaciones pendientes Ctrl+G

	CUES					TAMUNI	AREA
1	2162	292702162	10	46	250	6	0
2	2163	292702163	10	46	250	6	0
3	2164	292702164	10	46	250	6	0
4	2165	292702165	10	46	250	6	0
5	2166	292702166	10	46	250	6	0
6	2167	292702167	10	46	250	6	0
7	2168	292702168	10	46	250	6	0

Una vez escritos los seis comandos, conviene grabar el fichero de sintaxis, ejecutamos las instrucciones. A continuación vamos a estudiar la distribución de la nueva variable obtenida, para ello vamos a **Analizar: Estadística Descriptiva: Frecuencias**

En este caso además de la tabla de frecuencias y de los estadísticos habituales podemos también seleccionar un gráfico de dicha distribución. En la ventana de Frecuencias, aparece un botón **Charts...** que nos conduce a una nueva ventana. En dicha ventana vamos a seleccionar Histograma marcando la opción "Draw Histogram" y "Superimpose normal curve". Esta última opción añadirá al histograma la distribución normal que se corresponde con ese número de casos y estadísticos (media y varianza).

*CIS2927.sav — PSPPIRE Editor de Datos

Frecuencias: Charts

Chart Formatting
 Exclude values below 0
 Exclude values above 100
 Histograms
 Draw Histograms
 Superimpose normal curve
 Scale: Frecuencias Percentages
 Pie Charts
 Draw pie charts
 Include slices for missing values

Variable(s): ind1
 Estadísticas:
 Media
 Desviación Estándar
 Include missing values

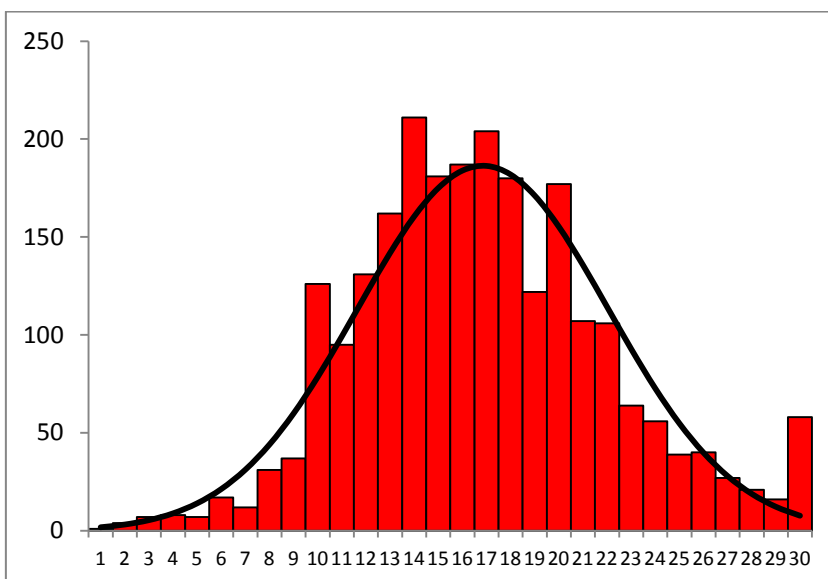
	MUN	TAMUNI	AREA	DISTR	SECCION	ENTR
6	2167	292702167	10	46	250	0
7	2168	292702168	10	46	250	0
8	2169	292702169	10	46	250	0
9	2170	292702170	10	46	250	0
10	2171	292702171	10	46	250	0
11	2172	292702172	10	46	250	0
12	2173	292702173	10	46	250	0
13	2174	292702174	10	46	250	0

Los resultados, que obtenemos son:

FRECUENCIAS
 FRECUENCIAS
 /VARIABLES= indi
 /FORMAT=AVALUE TABLE
 /HISTOGRAM=NORMAL PERCENT.

Etiqueta de Valor	Valor	Frecuencia	Porcentaje	Valid Percent	Cum Percent
	1,00	1	,04	,04	,04
	2,00	4	,16	,16	,21
	3,00	7	,28	,29	,49
	4,00	8	,32	,33	,82
	5,00	7	,28	,29	1,11
	6,00	17	,69	,70	1,81
	7,00	12	,48	,49	2,30
	8,00	31	1,25	1,27	3,57
	9,00	37	1,49	1,52	5,09
	10,00	126	5,08	5,18	10,27
	11,00	95	3,83	3,90	14,17
	12,00	131	5,28	5,38	19,56
	13,00	162	6,53	6,66	26,21
	14,00	211	8,51	8,67	34,88
	15,00	181	7,30	7,44	42,32
	16,00	187	7,54	7,68	50,00
	17,00	204	8,23	8,38	58,38
	18,00	180	7,26	7,40	65,78
	19,00	122	4,92	5,01	70,79
	20,00	177	7,14	7,27	78,06
	21,00	107	4,31	4,40	82,46
	22,00	106	4,27	4,35	86,81
	23,00	64	2,58	2,63	89,44
	24,00	56	2,26	2,30	91,74
	25,00	39	1,57	1,60	93,34
	26,00	40	1,61	1,64	94,99
	27,00	27	1,09	1,11	96,10
	28,00	21	,85	,86	96,96
	29,00	16	,65	,66	97,62
	30,00	58	2,34	2,38	100,00
	,00	46	1,85	Perdidos	
Tota		2480	100,0	100,0	

<i>N</i>	<i>Válida</i>	2434
	<i>Perdidos</i>	46
<i>Media</i>		16,85
<i>Desv. Sto</i>		5,21
<i>Mínimo</i>		1,00
<i>Máximo</i>		30,00



Como podemos observar el índice varía de 1 a 30 (recuérdese que los 46 casos con valor “0” han sido excluidos). La media es 16,85 y la desviación típica 5,21. Que la media sea superior a 15 señala que hay una leve tendencia al optimismo, si bien parece existir un consenso en que no habrá ni mejoría ni empeoramiento de las condiciones sociales.

El gráfico nos muestra que el índice construido tiene una forma próxima a una distribución normal. Por lo general multitud de índices demoscópicos tienen este tipo de distribución que concentra las opiniones en posiciones centrales.

El índice que hemos construido tiene diferentes carencias. Por una parte resulta confuso en su interpretación, dado que varía de 1 a 30. Una solución sería la estandarización, por ejemplo acotar sus valores entre 0 y 10.

(En este caso simplemente dividiríamos los valores entre 3 y obtendríamos una variable con valores entre 0 y 10. La instrucción – llamamos al índice escalado como INDI10- sería:

COMPUTE INDI10=INDI/3.

Si bien, que varíe entre 0 y 10 nos permite de forma más intuitiva su interpretación, y dado el caso, su comparación con otros indicadores –que mantengan la misma escala-, nuestro índice no está normalizado.

La normalización de los índices resulta muy interesante, además de necesaria, de cara a la interpretación y comparación de medidas. En este caso estamos midiendo – optimismo/pesimismo- evidentemente no hay una unidad de medida para esta magnitud. (No utilizamos metros, ni kilos...) Mediante la normalización, que consiste en utilizar la varianza como norma, obtenemos una medida que:

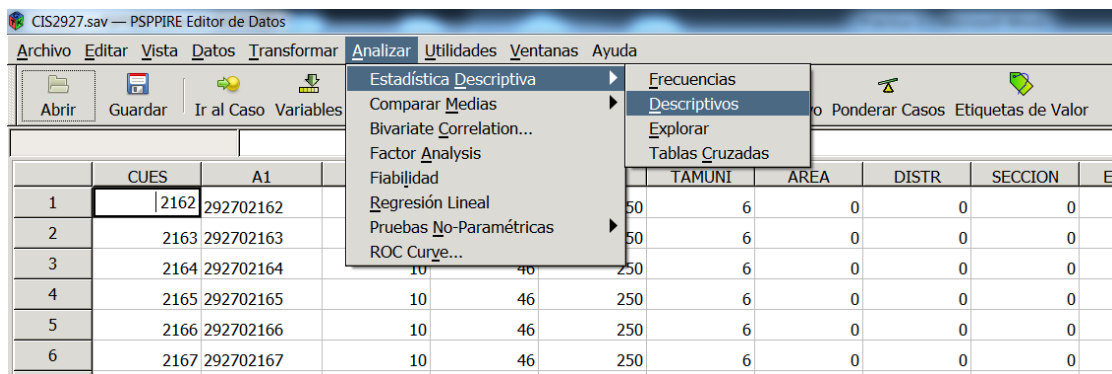
Por una parte, puede ser comparable con otras variables normalizadas. Por ejemplo, con un índice de confianza en el gobierno.

Por otra parte, podemos interpretar la diferencia de valores entre dos individuos como distancia. Por ejemplo, en nuestro índice podemos pensar a priori que quienes tienen un índice de 5, tienen la mitad de confianza que quienes tienen un 10, o que quienes tienen una confianza de 7,5 están a medio camino que quienes tienen una confianza de 5 y de 10. Sin embargo el problema es que los valores del índice no están escalados. (Un valor de 20 no significa que se tenga doble confianza sobre quien tiene un valor de 10, simplemente que tiene más confianza). Mediante la normalización conseguimos ofrecer una medida interpretable respecto a la media.

La normalización además resulta interesante cuando medimos dimensiones subjetivas como es la confianza. Que el valor 15 sea la mitad de la escala que hemos utilizado no quiere decir que se corresponda con el grupo de quienes piensan que todo seguirá igual. De hecho, en nuestro caso un valor de 15 está por debajo de la media, es decir realmente decir que todo sigue igual es en el conjunto de la opinión una posición discretamente pesimista.

Segunda parte; Normalización del índice

La forma habitual de normalización es el uso de unidades Z. Para transformar la variable a unidades Z podemos hacerlo mediante el recurso a los menús de PSPP. Esta vez sin embargo en vez de Frecuencias vamos a Descriptivos. Desde **Analizar: Estadística Descriptiva: Descriptivos**



En la ventana que se abre seleccionamos la variable (INDI) los estadísticos que sean de nuestro interés y marcamos la última casilla en la parte de Opciones: **Guardar puntuaciones-Z...**

908	PA909	PA910	INDI	ZINDI
2.00	1.00	2.00	15.00	-.35
3.00	1.00	2.00	17.00	-.03
3.00	3.00	2.00	22.00	.99
2.00	1.00	1.00	14.00	-.55
2.00	1.00	2.00	15.00	-.35
2.00	1.00	2.00	15.00	-.35
2.00	2.00	3.00	24.00	1.37
3.00	1.00	1.00	15.00	-.35
2.00	1.00	1.00	16.00	-.16
2.00	1.00	2.00	21.00	.80
3.00	2.00	.00	14.00	-.55
2.00	.00	3.00	17.00	.03
.00	1.00	.00	10.00	-1.31
1.00	2.00	.00	10.00	-1.31
3.00	3.00	3.00	30.00	2.53
1.00	1.00	2.00	16.00	-.16
1.00	1.00	2.00	11.00	-1.17

Ahora podemos analizar la variable ZINDI solicitando una tabla de frecuencias y los principales estadísticos. Como sabemos el valor Z en unidades Z es la diferencia del valor de la variable estandarizado por su desviación típica. Por ejemplo, los entrevistados que han obtenido un valor 10 en el índice, tendrán como valor Z:

$$Z_x = \frac{(x - \bar{x})}{S_x} = \frac{10 - 16,85}{5,21} = -1,31$$

<i>Etiqueta de Valor</i>	<i>Valor</i>	<i>Frecuencia</i>	<i>Porcentaje</i>	<i>Valid Percent</i>	<i>Cum Percent</i>
	-3,04	1	,04	,04	,04
	-2,85	4	,16	,16	,21
	-2,66	7	,28	,29	,49
	-2,47	8	,32	,33	,82
	-2,27	7	,28	,29	1,11
	-2,08	17	,69	,70	1,81
	-1,89	12	,48	,49	2,30
	-1,70	31	1,25	1,27	3,57
	-1,51	37	1,49	1,52	5,09
	-1,31	126	5,08	5,18	10,27
	-1,12	95	3,83	3,90	14,17
	-,93	131	5,28	5,38	19,56
	-,74	162	6,53	6,66	26,21
	-,55	211	8,51	8,67	34,88
	-,35	181	7,30	7,44	42,32
	-,16	187	7,54	7,68	50,00
	,03	204	8,23	8,38	58,38
	,22	180	7,26	7,40	65,78
	,41	122	4,92	5,01	70,79
	,61	177	7,14	7,27	78,06
	,80	107	4,31	4,40	82,46
	,99	106	4,27	4,35	86,81
	1,18	64	2,58	2,63	89,44
	1,37	56	2,26	2,30	91,74
	1,57	39	1,57	1,60	93,34
	1,76	40	1,61	1,64	94,99
	1,95	27	1,09	1,11	96,10
	2,14	21	,85	,86	96,96
	2,33	16	,65	,66	97,62
	2,53	58	2,34	2,38	100,00
	.	46	1,85	Perdidos	
<i>Tota</i>		2480	100,0	100,0	

<i>N</i>	<i>Válida</i>	2434
	<i>Perdidos</i>	46
<i>Media</i>		,00
<i>Desv Sta</i>		1,00
<i>Mínima</i>		-3,04
<i>Máxima</i>		2,53

Podemos observar que ahora nuestra distribución tiene como media "0" y desviación típica "1", es decir es una distribución normal estándar. Quienes tienen valores negativos tienen una opinión más pesimista que la media mientras que quienes alcanzan valores positivos son más optimistas que la media. Quienes tienen valores entre -1 y +1 son, como sabemos, aproximadamente el 68% del conjunto. Quienes tienen valores extremos inferiores a -2 o superiores a 2, serán aproximadamente el 5% del total y serán o extremadamente pesimista o extremadamente optimistas.

Una vez obtenido el índice normalizado podemos utilizarlo como variable analítica. Vamos a considerar ahora la relación que existe entre posición ideológica y confianza en las condiciones sociales. Para ello vamos a obtener las medias en función de los valores de la pregunta 27.

P.27 Cuando se habla de política se utilizan normalmente las expresiones izquierda y derecha. En esta tarjeta hay una serie de casillas que van de izquierda a derecha. ¿En qué casilla se colocaría Ud.? **(MOSTRAR TARJETA ESCALA)**. **(PEDIR A LA PERSONA ENTREVISTADA QUE INDIQUE LA CASILLA EN LA QUE SE COLOCARÍA Y REDONDEAR EL NÚMERO CORRESPONDIENTE)**.

(159)(160)

Izda.

Dcha.

											NS	NC
01	02	03	04	05	06	07	08	09	10		98	99

Para obtener las medias del índice respecto a los valores de otra variable vamos a utilizar el siguiente comando que no es accesible desde los menús.

MEANS TABLES=ZINDI BY P27 /CELLS=MEAN STDDEV.

Dicho comando lo escribimos en el fichero de sintaxis.

El comando MEANS permite la obtención de medias (y otros estadísticos de una variable). Mediante el subcomando TABLES indicamos la variable. Por ejemplo, si escribimos:

```
MEANS TABLES=ZINDI.
```

Obtendremos la media de la variable ZINDI.

En nuestro caso, hemos escrito

```
MEANS TABLES=ZINDI BY P27.
```

Con el término BY indicamos la variable independiente. En este caso vamos a solicitar las medias del índice de confianza en el gobierno (ZINDI) para cada valor de la escala de ideología (P27).

El subcomando CELLS nos permite especificar que estadísticos deseamos. En este caso (CELLS=MEAN STDDEV) la media aritmética y la desviación típica. Otras opciones que podemos añadir son:

- MEAN Media aritmética.
- COUNT Frecuencia.
- STDDEV Desviación típica.
- SEMEAN Error típico de la media.
- SUM Suma de valores.
- MIN Valor mínimo.
- MAX Valor máximo.
- RANGE Rango, diferencia entre el valor máximo y el mínimo.
- VARIANCE Varianza.
- HARMONIC Media armónica.
- GEOMETRIC Media geométrica.

Una vez ejecutado obtenemos los siguientes resultados. Si nos fijamos en las medias obtenidas podemos comprobar que en las posiciones de izquierda hay un pesimismo relativo, son valores negativos. Es decir la puntuación media que dan las personas que se declaran de izquierda es inferior a la media global. Por el contrario las posiciones ideológicas de derechas resultan claramente optimistas. Obsérvese que las posiciones más extremas de derecha -8 y 9- ofrecen medias positivas y de valor superior a 0,5 (Téngase en cuenta que entre Z=0 y Z=0,5 hay un 20% de casos). Es decir son diferencias muy relevantes sobre el conjunto de la población.

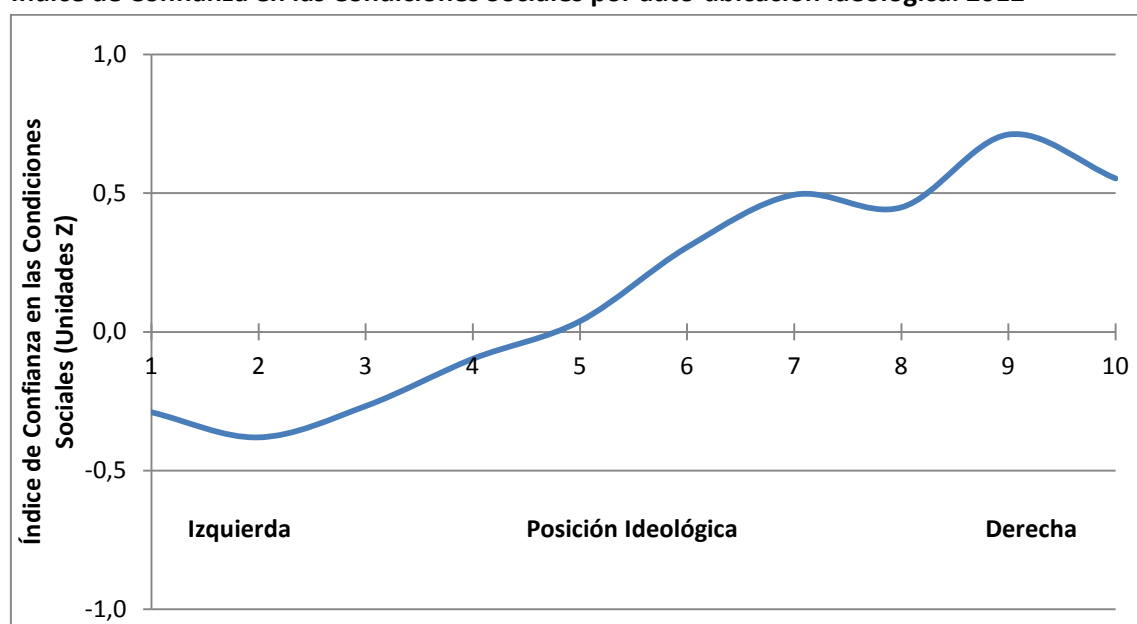
	Pregunta 27	Media	Desviación Típica
Izquierda	1	-0,29	0,85
	2	-0,38	0,74
	3	-0,27	0,80
	4	-0,10	0,80
	5	0,04	0,97
	6	0,31	0,98
	7	0,49	1,10
	8	0,45	1,16
	9	0,71	1,14
Derecha	10	0,55	1,07

El análisis de los resultados contenidos en la tabla nos está indicando que la posición ideológica guarda relación con la confianza sobre el futuro de las condiciones sociales. Si observamos el gráfico siguiente podemos comprobar que en términos generales, cuanto más a la izquierda mayor pesimismo y cuanto más a la derecha mayor optimismo. Podemos concluir que la posición ideológica condiciona la percepción de la realidad.

No obstante tenemos que tener en cuenta que en las relaciones entre variables intervienen terceras variables. Por ejemplo, en este caso resulta relevante el contexto –el gobierno es de derechas-. Por ello quienes se consideran más afines al gobierno probablemente, por ello, sean más optimistas.

Es importante tener presente que los datos solo pueden interpretarse cuando existen marcos teóricos que soporten la investigación. Los estadísticos nos informan de relaciones pero no explican las causas.

Índice de Confianza en las Condiciones Sociales por auto-ubicación ideológica. 2012



Fuente: CIS 2927. Febrero de 2012. Elaboración propia.