

Proyecto Observatorio Económico Social
Guía para actualizar la base de datos del mercado laboral

Introducción

La base de datos del mercado laboral, denominada Base de Datos IIES, presenta estimaciones para una selección de variables de la Encuesta de Hogares por Muestreo – Instituto Nacional de Estadística (INE) desde el semestre primero de 1975. La misma es producto del procesamiento de los cuestionarios originales codificados y digitalizados, labor ejecutada por el Instituto de Investigaciones Económicas y Sociales (IIES) de la Facultad de Ciencias Económicas y Sociales de la Universidad Central de Venezuela.

El procesamiento de los cuestionarios es una tarea compleja, con elevado nivel de detalle y utilización de variadas herramientas computacionales que se aplican a la operación de grandes volúmenes de datos contenidos en diversos archivos y formatos.

Para hacer posible esta labor, en el presente documento se describen los archivos, programas y secuencia de acciones que deben observarse para el procesamiento de un nuevo semestre de la Encuesta de Hogares por Muestreo y su incorporación a la Base de Datos IIES.

I. Archivos originales de la Encuesta de Hogares-INE

Los archivos de la Encuesta de Hogares contienen los cuestionarios codificados correspondientes a cada semestre de la Encuesta de Hogares. Por lo general, dichos archivos son suministrados por el INE en formato DBF y desglosados en:

Datos sobre la vivienda	EHVIV2006I.dbf
Datos sobre los hogares	EHHOG2006I.dbf
Datos sobre las personas	EHPER2006I.dbf

Para la identificación de los archivos, como en el caso anterior, se observan las reglas de construcción de nombres adoptadas en el proyecto de elaboración de la Base de Datos IIES, por lo general, se incluye una referencia a la encuesta de Hogares por Muestreo (EH), la clase de datos que contiene el archivo (VIV, HOG, PER), el año y semestre de referencia (AAAASS) o (AASS)

Después de un carácter punto (.), el formato del archivo que sea objeto de procesamiento o resulte como consecuencia de las operaciones que se realicen: DBF: Formato de base de datos, DBASE y otros programas, SAS: Formato de datos del programa SPSS, SYS: Formato de archivos sintaxis del programa SPSS, SPO: Formato de salida visual SPSS, XLS: Formato de hoja de cálculo EXCEL.

En todos los ejemplos se utiliza en semestre primero del año 2006, el cual se representa como 2006I ó 06I, según el contexto.

1) La primera tarea es transferir los archivos en formato DBF a formato SAV del paquete estadístico SPSS, acción que se debe realizar con las facilidades que ofrecen los paquetes de manejo de datos y/o procesamiento estadístico como DBASE, FOX o STATA:

Archivo DBF	Manejador de datos archivo txt	SPSS archivo SAV
EHVIV2006I.dbf	EHVIV2006I.txt	EHVIV2006I.sav
EHHOG2006I.dbf	EHHOG2006I.txt	EHHOG2006I.sav
EHPER2006I.dbf	EHPER2006I.txt	EHPER2006I.sav

2) Luego, se consolidan los tres archivos en uno que contenga todos los datos necesarios para el procesamiento con el programa SPSS. Esta consolidación se ejecuta con el programa SyntaxVivHog2006II.sps

El programa SyntaxVivHog2006II.sps requiere que se actualice la información relativa a la identificación de los archivos de vivienda, hogares y personas.

```
EHVIV2006I.sav  
EHHOG2006I.sav ----- SyntaxVivgoh2006II.sps ----- EH2006I.SAV  
EHPER2006I.sav
```

Este último archivo, archivo de datos INE consolidado, contiene los datos de personas, información sobre las variables Tipo de vivienda, Tenencia de la vivienda y las variables auxiliares Cedula y Cedulaviv que facilitan las operaciones de agregación de datos para producir indicadores de hogares. Dicho archivo consolidado se utiliza en la etapa subsiguiente pero no se altera para preservar la integridad de los datos INE.

II. Generación de códigos sumarios y variables según definición IIES

En la fase de generación de códigos sumarios y variables según definición IIES, se construye para cada individuo u observación en el archivo, variables relativas a su posición en la Fuerza de Trabajo (Código sumario INE para Fuerza de trabajo), se reagrupan según los objetivos del IIES las variables sexo, edad, ocupación, actividad económica y se define operativamente la variable formalidad/informalidad.

También, se genera una base de datos de hogares consolidando la información correspondiente a los individuos, casos u observaciones, de un mismo hogar, dicha base de datos contiene información referida al Jefe del hogar, Aportante máximo al ingreso del hogar, Ingresos del hogar y otras características relacionadas.

Se crean dos archivos que se diferencian del archivo de datos INE consolidado porque incluyen en su nombre la letra m y la expresión HOG.

Adicionalmente, se genera una salida donde se visualiza un cuadro estadístico de códigos sumarios INE cruzados con códigos sumarios IIES, esta salida debe guardarse como un archivo SPSS con extensión SPO.

El archivo sintaxis SPSS que interviene en esta fase, se denomina Syntaxpob2005II.sys y demanda la corrección de la ruta del archivo, año y semestre de la Encuesta de Hogares que se esté procesando.

EH2006I.SAV ----- Syntaxpob2005II.sys ----- EH2006Im.SAV
 EH2006IHOG.SAV
 Salida visual SPSS

La salida visual SPSS debe grabarse en el directorio de trabajo con el nombre genérico utilizado para identificar el año y semestre, seguido de la extensión spo.

Salida visual SPSS ----- EH2006I.SPO

III. Resultados

Los cuadros estadísticos que se generan en esta fase, deben guardarse en la salida visual SPSS generada y guardada en la fase anterior, por tanto, dicho archivo (EH2006I.SPO) debe estar activo antes de iniciar el proceso.

En esta fase interviene el archivo de sintaxis SPSS Syntaxresultot2005.sys .

EH2006Im.SAV ----- Syntaxresultot2005.sys ----- EH2006I.SPO
 EH2006I.SPO

EH2006IHOG.SAV (Llamado por programación)

Igualmente, se graba, con el mismo nombre de archivo, la salida visual SPSS que contiene todos los cuadros estadísticos del semestre en referencia.

Salida visual SPSS ----- EH2006I.SPO

IV. Modificación y exportación del archivo salida visual SPSS

En la salida visual SPSS los cuadros correspondientes a Fuerza de Trabajo se disponen en forma de columna, tal como se muestra a continuación:

		pttq... Quince y +	pttq... Pob/activa	pttq... Pob/ocupada	pttq... Pob/desoc.	pttq... DQNBT
N	Valid	18693110	12222910	11115313	1107597	32293
	Missing	0	6470200	7577797	17585513	18660817
Sum		18693110,0 0	12222910,0 0	11115313,00	1107597,00	32293,00

Para su inclusión en la base de datos histórica de la Encuesta de Hogares, toda la información debe ser dispuesta en filas, por tanto, se debe transponer cada uno de los cuadros de Fuerza de Trabajo. Al efecto, se utilizan las instrucciones PIVOT y Transpuesta del SPSS, tal como se muestra.

	N		Sum
	Valid	Missing	
pttq... Quince y +	18693110	0	18693110,00
pttq... Pob/activa	12222910	6470200	12222910,00
pttq... Pob/ocupada	11115313	7577797	11115313,00
pttq... Pob/desoc.	1107597	17585513	1107597,00
pttq... Pob/cesante	1008693	17684417	1008693,00
pttq... Pob/BTPPV	98904	18594206	98904,00
pttq... Pob/inactiva	6425647	12267463	6425647,00
pttq... Estudios	2172727	16520383	2172727,00
pttq... Hogar	2941862	15751248	2941862,00
pttq... Incapac.	280724	18412386	280724,00
pttq... Ot/situación	490655	18202455	490655,00
pttq... Jubilados	494616	18198494	494616,00
pttq... Rentista	12770	18680340	12770,00
pttq... DQNBT	32293	18660817	32293,00

Es importante verificar que esta operación se cumple para todos y cada uno de los cuadros: Total de personas, varones por tramo de edad y hembras por tramos de edad.

Una vez transpuestos todos los cuadros de Fuerza de Trabajo, se salva la salida visual SPSS y, además, se exporta dicha salida a formato HTLM con la facilidad EXPORT del mismo paquete SPSS.

EH2006I.SPO ----- SPSS/EXPORT ----- OUTPUT2006I.HTLM

Vale resaltar que el archivo exportado es una imagen en formato HTLM de la salida visual SPSS, esta información no puede anexarse directamente en la base de datos de hogares, por tanto, debe procesarse a objeto de convertir los cuadros estadísticos en filas y columnas de variables y datos.

Por otra parte, este archivo puede ser leído directamente por EXCEL.

V. Conversión de cuadros estadísticos en archivo EXCEL de filas y columnas

El proceso de conversión de los cuadros estadísticos en filas y columnas de variables y datos, se realiza a partir del archivo HTLM de cuadros estadísticos (exportado en la fase anterior) Para ello, se utiliza un archivo Excel auxiliar, INTERMEDIO.XLS, que es un formato para hacer la referida conversión.

El archivo INTERMEDIO.XLS consta de las Hojas 1, Columna B, Columna C, Percentil y Hoja 5, cuyos formatos e información están diseñados para facilitar la conversión en referencia.

OUTPUT2006I.HTLM ----- EXCEL ----- EH2006I.XLS
 INTERMEDIO.XLS
 (Cargados simultáneamente)

La conversión de los cuadros estadísticos en filas y columnas es una actividad que requiere un nivel elevado de detalle y manipulación directa de los datos, consta de varias fases, las cuales se describen a continuación:

1). Copia de la Hoja 1 del archivo HTLM en el archivo INTERMEDIO, Hoja 1, a partir de la segunda columna, columna B.

INTERMEDIO, Hoja1 contiene en la primera columna los números 1, 2, 3,... razón por la cual se “pegan” los datos a partir de la columna B.

Comandos: Copiar (CTLc) y Pegado especial-Valores.

2). Identificación de columnas en archivo INTERMEDIO Hoja 1

En el archivo INTERMEDIO: borrar la Hoja B fila 2 y copiarla en la Hoja 1 fila 1.

Comandos: Borrar (CTLx) y Pegar (CTLv)

Las primeras filas del archivo INTERMEDIO Hoja 1 debe tener un aspecto similar al que se presenta a continuación:

secuencia	Identific b	Identif c	Frecuencia			
1						
2	Frequencies					
3						
4	Statistics					
5	Actividad económica					
6	N	Valid	65206			
7		Missing	0			
8	Sum		4011398			
9						
10	Actividad económica					
11				Frequency	Percent	Valid Percent Cumulative Percent
12	Valid	mttqo.. 01 Agricultura	7612	11,7	11,7	11,7
13		mttqo.. 02 Pet/gas	231	0,4	0,4	12
14		mttqo.. 03 Min/canteras	177	0,3	0,3	12,3
15		mttqo.. 04 Manufact	6944	10,6	10,6	22,9
16		mttqo.. 05 Elect/gas/agua	328	0,5	0,5	23,5
17		mttqo.. 06 Construcción	5227	8	8	31,5
18		mttqo.. 07 Com m/mayor	13003	19,9	19,9	51,4
19		mttqo.. 08 Rest/hotel	2325	3,6	3,6	55
20		mttqo.. 09 Transporte	4479	6,9	6,9	61,8
21		mttqo.. 10 Comunicac	414	0,6	0,6	62,5
22		mttqo.. 11 Banca	362	0,6	0,6	63
23		mttqo.. 12 Seg/S.empres	2370	3,6	3,6	66,7
24		mttqo.. 13 Servicios	21486	33	33	99,6
25		mttqo.. 14 Actividad NEP	248	0,4	0,4	100
26		Total	65206	100	100	
27						

3). Ordenar archivo INTERMEDIO Hoja 1 por la columna B, Identificación b.

Comandos: DATOS/ORDENAR/Encabezado=Si/Identificación b/Ascendente.

4). En el archivo INTERMEDIO Hoja A: borrar y copiar en Hoja Percentil, las filas de datos que en la columna B se identifican como:

Maximum, Mean, Median, Minimun, Mode

El archivo INTERMEDIO Hoja Percentil tiene un aspecto similar al que se presenta a continuación, por supuesto, con más filas (son alrededor de 75 medias y modas)

secuencia	Nombre		Frecuencia
12980	Maximum		11635000
2138	Mean		108760,381
.....
12977	Median		350000
12979	Minimum		1000
.....
2139	Mode		0
2267	Mode		400000

Nota: Se trata de borrar y copiar todas las filas que cumplan la condición.
Tener cuidado de no sobrescribir ninguna fila de la Hoja percentil.

5). En archivo INTERMEDIO Hoja 1 columna B, borrar y copiar en Hoja B, las filas de datos que se identifican como:

Pob/activa, Pob/ocupada, Pob/cesante, Pob/BTPPV,
Pob/inactiva, Estudios, Hogar, Incapac., Rentista, Jubilados, Ot/situación, DQNBT

Esta información se dispone en dos bloques y, además, se reconoce porque los nombres se duplican.

La Hoja B INTERMEDIO luce como se presenta a continuación:

secuencia	Nombre	Frecuencia		
1671	pthc... DQNBT	1371	2149404	
1751	pthc... DQNBT			
1665	pthc... Estudios	6231	2144544	
1715	pthc... Estudios			
....
287	pttq... Quince y +	18473790	0	18473790
304	pttq... Quince y +			
299	pttq... Rentista	14235	18459555	14235
374	pttq... Rentista			
....
875	ptvc... DQNBT	966	2102806	966
955	ptvc... DQNBT			
736	ptw... Rentista	501	3850809	501
811	ptw... Rentista			

Obsérvese que hay datos en columnas no identificadas, estos se borrarán al final del proceso.

6). Ordenar archivo INTERMEDIO Hoja 1 por la columna C, Identificación c.
Comandos: DATOS/ORDENAR/Encabezado=Si/Identificación c/Ascendente.

7). En archivo INTERMEDIO Hoja 1, borrar y copiar en archivo INTERMEDIO Hoja Percentil las filas de datos que en la columna C se identifican como:

Percentiles 05/10/15/20/25/30/35/40/45/50/55/60/65/70/75/80/85/90/95

Ahora, la Hoja Percentil tiene una estructura similar a la que se presenta. Considérese que las filas de percentiles desde el 05 al 95 son más de 1400.

secuencia	Nombre		Frecuencia
12980	Maximum		11635000
2138	Mean		108760,381
-----	-----	-----	-----
12977	Median		350000
12979	Minimum		1000
-----	-----	-----	-----
2139	Mode		0
2267	Mode		400000
-----	-----	-----	-----
2140	Percentiles	5	0
2268	Percentiles	5	80000
-----	-----	-----	-----
12748		95	1200000
13008		95	2410000

8). Borrar en archivo INTERMEDIO Hoja 1 y copiar en archivo INTERMEDIO Hoja C las filas de datos que en la columna C se identifican con cualquier combinación de clave de identificación + nombre de variable:

Comienzan con m, tales como:

mptqo.. 12 Seg/S.empres mitqo.. 13 Servicios mptqo.. 05 Elect/gas/agua

Comienzan con p, tales como:

pftqo60 09 Cuenta propia pgtqc99 05 Ed 25-44 pptqo99 10 Univ1-2

Comienzan con z, tales como:

z hog 01 Aportante máximo es jefe zhog 06y+ ocupad/hogar

9). Una vez copiada la información indicada en los pasos 4 al 8

Eliminar la Hoja 1.

Borrar en archivo INTERMEDIO Hoja C la columna B.

10). En el archivo INTERMEDIO ordenar la Hoja B en forma ascendente para la variable Frecuencia.

Comandos: DATOS/ORDENAR/Encabezado=Si/Frecuencia/Ascendente

Copiar, desde la fila 2 en adelante, todas las filas de datos con valor 0 o mayor y pegar dicha información en la Hoja C. Nótese que se excluyen las filas con blanco en la columna Frecuencia.

11). En el archivo INTERMEDIO borrar la Hoja B.

12). En el archivo INTERMEDIO ordenar la Hoja Percentil en forma ascendente por la variable Secuencia, luego:

Insertar una columna entre las columnas B y C.

En la columna insertada copiar desde la Hoja E la columna B.

Las primeras filas de la Hoja Percentil guardan semejanza con la siguiente representación.

secuencia	Nombre	percentil ingreso		Frecuencia
2138	Mean	pttdo00 media		108760,381
2139	Mode	pttdo00 moda		0
2140	Percentiles	pttdo00 percentil ingreso 05	5	0
2141		pttdo00 percentil ingreso 10	10	0
2142		pttdo00 percentil ingreso 15	15	0
2143		pttdo00 percentil ingreso 20	20	0
2144		pttdo00 percentil ingreso 25	25	0
2145		pttdo00 percentil ingreso 30	30	12000
2146		pttdo00 percentil ingreso 35	35	30000
2147		pttdo00 percentil ingreso 40	40	40000
2148		pttdo00 percentil ingreso 45	45	60000
2149		pttdo00 percentil ingreso 50	50	80000
2150		pttdo00 percentil ingreso 55	55	100000
2151		pttdo00 percentil ingreso 60	60	120000
2152		pttdo00 percentil ingreso 65	65	120000
2153		pttdo00 percentil ingreso 70	70	150000
2154		pttdo00 percentil ingreso 75	75	160000
2155		pttdo00 percentil ingreso 80	80	200000
2156		pttdo00 percentil ingreso 85	85	240000
2157	pttdo00 percentil ingreso 90	90	280000	
2158	pttdo00 percentil ingreso 95	95	400000	
2266	Mean	pttqo00 media		552686,023
2267	Mode	pttqo00 moda		400000

El final de la Hoja Percentil debe tener información para todas las columnas como en el ejemplo que se presenta a continuación.

13005	zhog c/ingreso percentil ingres	80	1395000
13006	zhog c/ingreso percentil ingres	85	1570000
13007	zhog c/ingreso percentil ingres	90	1868000
13008	zhog c/ingreso percentil ingres	95	2410000

Si la Hoja Percentil no tiene información para todas las columnas, significa que para alguna (s) rama (s) de actividad económica la Encuesta de Hogares no registra personal ocupado, deben realizarse dos acciones complementarias:

Verificar con el archivo OUTPUTEHaaaass cuáles son las ramas de actividad sin información de ocupados.

Eliminar en el archivo INTERMEDIO Hoja percentil columna percentil ingreso los datos de dichas ramas de actividad.

13). En el archivo OUTPUTEHaaaass, utilizar la herramienta buscar con la palabra mean y hacer repetidamente uso de la opción Buscar siguiente. El cursor se posiciona en la palabra mean de cada uno de los siguientes grupos:

Total	Público	Privado	Informal	Formal
pttdo00 media				
pttqo00 media	pgtqo00 media	pptqo00 media	pitqo00 media	pftqo00 media
10	10	10	10	10
20	20	20	20	20
21	21	21	21	21
30	30	30	30	30
40	40	40	40	40
50	50	50	50	50
60	60	60	60	60
61	61	61	61	61
70	70	70	70	70
71	71	71	71	71
80	80	80	80	80
81	81	81	81	81
90	90	90	90	90
99	99	99	99	99

14). Eliminar en el archivo INTERMEDIO Hoja percentil columna percentil ingreso los datos de las ramas de actividad de la lista anterior cuyas medias no se encuentren, por lo general, pitqo 20/pitqo 21 ó pitqo 99.

Supongamos que se trata de la rama de actividad pitqo 20, tal como se muestra a continuación:

secuencia	Nombre	percentil ingreso	Frecuencia
10044		pitqo10 percentil ingreso 90	90 600000
10045		pitqo10 percentil ingreso 95	95 800000
10088	Mean	pitqo20 media	380000
10089	Mode	pitqo20 moda	380000
10090	Percentiles	pitqo20 percentil ingreso 05	5 380000
10091		pitqo20 percentil ingreso 10	10 380000
10092		pitqo20 percentil ingreso 15	15 380000
10093		pitqo20 percentil ingreso 20	20 380000
10094		pitqo20 percentil ingreso 25	25 380000
10095		pitqo20 percentil ingreso 30	30 380000
10096		pitqo20 percentil ingreso 35	35 380000
10097		pitqo20 percentil ingreso 40	40 380000
10098		pitqo20 percentil ingreso 45	45 380000
10099		pitqo20 percentil ingreso 50	50 380000
10100		pitqo20 percentil ingreso 55	55 380000
10101		pitqo20 percentil ingreso 60	60 380000
10102		pitqo20 percentil ingreso 65	65 380000
10103		pitqo20 percentil ingreso 70	70 380000
10104		pitqo20 percentil ingreso 75	75 380000
10105		pitqo20 percentil ingreso 80	80 380000
10106		pitqo20 percentil ingreso 85	85 380000
10107		pitqo20 percentil ingreso 90	90 380000
10108		pitqo20 percentil ingreso 95	95 380000
10182	Mean	pitqo21 media	354887,502
10183	Mode	pitqo21 moda	300000

Se marca con el cursor.

Con el botón izquierdo o en menú se utiliza la opción eliminar,
Seguido de la opción desplazar las celdas hacia arriba.

Luego de esta operación de eliminación y desplazamiento de celdas, se reacomoda la información, obsérvese que pitqo 21 se posiciona en los indicadores adjudicados a la rama eliminada que le antecedió, pitqo 20. Las restantes ramas, también se desplazan, en el ejemplo, pitqo 30, ocupa la posición que deja vacante el desplazamiento de pitqo 21 y así sucesivamente.

secuencia	Nombre	percentil ingreso	Frecuencia
10044		pitqo10 percentil ingreso 90	90 600000
10045		pitqo10 percentil ingreso 95	95 800000
10088	Mean	pitqo21 media	380000
10089	Mode	pitqo21 moda	380000
10090	Percentiles	pitqo21 percentil ingreso 05	5 380000
10091		pitqo21 percentil ingreso 10	10 380000
10092		pitqo21 percentil ingreso 15	15 380000
10093		pitqo21 percentil ingreso 20	20 380000
10094		pitqo21 percentil ingreso 25	25 380000
10095		pitqo21 percentil ingreso 30	30 380000
10096		pitqo21 percentil ingreso 35	35 380000
10097		pitqo21 percentil ingreso 40	40 380000
10098		pitqo21 percentil ingreso 45	45 380000
10099		pitqo21 percentil ingreso 50	50 380000
10100		pitqo21 percentil ingreso 55	55 380000
10101		pitqo21 percentil ingreso 60	60 380000
10102		pitqo21 percentil ingreso 65	65 380000
10103		pitqo21 percentil ingreso 70	70 380000
10104		pitqo21 percentil ingreso 75	75 380000
10105		pitqo21 percentil ingreso 80	80 380000
10106		pitqo21 percentil ingreso 85	85 380000
10107		pitqo21 percentil ingreso 90	90 380000
10108		pitqo21 percentil ingreso 95	95 380000
10182	Mean	pitqo30 media	354887,502
10183	Mode	pitqo30 moda	300000

15). Después de “cuadrar” la Hoja percentil, se eliminan las columnas Nombre y la columna D, sin nombre en el ejemplo, y se trasladan todos los datos (desde la fila 2 hasta la última fila con información a la Hoja Columna C del mismo archivo INTERMEDIO.

Con esta operación, todos los datos del semestre en estudio se encuentran en el archivo INTERMEDIO Hoja C.

16). Se ordena el archivo INTERMEDIO Hoja C por la columna Nombre y se hacen los siguientes cambios de de identificación de columnas:

Secuencia	Secaass	(Sec año y semestre procesado, ejemplo: Sec06I)
Frecuencia	Frecaass	

sec06I	Nombre	Frec06I
12946	zhog Pob total media	4,3039
12947	zhog Pob total mediana	4
12948	zhog Pob total modo	4

La información en el archivo resultante está ordenada de manera tal que puede incorporarse a la base de datos que contiene la información de todos los semestres de la Encuesta de Hogares ya procesados, archivo Base Datos IIES. Con este fin, en un proceso posterior, se requiere la posición del último dato (resaltado en el cuadro anterior) según el sistema Excel de identificación de celdas. En el caso del semestre utilizado como ejemplo, el archivo EH2006I, esta posición se identifica como C6119.

17). Si no lo ha hecho antes, asigne un nombre al archivo y grábelo, en el ejemplo: EH2006I.XLS

VI. Ampliación del archivo Base de Datos IIES

En SPSS, el archivo de sintaxis ARCHEXCEL.SYS contiene los procedimientos para:

- Leer el archivo EXCEL con los datos de un semestre en particular convertidos en filas y columnas de una hoja de cálculo SPSS.
- Aparear (Hacer matching) las variables del semestre y las variables de un archivo ordenado según la presentación de la Base de Datos IIES.
- Exportar a EXCEL un archivo ordenado y listo para agregar a la Base de Datos IIES.

1). En el archivo SPSS ARCHEXCEL:SYS modificar y resaltar el bloque de transferencia de archivos EXCEL a SPSS.

La modificación debe incluir la ruta, el nombre del archivo EXCEL y el rango desde la celda A1 hasta la celda con el último dato del archivo.

***Copiar Excel a SPSS: modificar nombre de archivo y rango .**

```
GET DATA
  /TYPE = XLS
  /FILE = 'c:\users\UCVhogares\EH2006I.xls'
  /CELLRANGE = RANGE 'A1:C6119'
  /READNAMES = ON .
EXECUTE .
```

2). Para que se ejecute el bloque anterior resaltado, utilizar en el Menu la opción:
RUN/Selection

Se instala en SPSS una imagen del archivo EXCEL en proceso, EH2006I en el ejemplo.

En SPSS, la hoja Variable View debe ser como la siguiente representación. En caso de no coincidir los campos Width y Decimals, deben ajustarse.

Name	Type	Widtht	Decimals
sec06Ix	Numeric	11	0
Nombre	String	35	0
Frec06Ix	Numeric	11	0

3). Guardar el archivo SPSS con el nombre BASEDATHOGINT.SAV, para ello, si es necesario se sobrescribe cualquier versión anterior.

4). Cargar como archivo de trabajo SPSS al archivoBASEDATEXCEL.SAV

5). Resaltar y ejecutar sin ninguna modificación el bloque Merge que se presenta a continuación.

*** Merge archivos de semestres diferentes no coinciden las variables .**
UPDATE / FILE = *
/FILE = 'C:\users\UCVhogares\BASEDATHOGINT.sav'
/BY Nombre .
EXECUTE .

Comandos: RUN/Selection

6). Seguidamente, resaltar y ejecutar el bloque de exportación a Excel, el cual, crea un archivo BASEDATANEXO.XLS con las variables en el mismo orden que la Base de Datos IIES. Los comandos a utilizar son: RUN/Selection

*** Exportación a EXCEL .**
SAVE TRANSLATE OUTFILE = 'c:\users\UCVhogares\BASEDATANEXO.XLS' replace
/TYPE = XLS /VERSION = 8
/FIELDNAMES
/CELL = VALUES .
EXECUTE .

7). El archivo BASEDATANEXO.XLS consta de tres columnas cuya identificación es:

Nombre SecuenciaY Frecaass

Se ordena por SecuenciaY y se anexa la columna Frecaass al archivo Base Datos IIES, acción con la cual concluye la incorporación de un nuevo semestre a la base de datos IIES.

Alternativamente, se puede trasladar las columnas SecuenciaY y Frecaass al archivo Base Datos IIES y hacer la diferencia entre columnas que se indica:
Secuencia – SecuenciaY

Todas las celdas resultantes deben ser cero, de lo contrario el archivo BASEDATANEXO.XLS no está ordenado correctamente y debe rehacerse todos los pasos de este apartado.

En el caso de que todas las celdas resultantes sean cero, se borra la columna SecuenciaY cuyo valor es de comprobación únicamente.

Universidad Central de Venezuela
Facultad de Ciencias Económicas y Sociales
Instituto de Investigaciones Económicas y Sociales “Rodolfo Quintero”

Proyecto Observatorio Económico Social
Guía para actualizar la base de datos del mercado laboral

Noviembre, 2007