

EXPLOTACIÓN DE LA ENCUESTA DE POBLACIÓN ACTIVA

Admitido para publicación en *Metodología de Encuestas* nº 7, vol. 1 de 2006

Enviado el 16 de septiembre de 2005

[Revisado 3/8/06]

José Saturnino Martínez García,

Tlf.: 922 31 74 45, Fax: 922 31 7364

E.mail [josamaga@ull.es](mailto:josamaga@ull.es), <http://webpages.ull.es/users/josamaga/>

Departamento de Sociología, Campus de Guajara

Universidad de La Laguna

La Laguna, Tenerife (38071).

EXPLOTACIÓN DE LA ENCUESTA DE POBLACIÓN ACTIVA

Palabras clave: población activa, encuesta, explotación de datos, investigación, mercado de trabajo

Resumen: Recientemente el Instituto Nacional de Estadística español ha puesto en internet a libre disposición gran cantidad de ficheros de microdatos. En este artículo se presenta cómo explotar uno de ellos, el de la Encuesta de Población Activa del primer trimestre de 2005. Por un lado, a través de la explotación de este fichero es posible familiarizarse con los procedimientos necesarios para explotar el resto de ficheros. Por otro lado, se comentan algunas posibilidades y limitaciones particulares de dicha encuesta. El texto se centra en el uso de la aplicación informática SPSS, pero se presenta con la suficiente información como para que sea útil para usuarios de otras aplicaciones.

Keywords: Labour Force Survey, data management, research, labour market.

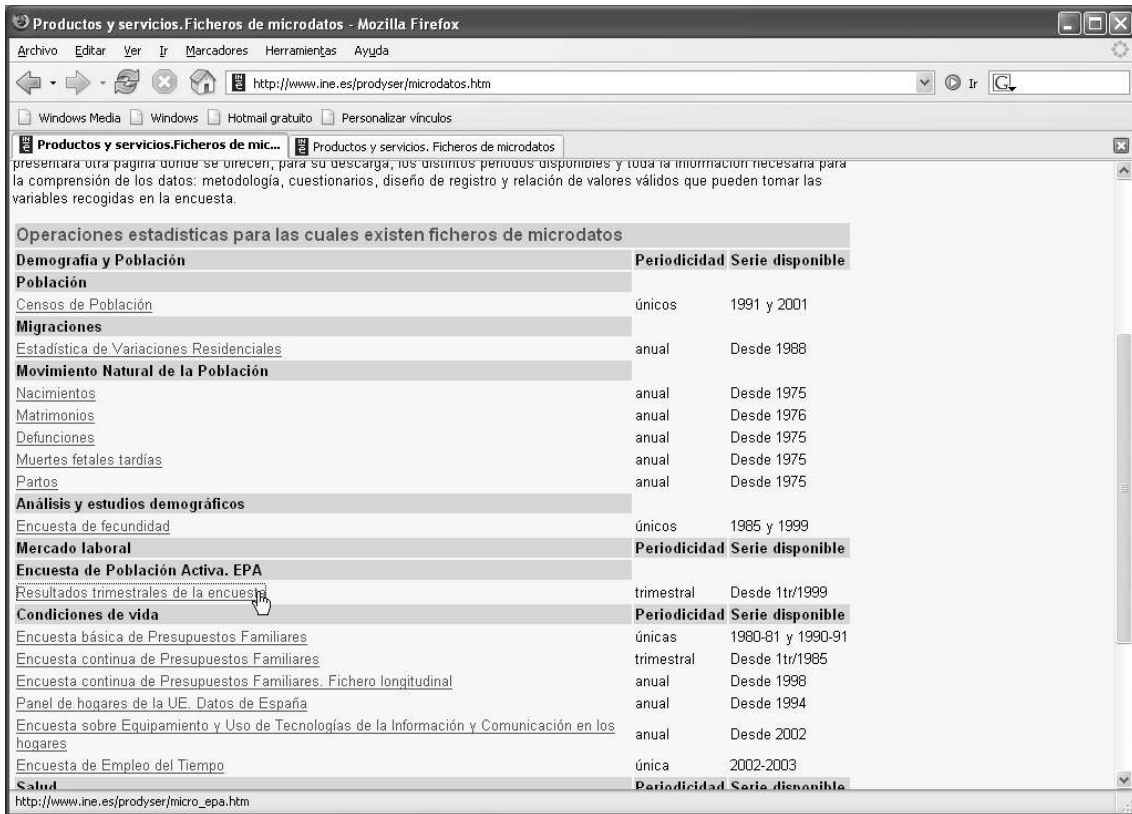
Abstract: Actually it's possible download data sets from Spaniard Statistical National Bureau (INE). This paper explains how management this data sets, with especial focus in Labour Force Survey (first trimester in 2005). For one hand, it is possible learn how know about this data sets. For another hand, it's showed some possibilities and limitations for this survey. This paper focus in SPSS, but there are enough information to be useful for other software.

Desde junio de 2005 el Instituto Nacional de Estadística español (INE) ha puesto a disposición de quien lo desee buena parte de sus ficheros de microdatos gratuitamente, mediante el procedimiento tan simple de descargarlos de Internet (<http://www.ine.es/prodyser/microdatos.htm>). El INE denomina fichero de microdatos a los ficheros de las encuestas que realiza, con los datos detallados a nivel individual, para diferenciarlos de los ficheros de datos agregados, cuyos resultados es posible obtener mediante las diversas herramientas de explotación de datos disponibles en su web (base de datos Tempus). La ventaja de trabajar con los ficheros de microdatos es obvia: la explotación es mucho más flexible, pues permite todo tipo de análisis estadísticos, así como relacionar variables que no siempre se pueden cruzar en otras aplicaciones *on line*. Pero también tiene sus inconvenientes, como que es necesaria una mayor formación en el uso de *software*, así como mejores conocimientos de estadística, para conocer no solo las posibilidades, sino también las limitaciones de trabajar con estas encuestas. Este artículo está pensado para facilitar el trabajo a todas aquellas personas formadas en el uso aplicaciones informáticas, pero que no conocen estas encuestas del INE. Estudiando solo una de ellas, la Encuesta de Población Activa (EPA), nos podremos hacer una idea de cómo explotar estos datos. También nos centraremos en una sola aplicación informática, el SPSS, pero con el suficiente detalle como para que aquellos que empleen otro *software* les resulte sencillo aplicar lo que aquí decimos.

### DESCARGAR LOS DATOS

El fichero general de los datos está en la dirección que hemos señalado previamente. En dicha página, nos encontramos con la siguiente información:

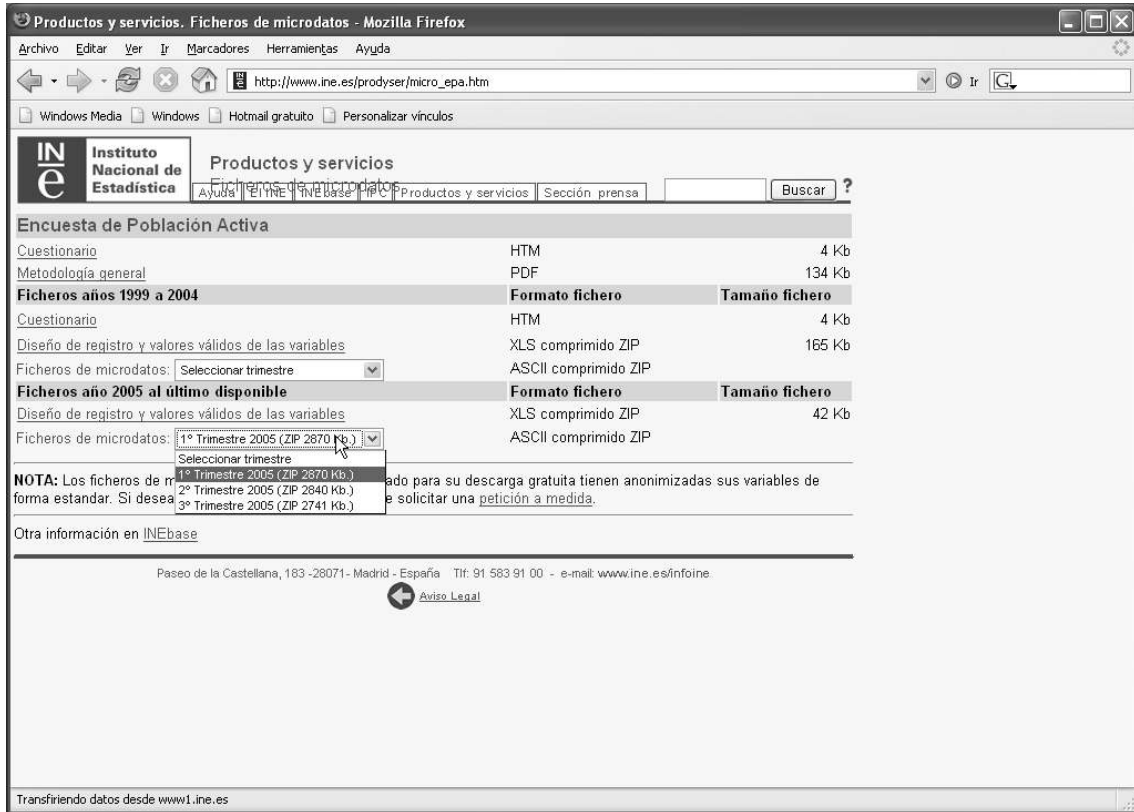
Pantalla 1



Como vemos, aparece un listado de las distintas encuestas de las que se puede descargar la información. Si pinchamos sobre *Resultados trimestrales de la encuesta*, en el apartado de la EPA, pasamos a la ventana en la que está toda la información que necesitamos. Si posamos el cursor sobre la línea de diálogo se nos despliegan las encuestas que podemos descargar, con información sobre su tamaño comprimido en ZIP<sup>1</sup> (descomprimido el *peso* del fichero es unas 10 veces mayor).

<sup>1</sup> Si no dispone de la aplicación, puede descargarla <http://www.winzip.com/downwzeval.htm> . La puede comprar, o si no, bajar la versión de prueba, suficiente para hacer esta descompresión.

Pantalla 2



Si selecciona el trimestre deseado, el primer trimestre de la EPA de 2005 en este caso, se abrirá un cuadro de diálogo en el que le pedirá si lo que desea es abrir el fichero o guardarlo; elegimos la opción de guardar. Por simplificar, crearemos el siguiente directorio de trabajo *C:\EPA*. Guardamos los datos en este directorio. Una vez guardados, debemos descomprimirlos, con lo que el resultado será el de dos ficheros, uno con los datos comprimidos, y el otro con los datos descomprimidos, en formato ASCII. Para poder leer los datos en ASCII necesitamos consultar el resto de la información que aparece en la Pantalla 2; **Error! No se encuentra el origen de la referencia.** Si no está familiarizado con la EPA es muy importante que consulte con detenimiento la Metodología General (INE 2005) en la que viene el mínimo necesario de conocimientos sobre el diseño de la encuesta y la definición de las principales variables. Y como en

toda explotación de datos, es fundamental que conozca bien el cuestionario, no solo para saber con exactitud cuál es el enunciado de cada pregunta que produce la variable con la que se va a trabajar, sino para saber qué preguntas se hacen a qué parte de la población, pues son muchas las que están filtradas por contestaciones previas.

Para poder leer el fichero en ASCII es necesario el “Diseño de registro y valores válidos de las variables”, que también está comprimido y en formato Excel. Con SPSS, teniendo esta información disponible, es posible leer el fichero de modo interactivo, pero no es recomendable, por dos motivos: es más difícil detectar los posibles errores que se cometan en la lectura y el procedimiento es válido solo una vez, y como esta encuesta se distribuye cuatro veces al año, conviene automatizar al máximo el proceso. Por ello, es mejor realizar esta lectura en el modo de sintaxis de SPSS. Si no está familiarizado con el uso de la sintaxis en SPSS puede consultar el anejo final.

### ESTRUCTURA DEL FICHERO

Si se abre el fichero Excel con la información sobre la estructura de los datos, encontramos dos tipos de registros; es decir, la información que se recoge del cuestionario se graba de dos formas diferentes, aunque no mucho. Por un lado están los registros para personas de 16 años o más, es decir, aquellas que legalmente pueden estar ocupadas o paradas. Por otro lado, los registros para personas menores de esa edad. La principal diferencia entre los dos registros es la cantidad de variables, mucho menor en el segundo caso. Esto hace que haya mucho espacio en blanco en el registro para menores. Se puede hacer frente a este diseño de registros diferentes de diferentes formas, aquí nos centraremos en la siguiente:

Procedemos a tratar ambos tipos de registros, con sus especificidades, de forma conjunta. Nos centraremos en este procedimiento, pues es el más sencillo cuando se sabe trabajar con ficheros no jerárquicos, un procedimiento que no suele aparecer en los principales manuales no oficiales de SPSS ni es habitual enseñar en los cursos de iniciación. Este tipo de procedimientos se emplea en ficheros en los que la forma de recoger la información varía según el tipo de casos. Otro uso de este procedimiento sucede cuando cada caso tiene un número de registros distintos. Por ejemplo, si a una persona se le pregunta cuantas veces se ha divorciado y luego se le pregunta por la información de cada una de sus parejas, tomando para cada una de ellas un registro, obviamente habrá casos en los que no haya información, casos en los que haya información para una persona o para más; cada conjunto de variables se puede recoger en un registro distinto, de forma que haya observaciones con un solo registro, con dos o más. Para este tipo de contingencias el procedimiento de lectura de SPSS difiere del habitual. Junto al comando `DATA LIST` debe emplearse `FILE TYPE`, tal y como puede encontrarse en el manual `BASE SYSTEM` de SPSS, anejo C.

La sintaxis del comando aparece en el siguiente recuadro (en cursivas los parámetros a personalizar en cada caso; todos los comandos de este artículo pueden encontrarse en <http://webpages.ull.es/users/josamaga/lectura-epa-2005-ME.sps>; si no está familiarizado con el trabajo en ventanas de sintaxis, vea el anejo).

```
FILE TYPE MIXED FILE="DIRECTORIO Y NOMBRE DEL FICHERO" RECO-
RD=NOMBRE DE VARIABLE QUE INDICA EL TIPO DE REGISTRO COLUMN(S) DE LA
VARIABLE.
```

```
*[DONDE X ES UN NÚMERO QUE IDENTIFICA EL TIPO DE REGISTRO].
```

```
DATA LIST RECORD=X
/ VAR1 POSICIÓN VAR2 POSICIÓN... .
```

```
DATA LIST RECORD=X
/ VAR1 POSICIÓN VAR2 POSICIÓN... .
```

```
*[SE REPITEN TANTOS PROCEDIMIENTOS DATA LIST COMO REGISTROS DISTINTOS
```

TIENE EL FICHERO. LOS NOMBRES DE LAS VARIABLES EN CADA PROCEDIMIENTO PUEDEN SER IGUALES O DISTINTOS, DEPENDIENDO DE CÓMO ESTÉN ESTRUCTURADOS LOS DATOS].

END FILE TYPE.

Teniendo en cuenta que la información sobre el tipo de registro está en la variable *NIVEL*, situada en la columna 13. Si el valor es 1, el registro es para personas con 16 años o más, y si es 2, es para menores de 16 años.

FILE TYPE MIXED FILE='C:\EPA\EPA.T0105' RECORD=NIVEL 13.

\*\*\*INDICAMOS DÓNDE ESTÁ EL FICHERO, Y LE AVISAMOS QUE SEGÚN LA INFORMACIÓN DISPONIBLE EN LA VARIABLE NIVEL, QUE ESTÁ EN LA COLUMNA 13, EL DISEÑO DEL REGISTRO ES DISTINTO\*\*\*.

RECORD TYPE 1.

\*\*CON ESTA INSTRUCCIÓN AVISAMOS AL SPSS QUE LEA AQUELLOS REGISTROS EN LOS QUE LA VARIABLE NIVEL TOMA EL VALOR 1 (POBLACIÓN CON 16 AÑOS O MÁS\*\*\*.

DATA LIST

/CICLO	1	-	3
CCAA	4	-	5
PROV	6	-	7
NVIVI	8	-	12
NIVEL	13		
NPERS	14	-	15
EDAD5	16	-	17
RELPP1	18		
SEXO1	19		
NCONY	20	-	21
NPADRE	22	-	23
NMADRE	24	-	25
RELLMILI	26		
ECIV1	27		
PRONA1	28	-	29
REGNA1	30	-	32
NAC1	33		
EXREGNA1	34	-	36
ANORE1	37	-	38
NFORMA	39	-	40
EDADES	43	-	45
CURSR	46		
NCURSR	47	-	48
CURSNR	49		
NCURNR	50	-	51
HCURNR	52	-	54
TRAREM	57		
AYUDFA	58		
AUSENT	59		
RZNOTB	60	-	61
VINCUL	62	-	63
NUEVEM	64		



## Explotación de la Encuesta de Población Activa

OCUP1	65		
ACT1	66	-	67
SITU	68		
SP	69		
DUCON1	70		
DUCON2	71		
DUCON3	72	-	73
TCONTM	74	-	75
TCONTD	76	-	77
DREN	78	-	80
DCOM	81	-	83
PROEST	84	-	85
REGEST	86	-	88
PARCO1	89		
PARCO2	90	-	91
HORASP	92	-	95 (2)
HORASH	96	-	99 (2)
HORASE	100	-	103 (2)
EXTRA	104		
EXTPAG	105	-	108 (2)
EXTNPG	109	-	112 (2)
RZDIFH	113	-	114
TRAPLU	115		
OCUPLU1	116		
ACTPLU1	117		
SITPLU	118	-	119
HORPLU	120	-	123 (2)
MASHOR	124		
DISMAS	125		
RZNDISH	126	-	127
HORDES	128	-	129
BUSOTR	130		
BUSCA	131		
DESEA	132		
FOBACT	133		
NBUSCA	134	-	135
ASALA	136		
EMBUS	137		
ITBU	138	-	139
DISP	140		
RZNDIS	141		
EMPANT	142		
DTANT	143	-	145
OCUPA	146		
ACTA	147		
SITUA	148	-	149
OFEMP	150		
SIDI1	151	-	152
SIDI2	153	-	154
SIDI3	155	-	156
SIDAC1	157		
SIDAC2	158		
MUN1	159		
PRORE1	160	-	161
REPAIRE1	162	-	164
TRAANT	165		
AOI	166	-	167
CSE	168	-	169
FACTOREL	170	-	176 (2).

\*\*\*EL DOS ENTRE PARÉNTESIS INDICA QUE LOS DOS ÚLTIMOS DÍGITOS DE LA VARIABLE SON DECIMALES; EN LAS VARIABLES QUE INDICAN TIEMPO LOS DECIMALES SE REFIEREN A MINUTOS.

RECORD TYPE 2.

\*\*\*IDEM QUE ANTERIOR, PERO AHORA CON EL VALOR 2 PARA LA VARIABLE NIVEL, ES DECIR, POBLACIÓN DE 16 AÑOS O MENOS\*\*\*.

DATA LIST

/CICLO	1	-	3
CCAA	4	-	5
PROV	6	-	7
NVIVI	8	-	12
NIVEL	13		
NPERS	14	-	15
EDAD2	16	-	17
RELPP2	18		
SEXO2	19		
NCONY	20	-	21
NPADRE	22	-	23
NMADRE	24	-	25
PRONA2	28	-	29
REGNA2	30	-	32
NAC2	33		
EXREGNA2	34	-	36
ANORE2	37	-	38
MUN2	159		
PRORE2	160	-	161
REPAIRE2	162	-	164
FACTOREL	170	-	176 (2).

END FILE TYPE.

\*\*\*SEÑALAMOS QUE FINALIZA LA LECTURA POR TIPO DE REGISTROS\*\*\*.

EXECUTE.

SAVE OUTFILE="C:\EPA\T0105.SAV".

\*GUARDAMOS EL FICHERO EN FORMATO SPSS\*.

Tras ejecutar este procedimiento correctamente, debemos observar tres ficheros en el directorio C:\EPA\, el de los datos comprimidos, los datos descomprimidos y los datos en formato SPSS (con extensión .sav). Este mismo fichero se puede aplicar a las EPA's de 1999 a 2004, teniendo en cuenta que éstas hay menos variables, y por tanto, algunas resultarán con todas las observaciones perdidas (debe comparar el fichero de registro de estas encuestas con el de las de 2005 y años posteriores).

LA EPA ES UNA ENCUESTA PONDERADA

El diseño muestral de la EPA está pensado para que la representatividad estadística sea provincial en algunas variables. Para hacer esto posible la muestra no es proporcional, pues es necesario alcanzar un mínimo de casos en cada provincia, independientemente de su población. Por ello, la muestra siempre debe explotarse ponderando el fichero. Esto se consigue con las variable `FACTOREL`<sup>2</sup>; si se trabaja desde el fichero de sintaxis, la orden es la que sigue:

```
WEI FACTOREL.
```

Con esta nueva variable la muestra está ponderada y elevada, es decir, al calcular las frecuencias absolutas se tiene en cuenta que cada individuo muestral cuenta como varios individuos poblacionales, de forma que el total de la muestra pasa de las 155.597 personas sobre las que se ha recogido información a 42.873.967 cuando se eleva la muestra. Si está interesado en conocer, pongamos por caso, el número de parados de cierta región, o con cierto nivel de estudios, tendrá que elevar la muestra. Pero para cualquier procedimiento en el que sea necesario emplear estadísticos inferenciales, el factor de elevación no permite un buen uso, pues para calcularlos es necesario el tamaño muestral, y no el poblacional. Un ejemplo sería comprobar si la tasa estimada de paro de dos provincias difiere significativamente en términos estadísticos. Para poder realizar este contraste debemos hacer uso del factor de ponderación. Este factor reequilibra la mues-

---

<sup>2</sup> En el formato en el que el INE vendía estas encuestas, antes de junio de 2005, para elaborar el factor de elevación era necesario calcular una nueva variable multiplicando dos que se facilitaban, tal como sigue:

```
COMPUTE FACTOREL=FACELE*FACREP.
```

Esta descomposición permitía emplear `FACELE` para elevar la muestra de viviendas. Tal y como se facilita actualmente la información, no parece que elevar las viviendas sea posible.

tra, teniendo en cuenta que el diseño no es proporcional, pero de forma que los totales coinciden con el número total de casos sin elevar.

Podemos hallar este factor teniendo en cuenta la relación que hay entre pesos muestrales y pesos poblacionales. Si usamos minúsculas para la información muestral y las mayúsculas para la poblacional, siendo  $w_i$  la cantidad de entrevistados en el estrato muestral,  $w$  el tamaño de la muestra,  $W_i$  la cantidad de personas en el estrato poblacional y  $W$  el tamaño de la población, se mantiene la siguiente proporción entre el tamaño de cada estrato y el tamaño de la población:

$$\frac{w_i}{w} = \frac{W_i}{W}$$

Despejando, obtenemos la información para la ponderación, tal que:

$$w_i = \frac{W_i}{W} w$$

En el caso de la EPA que estamos tratando, debemos tomar nota del tamaño de la muestra antes y después de elevarla, operación que puede realizarse con cualquier variable. Tomemos la variable AOI, una de las claves de esta encuesta, pues contabiliza el número oficial de de ocupados parados e inactivos del INE, aunque obviamente se puede tomar cualquier otra variable.

```
WEI OFF.
*[EL COMANDO ANTERIOR ES PARA ASEGURARNOS QUE NO ESTÁ ACTIVO NINGÚN
PESO].
FRE AOI.
WEI FACTOREL.
FRE AOI.
```

Con los totales de la muestra sin ponderar ( $w$ ) y de la muestra elevada ( $W$ ), podemos calcular el factor con los comandos de SPSS como sigue:

```
COMPUTE PONDERA=FACTOREL*155997/42873967.
VAR LAB PONDERA 'FACTOR DE PONDERACIÓN MUESTRAL'.
WEI PONDERA.
FRE AOI.
```

En la Tabla 1 tenemos los resultados de la muestra, en frecuencias absolutas y en porcentajes, tal cual es (columna A), elevada (columna B) y ponderada (columna C). La información de la columna A nos indica el número de personas realmente entrevistado que responde a cada categoría. La columna B *eleva* esas frecuencias para permitirnos realizar una estimación del número de personas clasificadas en cada valor para toda la población. Y la columna C tiene en cuenta la proporción de personas que habría respondido a cada categoría si las probabilidades de cada persona de entrar en la muestra hubiesen sido las mismas, en vez de variar dependiendo de la provincia en la que vivan.

**Tabla 1**

AOI Clasificación de los entrevistados

		A) MUESTRA		B) MUESTRA ELEVADA		C) MUESTRA PONDERADA	
		Frecuencia	%	Frecuencia	%	Frecuencia	%
3	Ocupados subempleados por insuficiencia de horas	4975	3,2	1592462	3,7	5794	3,7
4	Resto de ocupados	57852	37,1	16900196	39,4	61491	39,4
5	Parados que buscan primer empleo	930	0,6	257124	0,6	936	0,6
6	Parados que han trabajado antes	6293	4,0	1841891	4,3	6702	4,3
7	Inactivos 1 -desanimados-	1124	0,7	264871	0,6	964	0,6
8	Inactivos 2 -junto con los desanimados forman los activosa p	1557	1,0	421819	1,0	1535	1,0
9	Inactivos 3 -resto de inactivos-	58899	37,8	14909227	34,8	54247	34,8
Válidos	Total	131630	84,4	36187589	84,4	131669	84,4
Perdidos	Sistema	24367	15,6	6686379	15,6	24328	15,6
<b>Total</b>		<b>155997</b>		<b>42873967</b>		<b>155997</b>	
			100		100		100

FUENTE: ENCUESTA DE POBLACIÓN ACTIVA, I TRIMESTRE 2005. INE

Es importante resaltar algunos datos de esta tabla. Primero, en los totales apreciamos que el tamaño de la muestra elevada coincide con el número de residentes en España, según datos del diseño de esta encuesta, mientras que el total de la muestra ponderada

coincide con el número de personas entrevistadas. Por otro lado, los porcentajes de la muestra elevada y ponderada coinciden. Por ello es por lo que la muestra siempre debe trabajarse “pesada”, ya sea por el factor de elevación que facilita el INE, FACELE, ya sea por el factor de ponderación que hemos construido, PONDERA. Como ya hemos señalado, para procedimientos estadísticos inferenciales, la muestra debe estar ponderada por PONDERA (columna C), mientras que en los procedimientos de estadística descriptiva se puede emplear ELEVA (columna B), pero nunca debe trabajarse con los datos sin ponderar (columna A).

### LA EPA ES UNA ENCUESTA A VIVIENDAS

El procedimiento de lectura que acabamos de apuntar solo es útil para leer los datos de forma que cada observación es una persona. Pero la encuesta es a viviendas, posibilidad poco explotada (Garrido, Requena y Toharia, 2000), debido tanto que los resultados que publica el INE sobre esta cuestión son escasos, como debido, a que, como veremos en este apartado, la explotación es más compleja. Para explotar la EPA como encuesta de viviendas es posible seguir al menos dos procedimientos. Uno, leerla directamente como fichero de viviendas. Otro, leerla como fichero de individuos, y luego reestructurarla. En el primer caso habría que proceder con la lectura de ficheros no jerárquicos, de forma un poco más compleja de la mostrada en el apartado anterior, pues habría que diseñar tantos tipos de registro como número máximo de personas haya en la vivienda de la encuesta en la que residen más personas. El segundo caso resulta más sencillo y es el que mostraremos aquí. Consiste en reestructurar el fichero de individuos, posible gracias a la información disponible en la variable NVIVI, el número de identificación de cada vivienda, es decir, a todos los residentes en la misma vivienda se les asigna el

mismo valor en esta variable. Con la nueva reestructuración, todas las personas de cada vivienda forman un solo caso, y hay tantas variables como variables individuales multiplicadas por el número máximo de individuos por viviendas, descontadas las variables que son iguales para todos los residentes en la vivienda, como, por ejemplo, la provincia de residencia. En el proceso de reestructuración SPSS (posible a partir de la versión 12.0) añade un punto y un número al nombre de cada variable, de forma que, si en el fichero de individuos hay una variable que se llama EDAD5, con la edad de la persona, en el fichero de viviendas tenemos las variables EDAD5 . 1, EDAD5 . 2, EDAD5 . 3, ..., EDAD5 . 12 (pues el máximo de personas que refleja la encuesta en una vivienda es de 12). De esta forma sabemos cuál es la edad de cada una de las personas que comparten la vivienda.

```
*ANTES DE PROCEDER A REESTRUCTURAR EL FICHERO, GUARDAMOS LOS DATOS.  
SAVE OUTFILE="C:\EPA\EPA105.SAV".
```

```
*PRIMERO ES NECESARIO ORDENAR EL FICHERO POR LA  
VARIABLE QUE VAMOS A REESTRUCTURAR, NVIVI, QUE  
ES LA IDENTIFICACIÓN DE LA VIVIENDA.
```

```
SORT CASES BY NVIVI.
```

```
**CON EL SIGUIENTE COMANDO AVISAMOS QUE HABRÁ CASOS QUE PASARÁN A SER  
NUEVAS VARIABLES, EN FUNCIÓN DE LA VARIABLE DE IDENTIFICACIÓN DE LA  
VIVIENDA, NVIVI  
**DEFINIMOS UNA NUEVA VARIABLE, CANPER, PARA SABER CUÁNTAS PERSONAS  
HAY EN LA VIVIENDA***.
```

```
CASESTOVARS  
/ID = NVIVI  
/GROUPBY = VARIABLE  
/COUNT = CANPER "CANTIDAD DE PERSONAS QUE RESIDEN EN LA VIVIENDA".
```

El resultado que produce este procedimiento<sup>3</sup> (Resultado 1) nos informa sobre las transformaciones experimentadas por el fichero de datos. Por un lado, nos informa de las nuevas variables generadas a partir del fichero individual. Indica la variable original en la primera columna, luego un número que señala las veces que esa variable se ha ido generando en el nuevo fichero, y en la última columna, el nombre de las variables nuevas que se han generado. Esta operación se repite para todas las variables del fichero individual con valores distintos para al menos dos personas, pero aquí solo presentamos la información para las tres primeras variables, suficiente para familiarizarse con el procedimiento, y evitamos el largo listado de salida para todas las variables del fichero individual.

**Resultado 1 Nomenclatura de tres variables del fichero de viviendas a partir del fichero de individuos**

VARIABLES GENERADAS

Variable original		Nombre
NIVEL	1	NIVEL.1
	2	NIVEL.2
	3	NIVEL.3
	4	NIVEL.4
	5	NIVEL.5
	6	NIVEL.6
	7	NIVEL.7
	8	NIVEL.8
	9	NIVEL.9

---

<sup>3</sup> Se puede realizar a través del menú, eligiendo Datos, Reestructurar y en el cuadro de diálogo que se abre, seleccionar “Reestructura los casos seleccionados en variables”, y luego seguir las opciones del cuadro de diálogo, teniendo en cuenta que la variable de identificación es NVIVI.



	10	NIVEL.10
	11	NIVEL.11
	12	NIVEL.12
NPERS	1	NPERS.1
	2	NPERS.2
	3	NPERS.3
	4	NPERS.4
	5	NPERS.5
	6	NPERS.6
	7	NPERS.7
	8	NPERS.8
	9	NPERS.9
	10	NPERS.10
	11	NPERS.11
	12	NPERS.12
EDAD5	1	EDAD5.1
	2	EDAD5.2
	3	EDAD5.3
	4	EDAD5.4
	5	EDAD5.5
	6	EDAD5.6
	7	EDAD5.7
	8	EDAD5.8
	9	EDAD5.9
	10	EDAD5.10
	11	EDAD5.11
	12	EDAD5.12

El otro resultado (Resultado 2) nos indica la transformación del número de casos. Los casos de entrada son los individuos que componen el fichero original, mientras que los de salida son el número de casos del fichero reestructurado, es decir, el número de viviendas en las que se ha recabado información. La relación entre casos de entrada y de salida es de 2,8 (resultado de dividir el número de individuos entre el número de viviendas, es decir,  $155.997/54.955$ ). También nos informa sobre la cantidad original de variables, en el fichero de individuos y el nuevo número de variables, en el fichero de viviendas. Los “valores de índice” indican la cantidad máxima de personas que presenta el

fichero en una vivienda, 12 en este caso. Para cada variable que presenta valores distintos para los diferentes moradores, se generan nuevas variables. Si las variables originales toman los mismos valores para todos los residentes en la vivienda, no se construye una nueva variable, como sucede con las que aparecen al comienzo del nuevo fichero: NVIVI (identificación de la vivienda), CICLO (identificación del trimestre y año en que se ha realizado la encuesta), CCAA (comunidad autónoma de residencia), PROV (provincia de residencia), RELLMILI (espacio en blanco), REGEST (lugar donde está ubicada la persona), MUN2 (municipio donde reside la persona), REPAIRE2 (región de residencia en país extranjero en el periodo anterior). El número total de variables del fichero reestructurado a partir de las 103 variables originales será el de estas 8 variables más la 95 restantes multiplicadas por 12, en total 1.148 variables.

**Resultado 2 Información sobre la reestructuración del fichero de individuos en fichero de viviendas**

**Estadísticos de procesamiento**

Casos de entrada	155997
Casos de salida	54955
Casos de entrada/Casos de salida	2,8
Variables de entrada	103
Variables de salida	1148
Valores de índice	12

Trabajar con los ficheros de vivienda es más complejo que con los ficheros individuales, pues normalmente para hallar cualquier información debemos revisar las variables para cada individuo de la vivienda. Por ejemplo, si queremos conocer cuál es la relación laboral del cónyuge de la persona principal, debemos construir una nueva variable y

asignarle el valor correspondiente tras revisar la información de todos los miembros de la vivienda.

```
*AOI: VARIABLE SOBRE RELACIÓN CON EL MERCADO DE TRABAJO
*RELPP: VARIABLE DE RELACIÓN CON EL SUSTENTADOR PRINCIPAL,
CON EL VALOR 1 PARA PERSONA PRINCIPAL.
```

```
COMPUTE AOIP=0.
VAR LAB AOIP "RELACIÓN CON LA ACTIVIDAD DE LA PERSONA PRINCIPAL".
```

```
IF (RELPP.1=1) AOIC=AOI.1.
IF (RELPP.2=1) AOIC=AOI.2.
IF (RELPP.3=1) AOIC=AOI.3.
IF (RELPP.4=1) AOIC=AOI.4.
IF (RELPP.5=1) AOIC=AOI.5.
IF (RELPP.6=1) AOIC=AOI.6.
IF (RELPP.7=1) AOIC=AOI.7.
IF (RELPP.8=1) AOIC=AOI.8.
IF (RELPP.9=1) AOIC=AOI.9.
IF (RELPP.10=1) AOIC=AOI.10.
IF (RELPP.11=1) AOIC=AOI.11.
IF (RELPP.12=1) AOIC=AOI.12.
```

```
*RELPP: VARIABLE DE RELACIÓN CON EL SUSTENTADOR PRINCIPAL,
CON EL VALOR 2 PARA CÓNYUGE.
```

```
COMPUTE AOIC=0.
VAR LAB AOIC "RELACIÓN CON LA ACTIVIDAD DEL CÓNYUGE".
```

```
IF (RELPP.1=2) AOIC=AOI.1.
IF (RELPP.2=2) AOIC=AOI.2.
IF (RELPP.3=2) AOIC=AOI.3.
IF (RELPP.4=2) AOIC=AOI.4.
IF (RELPP.5=2) AOIC=AOI.5.
IF (RELPP.6=2) AOIC=AOI.6.
IF (RELPP.7=2) AOIC=AOI.7.
IF (RELPP.8=2) AOIC=AOI.8.
IF (RELPP.9=2) AOIC=AOI.9.
IF (RELPP.10=2) AOIC=AOI.10.
IF (RELPP.11=2) AOIC=AOI.11.
IF (RELPP.12=2) AOIC=AOI.12.
```

Tras estas operaciones podríamos proceder a cruzar las dos nuevas variables, AOIP y AOIC, para saber, por ejemplo, en cuantos hogares el sustentador principal y su cónyuge están en paro, u ocupados, o cualquier otro tipo de relación.

Una vez que tenemos variables definidas en el fichero de viviendas podemos agregarlas al fichero de individuos, y así sabemos cuáles son las características de los hogares de los entrevistados. En primer lugar procederemos a salvar el nuevo fichero de viviendas,

y luego recuperar el fichero de individuos; de esta forma, las variables de los individuos serán las primeras que aparezcan en el nuevo fichero, lo que puede simplificar posteriores trabajos. Luego procedemos a fundir los dos ficheros, mediante el procedimiento de añadir variables de un *fichero de claves*, es decir, un fichero con información común para varias observaciones de otro fichero. En este caso el fichero de claves es el fichero de viviendas y la variable de identificación es NVIVI.

```
*PASAMOS A GUARDAR EL NUEVO FICHERO DE VIVIENDAS.  
SAVE OUTFILE="C:\EPA\V-EPA105.SAV".
```

```
*RECUPERAMOS EL FICHEROS DE INDIVIDUOS.  
GET FILE="C:\EPA\EPA105.SAV".
```

```
*AMBOS FICHEROS DEBEN ESTAR ORDENADOS POR  
LA MISMA VARIABLE.  
SORT BY NVIVI.
```

```
*EL COMANDO SIGUIENTE UNE FICHERO DE INDIVIDUOS CON LA INFORMACIÓN DE  
SUS VIVIENDAS. EN ESTA FUSIÓN SE ELIMINAN LAS VARIABLES CON EL MISMO  
NOMBRE, Y POR TANTO, LA MISMA INFORMACIÓN, EN AMBOS FICHEROS (SUBCO-  
MANDO DROP).
```

```
MATCH FILES /FILE=*  
/TABLE='D:\DATOS\EPA\DATOS\F105.SAV'  
/RENAME (CCAA CICLO MUN2 PROV REGEST RELLMILI REPAIRE2 = D0 D1 D2 D3  
D4 D5 D6)  
/BY NVIVI  
/DROP= D0 D1 D2 D3 D4 D5 D6.  
EXECUTE.
```

Un problema de la EPA de distribución gratuita es que no facilita el factor de elevación para la muestra de viviendas, a diferencia de cómo venía haciendo el INE hasta ahora. Por tanto, de momento solo se puede explotar esta información si fusionamos el fichero de hogares al de individuos, es decir, en vez de estudiar los hogares como tales, solo los podemos estudiar como características de los individuos<sup>4</sup>.

---

<sup>4</sup> Si se realiza una explotación de tipo inferencial con estos datos, téngase en cuenta que el supuesto habitual de independencia de las observaciones no se cumpliría, por formar los individuos parte del mismo hogar. De todas formas, dado el tamaño de la muestra y lo pequeñas que son las familias en com-

## LA EPA EN PERSPECTIVA HISTÓRICA

La EPA comenzó en 1964 y desde entonces son varios los cambios producidos, tanto el tipo de variables recogidas como en la operacionalización de las mismas y en el trabajo de campo. Algunas de estas transformaciones son el resultado de adaptar los datos españoles a directrices de organismos como la Organización Internacional del Trabajo o de Eurostat. Aquí comentamos brevemente algunos de estos cambios. Los principales se produjeron en 1976, 1987, 1999, 2002 y 2005. De cara a la explotación de los datos, la relevancia de esta cuestión es obvia: en los periodos en los que la encuesta se ha mantenido constante se les puede aplicar el mismo fichero de lectura de datos, con el único cambio de especificar correctamente la carpeta del fichero a leer y su nombre. Por ejemplo, desde 1976 hasta 1986 los ficheros permanecen con el mismo diseño. En 1987 se introducen algunas variables nuevas, así como nuevos valores para las ya existentes y se modifica la definición de parado. En 1999 se procede otra reestructuración profunda de la encuesta, añadiendo variables, cambiando el nombre a variables anteriores y añadiendo más valores a algunas variables, así como a cambios importantes en el diseño del trabajo de campo. En 2002 se modifica nuevamente la definición de parado, así como se calculan los factores de elevación poblacional a partir de nuevas proyecciones del censo de 1991. En 2005 se procede a una nueva adaptación de la encuesta a ciertas normas internacionales, se toman en cuenta los datos del censo de 2001 y se mejora notablemente el trabajo de campo (INE 2005), que estaba produciendo ciertos sesgos (G<sup>a</sup> Serrano, Garrido y Toharia, 1999). El INE ha propuesto una homogenización en la variable de *relación con la actividad* (parado, ocupado, inactivo), según la clasificación de

---

paración, violar este supuesto no tendría por qué producir problemas.

paro de 2002 (Trejo y Ortega, 2005), así como se explica el paso entre la codificación distinta de las mismas variables en sucesivos periodos (INE, 2005). También, se pueden encontrar procedimientos para trabajar con *EPAs* de distintos periodos tal y como las venía ofreciendo el INE hasta ahora en Fernández Enguita, Las Heras Pérez y Martínez García (1996a), así como comentarios sobre las modificaciones en la Clasificación Nacional de Ocupaciones (Fernández Enguita, Las Heras Pérez y Martínez García, 1996), realizada en el segundo trimestre de 1994 y en la Clasificación Nacional de Actividades, realizada en el primer trimestre de 1993 (Fernández Enguita, Martínez García y Las Heras Pérez, 1996b). A partir de estos estudios cabría hacer otro que tuviese en cuenta las profundas modificaciones de la EPA en 1999, tarea que sale de los límites de este artículo.

Pero sí podemos comentar de momento algunas desagradables sorpresas. Por ejemplo, en los datos que el INE venía ofreciendo hasta ahora, se facilitaba la edad de las personas desagregada, mientras que ahora se proporciona en grupos quinquenales. Para muchos estudios esto puede ser irrelevante, pero no para quienes nos dedicamos a investigar cuestiones educativas, pues entre los 16 y 20 años los jóvenes pueden estar en la Educación Secundaria Obligatoria, cursando estudios post-obligatorios secundarios o en la Universidad. No poder distinguir en estas edades arruina el estudio sobre decisiones educativas. Otro cambio a peor consiste en que la Clasificación Nacional de Ocupaciones está a un dígito, lo que imposibilita estudiar con detalle la evolución de las ocupaciones en España. Es cierto que algunas ocupaciones no alcanzan el número suficiente de casos en algunas submuestras regionales como para ser representativas, y que el INE debe velar por el rigor de los análisis que se hace con sus datos. Pero también deben

confiar en el buen juicio de quienes los explotamos, así como en el autocontrol de la comunidad investigadora. Y por último, tal como señalábamos anteriormente, no se facilita el factor de elevación de las viviendas, con lo que dificulta el estudio de los hogares.

### OTRAS POSIBILIDADES DE LA EPA

Entre las posibilidades de la EPA nos gustaría señalar cómo explotar submuestras demasiado pequeñas, la EPA como encuesta de panel y los módulos de los segundos trimestres. Empecemos por las submuestras, que ya señalábamos en el apartado anterior. Cuando deseamos explotar una submuestra de la EPA corremos el riesgo de disponer de pocos casos, que dificultan el análisis inferencial, en tanto que los márgenes de error de los resultados pueden llegar a ser tan grandes que es difícil encontrar diferencias estadísticamente significativas. Supongamos que deseamos averiguar las tasas de paro por nivel de estudios y provincia. Podríamos tener tan pocos casos en algunas provincias que los resultados no serían significativos. Una forma de hacer frente a este problema es agregar varios trimestres de la encuesta, en lo que se conoce como *roll-sample* (Kish, 1998). La recogida de la información de la EPA está diseñada de modo que el número de entrevistas se distribuye uniformemente a lo largo del trimestre, de manera que los resultados de la encuesta deben interpretarse como válidos para la mitad del trimestre. Por ejemplo, los datos de paro de la EPA del primer trimestre se refieren al paro a mitad del trimestre, es decir el 14 de febrero, y no al 1 de enero ni al 31 de marzo. Si decidimos agregar dos trimestres, el resultado sería válido para el día equidistante entre el día medio del primer trimestre y del segundo trimestre, es decir, el 1 de abril, y así sucesivamente. De esta forma, cuanto más pequeña sea la submuestra que deseamos investi-

gar, mayor será el periodo de referencia que debemos abarcar. El inconveniente de este procedimiento es que puede resultar un tanto inestable para estudiar variables de tipo más coyuntural que dependan del ciclo económico, como por ejemplo horas extraordinarias, pues, cuanto mayor sea el periodo de estudio, mayor será la heterogeneidad de la muestra, así que lo por un lado ganamos en mayor tamaño de muestra, por otro lo perdemos en mayor variabilidad de la misma. Pero en variables de tipo sociodemográfico más estable, como nivel de estudios, estado civil, etc. este problema sería menor.

Otra posibilidad de la EPA es explotarla en tanto que panel de viviendas. La EPA se renueva en un sexto cada trimestre, es decir, una vivienda que es seleccionada en un trimestre es vuelta a entrevistar durante 5 trimestres más. Otra forma de decirlo es que cada seis trimestres, la muestra se renueva por completo. De momento no está disponible entre los datos que se ofertan gratuitamente, pero también es posible explotar este panel de hogares, a través de un fichero de enlace que se puede solicitar como petición a medida al INE. De esta forma se pueden estudiar variaciones en las mismas observaciones a lo largo de un periodo de tiempo, como por ejemplo, cuánto tarda un parado en encontrar trabajo o durante cuánto tiempo permanecen todos los miembros activos de una familia en el paro (con el periodo de observación siempre limitado a seis trimestres). El inconveniente de esta explotación es que la muestra se hace con viviendas, no con personas o familias. Esto quiere decir que si durante el periodo de observación, la familia se muda de vivienda y entra otra familia, se pasa a entrevistar a la nueva familia, y se pierde la pista de la familia saliente. Por ello es necesario tener cuidado con este tipo de explotación, para no producir sesgos o al menos calcular el tamaño de los posibles sesgos.



Por último, cabe mencionar los módulos especiales de los segundos trimestres de la EPA<sup>5</sup>. Debido a convenios internacionales, el INE, desde 1999, añade un módulo de preguntas *ad hoc* a la encuesta, sobre un tema que varía en cada ocasión. Para leer este módulo *ad hoc* sería necesario contar con el libro de códigos pertinente, y realizar los añadidos necesarios en el diseño de registros que hemos mostrado en este artículo. Los temas tratados hasta el 2004 son los siguientes:

- Módulo año 1999: Accidentes y enfermedades laborales
- Módulo año 2000: Transición de la educación al mercado laboral
- Módulo año 2001: Relaciones laborales especiales y de condiciones y horarios de trabajo
- Módulo año 2002: Personas con discapacidad y su relación con el empleo
- Módulo año 2003: Módulo de cursos de educación/formación recibidos en los últimos doce meses
- Módulo año 2004: Tiempo de trabajo

## OTRAS ENCUESTAS

Los procedimientos que hemos explicado aquí son similares a los necesarios para explotar las otras encuestas que el INE ha puesto gratuitamente a disposición del público. En

---

<sup>5</sup> El segundo trimestre tiene un valor especial, pues se supone que es el más estable a variaciones coyunturales: es el que está menos sujeto a vacaciones, incorporación de recién licenciados al mercado de trabajo, comportamientos estacionales, como las vacaciones de verano o picos de trabajo como las rebajas de enero o de verano. Es por esta estabilidad por lo que es recomendable trabajar con los segundos trimestres, cuando no se puede hacer con todos los trimestres, o no hay un interés especial para hacerlo.

todas ellas habría que seguir un procedimiento de lectura de los datos en ASCII a partir de la información de los registros que vienen en formato Excel. Si las encuestas son a viviendas, como ocurre con las Encuestas de Presupuestos Familiares, habría que proceder a la reestructuración del fichero de individuos, como aquí se ha mostrado. Y en el caso de los Censos, si se quiere disponer de una muestra del censo para toda España, habría que descargar cada provincia, y luego fusionarlas todas, con el procedimiento de añadir casos de SPSS. Todas aquellas encuestas que están ponderadas deben explotarse con sus pesos, ya sea con el factor de elevación, que facilita el INE, ya sea calculando el factor de ponderación, tal y como hemos mostrado previamente.

### CONCLUSIONES

Hemos presentado cómo leer la Encuesta de Población Activa, mediante los comandos de sintaxis de SPSS. La ventaja de emplear la sintaxis es que permite automatizar procedimientos repetitivos, con una gran ganancia de tiempo sobre el uso de ventanas, así como permite la detección y corrección de errores más fácilmente. Hemos mostrado algunos procedimientos que habitualmente no se enseñan en los cursos de iniciación al SPSS, como son la lectura de ficheros no jerárquicos o la reestructuración de variables que están en casos distintos a variables en un solo caso. Esto último permite la explotación por hogares de la encuesta, más complicada que la explotación por individuos, pero de mucha más relevancia sociológica, pues nos permite estudiar la situación de una persona en el contexto de su familia. También hemos presentado la forma de calcular los pesos de la muestra, para poder aplicar procedimientos inferenciales, pues el peso que facilita el INE solo permite procedimientos descriptivos.

Con toda esta información esperamos que resulte más sencillo disponer de la gran cantidad de información que el INE acaba de facilitar gratuitamente a todo el mundo. No podemos más que congratularnos por esta iniciativa, y esperar que la secunden otros centros de producción de datos, como los institutos autonómicos de estadística o el Centro de Investigaciones Sociológicas. Por último, señalar que, por desgracia, el INE ha limitado el nivel de detalle con el que estaba distribuyendo los datos, sin que sepamos muy bien los motivos, imposibilitando líneas de investigación que eran posibles hasta hace poco.

#### REFERENCIAS

- INE (2005). *Metodología de la EPA* URL: <http://www.ine.es/epa02/meto2002.htm> y en <http://www.ine.es/epa02/ultima%20nota%20metodologica.pdf> , <http://www.ine.es/daco/daco43/notaepa.htm?L=0>
- Fernández Enguita, Mariano, Soledad Las Heras Pérez y José S. Martínez García 1996. "Reagregación de la Clasificación Nacional de Ocupaciones" en <http://webpages.ull.es/users/josamaga/Papers/CNO.pdf>.
- Fernández Enguita, Mariano, José S. Martínez García y Soledad Las Heras Pérez 1996a. "Reagregación de diversas variables de la Encuesta de Población Activa" en <http://webpages.ull.es/users/josamaga/Papers/varias-var.pdf>.
- 1996b. "Reagregación de la Clasificación Nacional de Actividades" en <http://webpages.ull.es/users/josamaga/Papers/CNAE.pdf>.
- G<sup>a</sup> Serrano, Carlos , Luis Garrido y Luis Toharia. 1999. "Empleo y paro en España. Algunas cuestiones candentes." Pp. 23-50 en *Las relaciones de empleo en España*, editado por F. Miguélez y C. Prieto. Madrid: Siglo XXI.
- Garrido, Luis, Miguel Requena y Luis Toharia. 2000. "La Encuesta de Población Activa desde la perspectiva de los hogares." *Estadística Española* 42:115-152.
- INE 2005. "EPA. RDE.- Resultados detallados (EPA-94) Comentarios sobre el enlace de la EPA" en <http://www.ine.es/daco/daco42/daco4211/notamet.htm?L=0>, 27 julio.
- Kish, Leslie. 1998. "Space/Time Variations and Rolling Samples." *Journal of Official Statistics* 14:31-46.
- Trejo, Javier y Lourdes Ortega. 2005. "Enlaces de las series de paro 1976-2000 según la definición EPA 2002." INE.

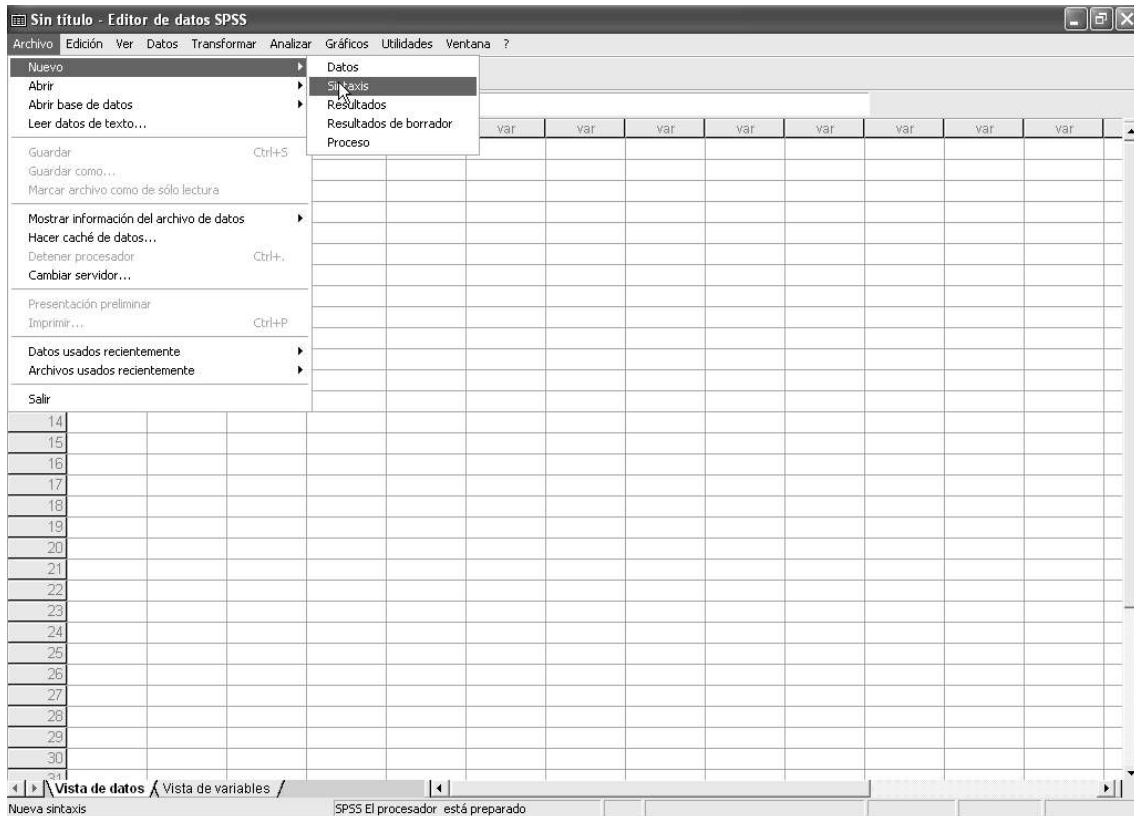
### ANEJO: EL USO DE SINTAXIS EN SPSS

Es habitual que quienes emplean SPSS de forma esporádica o solo lo conozcan a través de cursos de iniciación solo usen los procedimientos a través de los menús de ventanas. Aquí repasaremos brevemente cómo emplear los procedimientos de sintaxis, un poco más difíciles de manejar al principio, pero mucho más rápidos cuando se deben repetir tareas. Toda la información necesaria para los procedimientos de SPSS se encuentra excelentemente explicada en los manuales que se distribuyen con la aplicación. Si Vd. es usuario de este paquete y no encuentra los manuales, póngase en contacto con la persona responsable de SPSS en el lugar que lo tiene licenciado, pues estos manuales se distribuyen también en formato PDF, y es posible cargarlos en el ordenador en el momento de la instalación, a través de la instalación personalizada, o posteriormente, copiando los ficheros PDF del CD o directorio de instalación.

### LA VENTANA DE SINTAXIS

Dependiendo de cómo se tenga configurado SPSS, puede que se abra o no una ventana de sintaxis cuando empezamos a trabajar con él. Para comprobar si la ventana está abierta, en la barra de SPSS podemos picar con el botón izquierdo del ratón, y veremos si hay una ventana que se llame *Sintaxis1-Editor de sintaxis de SPSS*. Si es así, entramos en esta ventana, y en ella es donde iremos escribiendo los comandos de SPSS. Si no es así, hay que picar con el botón izquierdo del ratón en Archivo, luego posar el cursor en Nuevo, y se despliega un menú, una de cuyas opciones es Sintaxis, la cual debemos elegir.

## Pantalla A-1



Esto nos lleva a la ventana de edición, en la que podemos escribir normalmente, empleando comandos como copiar, marcar, cortar y pegar de la misma forma que en un procesador de textos tipo Word. Lo que escribamos en esta ventana serán órdenes de SPSS. Conviene resaltar las siguientes cuestiones sobre la sintaxis de SPSS:

- Cada orden de SPSS comienza con un comando, el cual se compone de diferentes opciones, y acaba con un punto o una línea en blanco que le separa del siguiente comando.
- Dentro del comando se pueden especificar a su vez subcomandos, cuyo comienzo se indica con el siguiente signo / , es decir, presionando mayúscula y el número 7. Entre el comando y sus subcomandos no puede haber ni puntos (a no ser en el nombre de variables o como señal de decimal, y en algunos casos, entre comillas) ni líneas en blanco.

-El tipo de letra que emplea por defecto SPSS en esta pantalla no es el mejor para escribir comandos, no solo porque las letras son un poco finas para trabajar desde la pantalla del ordenador, sino lo que es mucho peor, a veces cuesta diferenciar determinados caracteres que son muy distintos para el ordenador, pero no para el ojo humano, especialmente el 0 y la O. Por ello es recomendable cambiar el tipo de letra en esta pantalla, lo cual no varía para nada la forma en que se presenta la información en las pantallas de resultados. Para hacer este cambio, desde la propia pantalla de sintaxis, elegimos la opción VER del menú, que tras seleccionarla nos aparece la opción FUENTES. Picamos en fuentes y nos aparece un listado del tipo de fuentes, buscamos FIXEDSYS y la elegimos.

-Todo lo escrito en la ventana de sintaxis se puede grabar en un fichero (en formato ANSI). De esta forma, tenemos un conjunto de órdenes preparadas para ejecutar cada vez que queramos, sin necesidad de repetir todos los pasos a través de los distintos menús. Por ejemplo, en el caso de la lectura de datos que se presenta en este artículo, podemos salvar el fichero, y cada vez que descarguemos una nueva EPA del sitio del INE, abrimos el fichero, escribimos correctamente el nombre y la carpeta donde está el nuevo fichero de la EPA, y cambiamos el nombre con el que queremos guardar el fichero y lo ejecutamos. (Obviamente, primero tendríamos que asegurarnos que no ha habido variaciones en el diseño de los registros).

-Cada vez que desee ejecutar los comandos escritos en un fichero es suficiente con ir a *Archivo*, posar el cursor sobre *Abrir*, y en el menú que se despliega seleccionar *Sintaxis*. Se abrirá una ventana, por medio de la cual debemos seleccionar el fichero a ejecutar. Los comandos que mostramos en este artículo los puede descargar en su ordenador, y posteriormente abrir y ejecutar el fichero. Eso sí, debe asegurarse que los directorios y

nombres de ficheros coinciden con los de su ordenador. La dirección donde se encuentra este fichero es <http://webpages.ull.es/users/josamaga/lectura-epa-2005-ME.sps>; puede copiar todos los comandos que aparecen y pegarlos en la ventana de sintaxis de SPSS. Antes de ejecutarlo debe personalizar los directorios en los que se encuentran los ficheros.

-No es necesario comenzar sabiendo mucho de sintaxis. SPSS le permite trabajar con el menú de ventanas, y cuando ha terminado de definir todos los parámetros necesarios, siempre permite elegir entre “aceptar” y “pegar”. *Aceptar* ejecuta directamente el comando, mientras que *pegar, lo pega en la ventana de sintaxis*. Como ya hemos ido señalando, la ventaja que esto tiene es que una vez que hemos optado por un procedimiento, si lo pegamos, y queremos repetirlo, no es necesario volver a pasar por todas las ventanas y definir nuevamente los parámetros. Además, lo podemos salvar como fichero con extensión SPS y emplearlo en otra ocasión.

## EJECUCIÓN DE LA SINTAXIS

-Una vez escritos o pegados los comandos en la ventana de sintaxis es necesario ejecutarlos. Para ello debemos picar con el botón izquierdo del ratón sobre el botón de la barra de herramientas de SPSS con el icono ►. Tras esta operación, se ejecuta el comando sobre el que está activo el cursor. Si queremos ejecutar varios comandos, primero marcamos el texto con el ratón de la misma forma que se hace en la mayoría de los editores de texto: picamos con el botón izquierdo del ratón al comienzo del texto que queremos ejecutar, y sin soltar el botón, arrastramos el cursor hasta el final del texto. Una vez marcado el texto, que aparece resaltado, soltamos el botón izquierdo del ratón, y dirigimos el puntero hacia el botón de la barra de herramientas de SPSS con el icono ►, y

picamos una vez en este icono con el botón izquierdo del ratón. Se ejecutan los comandos y, dependiendo de cómo esté configurado el software, nos lleva a la ventana de resultados de SPSS. Si no, podemos llegar a esta ventana a través de la opción de *Ventanas*, en la barra superior de SPSS, la seleccionamos picando sobre ella con el botón izquierdo del ratón, se despliegan todas las ventanas abiertas de SPSS y seleccionamos la de resultados.

### Pantalla A-2

The screenshot shows the SPSS Syntax Editor window titled 'articulo-ME.sps - Editor de sintaxis SPSS'. The window contains a list of variables and their weights, followed by several SPSS commands. A mouse cursor is pointing to the 'Ejecutar comando actual' button in the toolbar.

Variable	Weight
NIVEL	13
NPERS	14
EDAD2	16
RELPP2	18
SEX02	19
NCONY	20
NPADRE	22
NMADRE	24
PRONA2	28
REGNA2	30
NAC2	33
EXREGNA2	34
ANDRE2	37
MUN2	159
PRORE2	160
REPAIRE2	162
FACTOREL	170

```

END FILE TYPE.
***SEÑALAMOS QUE FINALIZA LA LECTURA POR TIPO DE REGISTROS***.
EXECUTE.
*NOS ASEGURAMOS QUE LA MUESTRA NO TENGA NINGUNA PONDERACIÓN
CON EL SIGUIENTE COMANDO.
WEI OFF.
*ELEVAMOS LA MUESTRA.
WEI FACTOREL.
*PEDIMOS UNA FRECUENCIA PARA COMPROBAR EL FACTOR DE ELEVACIÓN
Y TOMAR NOTA DEL TOTAL DE CASOS.
FRE AOI.
*CÁLCULO DEL COEFICIENTE DE PONDERACIÓN.
COMPUTE PONDERA=FACTOREL*155997/42873967.
VAR LAB PONDERA 'FACTOR DE PONDERACIÓN MUESTRAL'.
WEI PONDERA.
FRE AOI.
SAVE OUTFILE='C:\EPA\DATOS\T0105.SAU'.
*GUARDAMOS EL FICHERO EN FORMATO SPSS, PONDERADO*.
    
```

The status bar at the bottom indicates 'Ejecutar comando actual' and 'SPSS El procesador está preparado'.



## SEGURIDAD

-Un consejo elemental es no hacer ninguna transformación sobre el fichero original de datos. Guárdelo en varios sitios, y realice su trabajo sobre copias del fichero, nunca sobre el original. Por ejemplo, en este caso el fichero original se llama EPA0105, pues copie el fichero y póngale el mismo nombre con la extensión EPA0105.DAT. Así, si en el proceso de datos comete algún error, siempre podrá volver a recuperar el fichero original.

-Guarde los ficheros de sintaxis de forma que si quiere llegar a un punto en el proceso de datos siempre sea fácil hacerlo. Así, si hay algún error, se puede repetir todo el proceso desde el origen, ejecutando los sucesivos archivos de trabajo, de forma que se pueda localizar y subsanar el error. Esto exige ser sistemático y ordenado en la grabación de los ficheros.

-Conviene añadir comentarios a cada paso, así como información general sobre lo que se pretende realizar con el fichero de comandos. Los comentarios siempre deben comenzar con un \* y finalizar con un punto o/y separados con una línea en blanco del siguiente comando.