

Universidad Católica del Uruguay
Facultad de Ciencias Humanas

Autotexto
Serie Estadística



Universidad
Católica

DAMASO A. LARRAÑAGA • URUGUAY

Transformación de bases
de datos: Fundir Archivos y Agregar

Laboratorio Metodológico

Autores

M. J. Álvarez -A. de León – M. Dodel – C. Rafaniello

2008

Módulo de Práctica de Análisis

Serie Estadística

**Tema: Transformación de Bases de Datos.
Fundir Archivos (añadir casos y variables) y Agregar.**

Descripción

En esta guía explicaremos los pasos que deben seguirse para relacionar dos bases de datos. Lógicamente, para que ello sea posible, ambos archivos deben poder vincularse mediante alguna variable o combinación de variables, pero eso lo veremos caso a caso.

A continuación expondremos los pasos a seguir para Fundir y Agregar Archivos.

Usos de los comandos

Fundir Archivos

La función “*Fundir Archivos*” se emplea cuando se desean añadir casos o variables de una base a otra. Este comando es utilizado en conjunto con otros comandos (agregar) o para unificar distintos conjuntos parciales de una misma base de datos (por ejemplo la ECH trabaja con bases separadas de Montevideo e Interior, si queremos generar una base para el total del país tenemos que utilizar este comando).

Agregar

La función de “Agregar” es una de las más usadas para manipular bases. Nos permite condensar una base colapsando datos, basándonos en una o más variables de agrupación.

A modo de ejemplo, si tenemos la base de datos de los empleados de una compañía, y queremos investigar las características de la población de la misma basándonos en su educación, podemos agrupar todas las variables que queramos en filas definidas por años de educación:

Cada fila será, por ejemplo, el promedio de horas trabajadas para los que tienen 7 años de educación, o el promedio de sueldo de los que tienen 14 años de educación.

Esta función sirve no sólo para calcular el promedio, sino que también podemos agregar o agrupar en base a otras medidas, como porcentajes, valores máximos y mínimos, etc.

Procedimiento

1. Añadir registros (casos) de un archivo de datos a otro
2. Agregar variables y añadirlas de un archivo de datos a otro

1. Añadir registros (casos) de un archivo de datos a otro

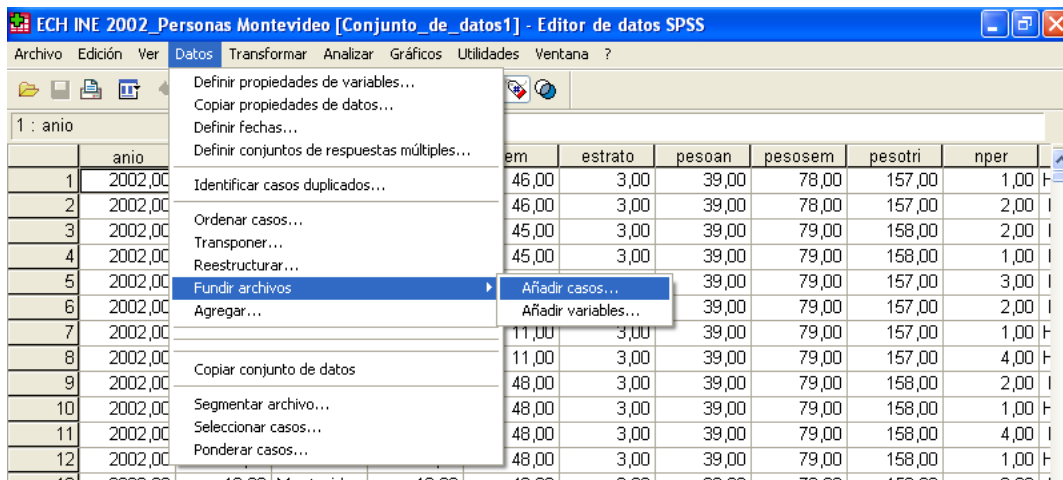
Supongamos que conseguimos los archivos de datos de las personas de la ECH y nos han entregado uno para Montevideo y otro para el Interior Urbano, ambos conteniendo las mismas variables, ya que el cuestionario es idéntico para todo el país. Al momento de comenzar a explorar la información que contienen los archivos sería muy engorroso trabajar ambas bases por separado, por lo cual intentaremos añadir los casos de un archivo a otro.

Procedimiento para añadir registros

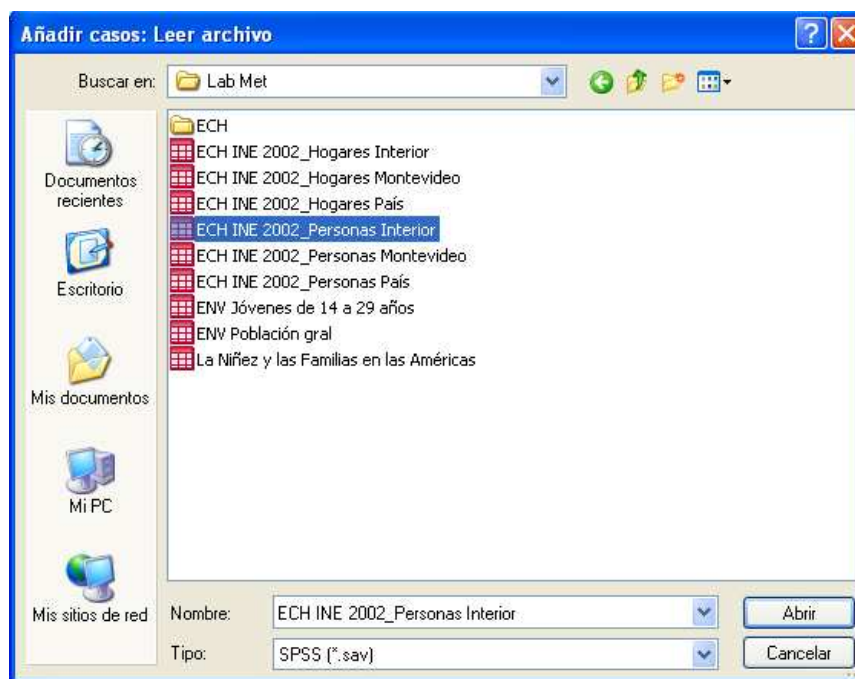
El primer paso es abrir uno de los dos archivos. Comencemos por Montevideo.

	anio	correlat	dpto	mes	sem	estrato	pesoan	pesosem	pesotri	nper
1	2002,00	1,00	Montevideo	12,00	46,00	3,00	39,00	78,00	157,00	1,00
2	2002,00	1,00	Montevideo	12,00	46,00	3,00	39,00	78,00	157,00	2,00
3	2002,00	8,00	Montevideo	12,00	45,00	3,00	39,00	79,00	158,00	2,00
4	2002,00	8,00	Montevideo	12,00	45,00	3,00	39,00	79,00	158,00	1,00
5	2002,00	15,00	Montevideo	3,00	11,00	3,00	39,00	79,00	157,00	3,00
6	2002,00	15,00	Montevideo	3,00	11,00	3,00	39,00	79,00	157,00	2,00
7	2002,00	15,00	Montevideo	3,00	11,00	3,00	39,00	79,00	157,00	1,00
8	2002,00	15,00	Montevideo	3,00	11,00	3,00	39,00	79,00	157,00	4,00
9	2002,00	16,00	Montevideo	12,00	48,00	3,00	39,00	79,00	158,00	2,00
10	2002,00	16,00	Montevideo	12,00	48,00	3,00	39,00	79,00	158,00	1,00
11	2002,00	18,00	Montevideo	12,00	48,00	3,00	39,00	79,00	158,00	4,00
12	2002,00	18,00	Montevideo	12,00	48,00	3,00	39,00	79,00	158,00	1,00
13	2002,00	18,00	Montevideo	12,00	48,00	3,00	39,00	79,00	158,00	2,00
14	2002,00	18,00	Montevideo	12,00	48,00	3,00	39,00	79,00	158,00	3,00
15	2002,00	20,00	Montevideo	12,00	48,00	3,00	39,00	79,00	158,00	2,00
16	2002,00	20,00	Montevideo	12,00	48,00	3,00	39,00	79,00	158,00	4,00
17	2002,00	20,00	Montevideo	12,00	48,00	3,00	39,00	79,00	158,00	3,00
18	2002,00	20,00	Montevideo	12,00	48,00	3,00	39,00	79,00	158,00	5,00
19	2002,00	20,00	Montevideo	12,00	48,00	3,00	39,00	79,00	158,00	1,00
20	2002,00	20,00	Montevideo	12,00	48,00	3,00	39,00	79,00	158,00	6,00
21	2002,00	23,00	Montevideo	11,00	43,00	3,00	39,00	79,00	158,00	1,00
22	2002,00	43,00	Montevideo	7,00	27,00	3,00	39,00	78,00	157,00	1,00

Luego de abierto el archivo de Montevideo, para poder añadir los casos del Interior utilizamos el comando *Datos / Fundir Archivos / Añadir Casos*. De esta forma le estamos pidiendo al programa que adicione los casos que hay en otro archivo al que ya tenemos abierto.



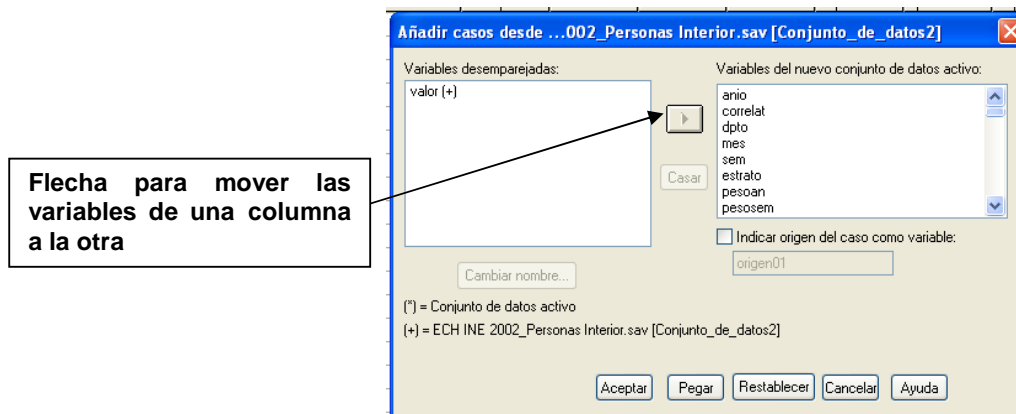
En ese momento se nos abre un nuevo recuadro de diálogo, en el que debemos seleccionar el archivo que contiene la información a incorporar. En este caso debemos marcar **ECH INE 2002_Personas Interior.sav**.



Una vez seleccionado el archivo de datos a incorporar se abre otro recuadro de diálogo, el cual se encuentra dividido en dos grandes columnas: “*Variables no apareadas*” y “*Variables en el nuevo archivo de datos*”.

Como el título lo indica, en esta última columna se presentan todas las variables que serán incorporadas en el nuevo archivo de datos. Por el contrario, en la primer columna se presentarán todas aquellas variables que no coinciden en ambos archivos de datos, las que serán marcadas al costado con (*) o con (+) según cuál sea su archivo de origen. En cualquier caso, el programa admite que “forcemos” su incorporación al nuevo archivo de datos

seleccionando la variable excluida y pasándola con la flecha del medio a la segunda columna.¹



Una vez que estemos de acuerdo con las variables que serán incorporadas en el nuevo archivo presionamos el botón de “Aceptar” para añadir los registros de la base del Interior Urbano. Cumplida toda la operación podemos salvar el archivo con el nombre **ECH_2002 - Personas Todo el País.sav**.

Recomendaciones

Debemos estar seguros de que el procedimiento se ejecutó correctamente, y para ello basta verificar que la suma de los registros que contiene cada uno de los archivos parciales (Montevideo e Interior Urbano) sea igual a la cantidad que contiene el nuevo archivo de datos.

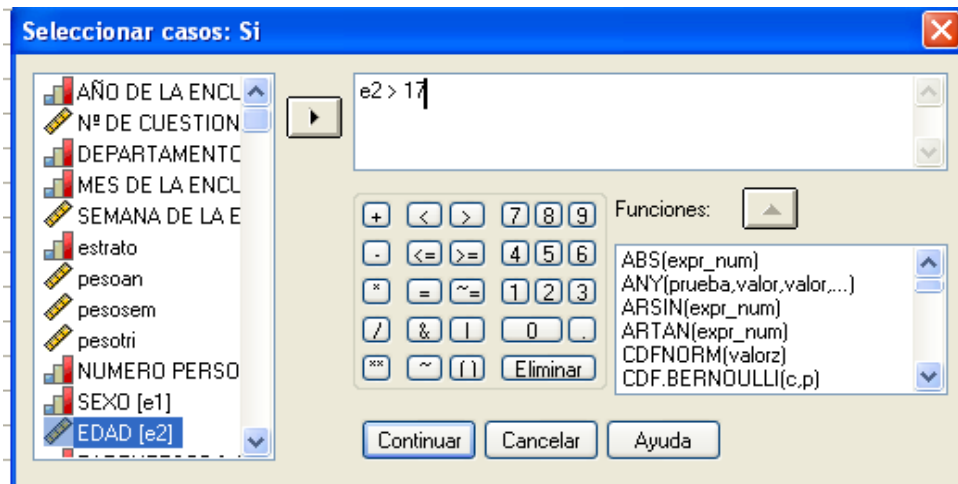
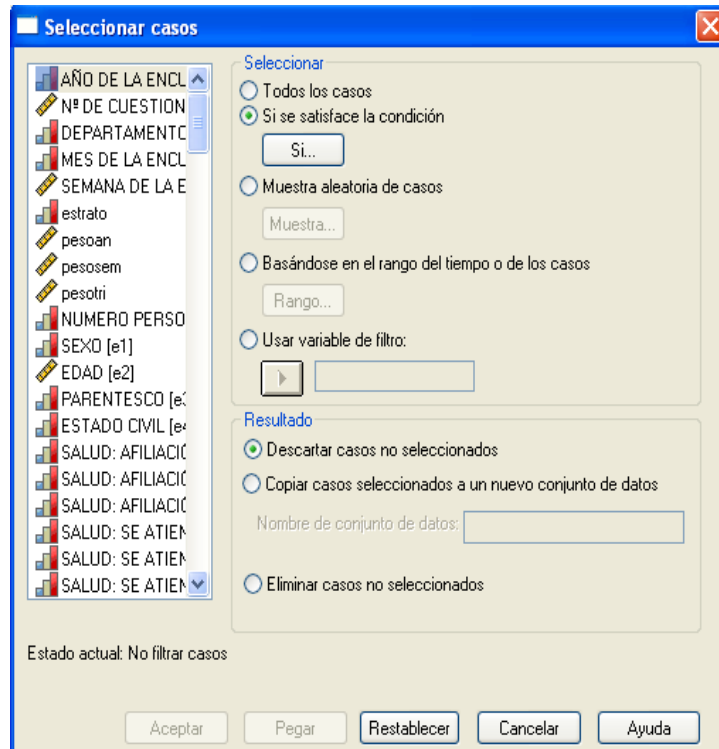
2. Agregar variables y añadirlas de un archivo de datos a otro

Con la ECH esta función nos es de mucha utilidad si queremos atribuirles a las personas características de los hogares. Por ejemplo, si queremos generar una variable en la base de personas que sea clima educativo del hogar (promedio de años de educación de los mayores de 18).

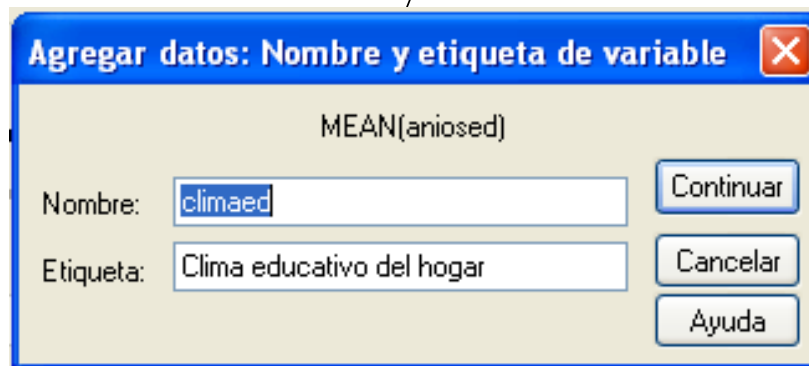
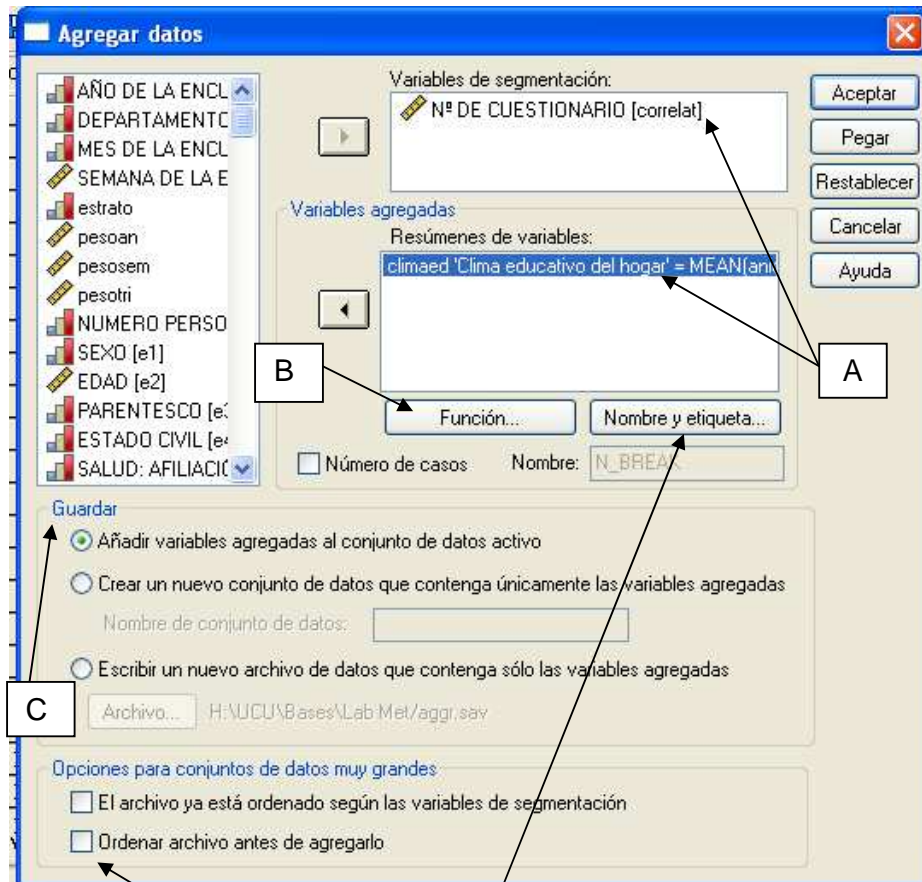
Procedimiento para agregar variables:

1. Lo primero que tenemos que hacer, en la base de personas, es filtrar a los mayores de 17. Para ello debemos ir a *Datos, Seleccionar Casos*.

¹ En este caso, la variable asumirá valor nulo (missing) cuando se trate de los registros del archivo de datos donde no estaba incluida la variable original.



2. Luego, debemos de ir a *Datos, Agregar*



A- Seleccionamos *Nº DE CUESTIONARIO (correlat)* -número único que identifica cada hogar- para que nos agrupe la información resumida para cada hogar. Y seleccionamos *Años de educación formal (añosed²)* -años de educación de cada persona- como la variable a resumir.

² Suma de todos los años de educación formal aprobados

Agregar datos: Función de agregación

Estadísticos de resumen	Valores específicos	Número de casos
<input checked="" type="radio"/> Media	<input type="radio"/> Primero	<input type="radio"/> Ponderado
<input type="radio"/> Mediana	<input type="radio"/> Último	<input type="radio"/> Perdido ponderado
<input type="radio"/> Suma	<input type="radio"/> Mínimo	<input type="radio"/> Sin ponderar
<input type="radio"/> Desviación típica	<input type="radio"/> Máximo	<input type="radio"/> Perdido sin ponderar

Porcentajes

Por encima Valor:

Por debajo

Dentro Menor: Mayor:

Fuera

Fracciones

Por encima Valor:

Por debajo

Dentro Mínimo: Superior:

Fuera

Continuar Cancelar Ayuda

- B-** Como mencionamos al comienzo del autotexto, si bien esta función nos da por defecto el promedio de la variable que queremos agregar, podemos cambiar el estadístico si así lo preferimos mediante el botón de Función.

En este caso, sin embargo, nos interesa el promedio.

- C-** Luego de seleccionar las variables de corte y de agregación, tenemos dos opciones a la hora de guardar las nuevas variables: Podemos adicionarlas a la base con la que estamos trabajando o crear un nuevo archivo que contenga únicamente las variables agregadas.

En caso de elegir la segunda opción tenemos, a su vez, dos alternativas: *“crear un nuevo conjunto de datos que contenga únicamente las variables agregadas”*—lo que implica generar un archivo nuevo y mantenerlo abierto al mismo tiempo que la base con la que estábamos trabajando— o *“escribir un nuevo archivo de datos que contenga sólo las variables agregadas”*—que se diferencia de la anterior alternativa en que en vez de mantenerlo en pantalla se guarda el nuevo archivo en el lugar que el usuario determine—.

- D-** Clickear la opción “Ordenar archivo antes de agregarlo”.

- E-** Finalmente, debemos de cambiar el nombre de la nueva variable (opcionalmente podemos etiquetarla también), para que no hayan conflictos en los pasos siguientes.

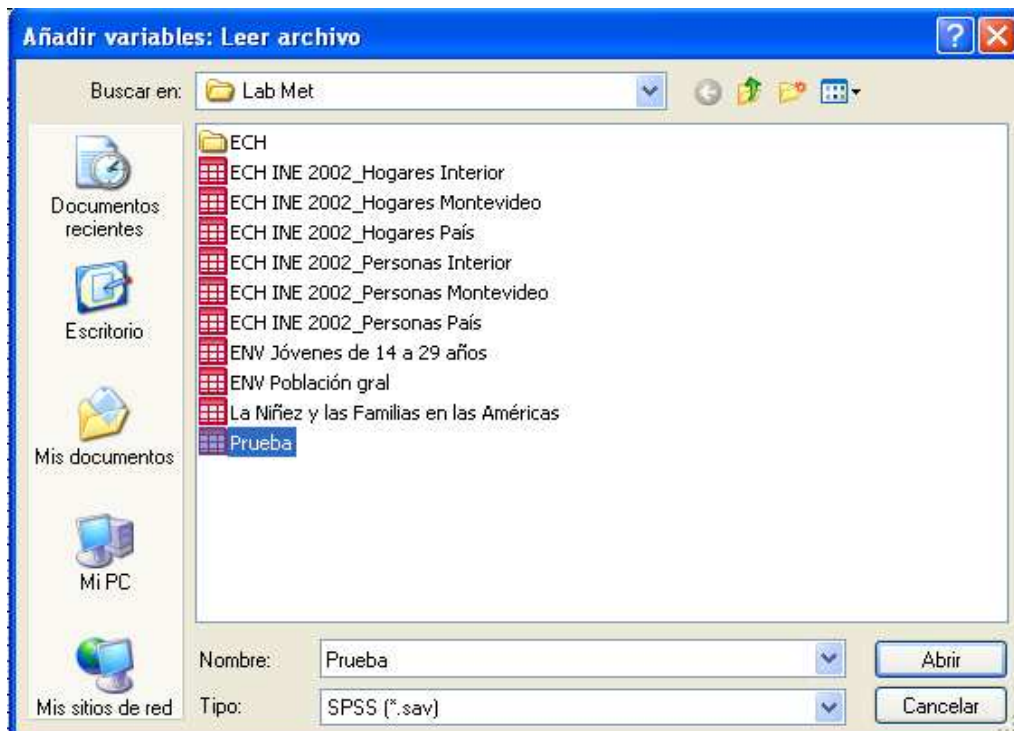
Al llegar a este punto ya tenemos generada una variable correspondiente del clima educativo del hogar, pero en la base de personas de la ECH.

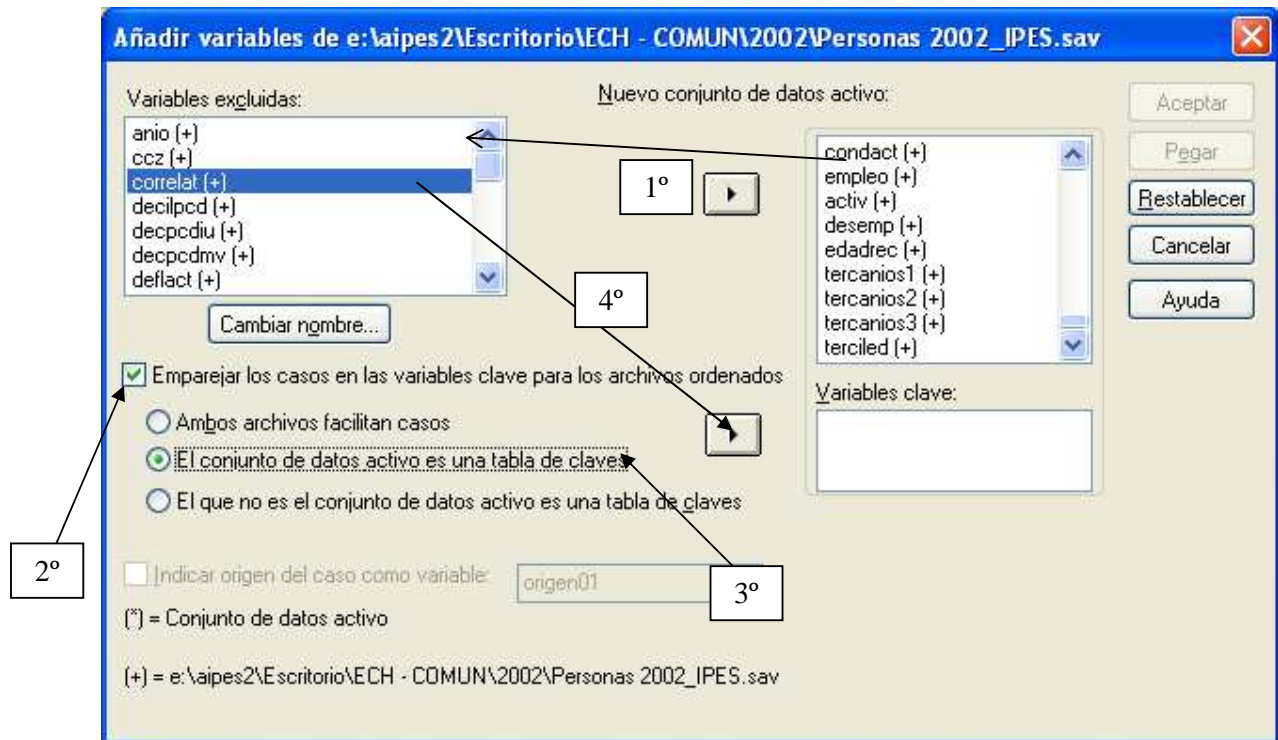
Supongamos que luego nos interesa indagar como se relaciona esta nueva variable con ciertos atributos del hogar, como por ejemplo, el total de ingresos del mismo.

Para ello, lo que deberemos de hacer es pasar esta nueva variable a la base de hogares. Esto se realiza mediante el comando *Fundir archivos / Añadir variables*.

3. Abrimos la base de personas y seleccionamos “Fundir Archivos”.

Debemos de ir a *Datos, fundir archivos, añadir variables*





Advertencia:

- Para que sea posible *añadir variables* entre dos bases de datos es necesario que exista una variable o combinación de variables común a ambas (clave), que permita identificar/distinguir de forma inequívoca un caso del otro. En el ejemplo, utilizaremos la única variable que estas bases tienen en común: *correlat*. O sea, el número del cuestionario del hogar.
- En el caso de la adición de variables a raíz de la agregación de variables, debemos usar como tabla de claves aquella base de datos con menor nivel de desagregación.

Para ello debemos realizar los siguientes cuatro pasos:

1. Asegurarse que la única variable del conjunto de datos no activo (personas) que se añada, sea la variable que deseamos: *climaed*. Para ello debemos de pasar todas las variables pertenecientes a este conjunto de datos (señaladas con el símbolo de +) menos *climaed* y pasarlas de la ventana "Nuevo conjunto de datos activo" a "variables excluidas".
2. Seleccionar la opción de *emparejar los casos en las variables clave para los archivos ordenados*
3. Si estamos trabajando desde la base de hogares como en este caso, el archivo con un menor nivel de desagregación es el actual. Por lo tanto seleccionamos la opción: *El conjunto de datos activo es una tabla de claves*

4. Como ya mencionamos, la variable “correlat” es la que servirá para vincular los dos archivos que se desea fundir.

Recomendaciones

Al finalizar el proceso, si no apareció ningún mensaje de error, es conveniente comprobar que el procedimiento se ejecutó correctamente. Esto lo realizamos sencillamente verificando que para todos los casos de un mismo hogar –igual valor de la variable *Nª DE CUESTIONARIO (correlat)*– la nueva variable *Clima educativo del Hogar (climaed)* tenga un mismo valor.

Ejercicios

Ejercicio 1:

- a. Describa el clima educativo de los hogares con jóvenes de 13 a 18 años que no estudian (use gráficas, medidas de tendencia central y de dispersión)
- b. ¿Es distinto al clima educativo de los hogares con jóvenes de la misma edad que si estudian? (además de tablas o gráficos, utilice la prueba de hipótesis adecuada para el tipo de variables)?

Ejercicio 2:

- a. A partir de dos archivos de datos, uno de hombres y otro de mujeres, genere un único archivo que contenga a todas las personas sin importar el género.
 - b. Una vez generado el archivo de personas, agregue a cada una de ellas la siguiente información contenida en el archivo “personas-educación.sav”: asistencia a la educación formal, tipo de establecimiento educativo al que asiste y años de educación completados.
-