Attaching spouse information using the Current Population Survey

The Current Population Survey has a three-level structure containing information on the individual, the family and the household. We are interested in creating a file of married people with information on their spouse attached. The following example show how you would attach information on an individuals spouse so that it is all in the same line. In other words, rather than having a hierarchical data structure, each line of data will contain information on the respondent and the respondent's spouse.

```
(1)      data cps03;
(2)      infile 'C:\DATA\cpsmar03\CPSMAR03.dat'  lrecl = 1300 missover ;
(3)      retain foodstamps hhid poverty;
(4)      INPUT @1 rectype 1. @;
(5)      if rectype = 1 then do;
(6)      INPUT
(7)      @2   hhid          5.
(8)      @76  foodstamps    1. ;
(9)      end;
(10)     else if rectype = 2 then do;          /*family record*/
(11)     INPUT
(12)     @37   poverty        1. ;
(13)     end;
(14)     else if rectype=3 then do; /*person record*/
(15)     INPUT
(16)     @9   lineno        2.
(17)     @15  age           2.
(18)     @17  marital       1.
(19)     @18  spouse        2.
(20)     @20  sex           1.
(21)     @722 pob           3. ;
(22)     if marital = 1 then  output;
(23)     end;
(24)     drop rectype;
```

Lines 1 through 31 input the raw data and create a person-level data file.

Line 1 creates a new dataset called 'cps03'.

Line 2 tells SAS where to go to get the raw data (a file called 'CPSMAR03.dat'), as well as the width of the raw data file (lrecl=1300) and to input until the end of the line of data no matter what.

Line 3 tells SAS to hold on to the values of the variables listed through the course of each record. The retain statement is necessary for any variables that you have listed in the household or family level records.

Line 4 uses the @ symbol twice. The first '@' tells SAS to input the variable 'rectype' in the first column of data, the '1.' tells SAS that the variable 'rectype' is only one character in length, and the final '@' holds the line of data for the subsequent INPUT statements.

Line 5 is telling SAS to input the following column locations only if the variable 'rectype'= 1.

Lines 7-8 provide the column locations, variable length and names for the variables on the household record.

Line 9 tells SAS to stop the inputting when 'rectype' no longer equals 1.

Line 10 tells SAS what to do when 'rectype' equals 2.

Line 11-21 is more of the same for the family and person level records.

Line 22 tells SAS that we only want people who are married (marital=1).

Line 23 tells SAS to stop when it gets to the end.

Line 24 drops an unnecessary variable, rectype. We don't need it anymore.

Try it yourself! The raw data and the SAS program are located on the CFDR server:\Public\Practice Data\Spouse CPS

This is what the first few observations look like.

foodstamps	hhid	poverty	lineno	age	marital	spouse	sex	pob
2	3683	4	1	54	1	2	2	57
2	3683	4	2	59	1	1	1	57
2	4183	4	1	45	1	2	2	57
2	4183	4	2	47	1	1	1	57
2	4184	4	1	51	1	2	1	57
2	4184	4	2	46	1	1	2	57

Now that we have our person level file with only married people in it, we need to attach information on their spouse.

```
(25)      data d1 (keep=spouse age pob hhid);
(26)      set cps03;
(27)      data d2 (rename =(spouse=lineno age=spouseage pob=spousepob));
(28)      set d1;
(29)      proc sort data=cps03; by hhid lineno;
(30)      proc sort data=d2;by hhid lineno;
```

Line 25 creates a new dataset called 'd1' that only contains the household identifier (hhid), the line number of person's the spouse (spouse), age and the place of birth (pob).

Line 26 sets 'd1' equal to our original 'cps03' dataset.

Line 27 creates another dataset called 'd2', and renames several variables. In this dataset, the variable 'spouse' is now called 'lineno', the variable 'age' is now called 'spouseage' and the variable 'pob' is now called 'spousepob'.

Line 28 sets 'd2' equal to 'd1'.

Line 29 sorts our original 'cps03' file by household id (hhid) and respondent line number (lineno).

Line 30 sorts 'd2' the same way.

Here is what the 'd1' looks like:

hhid	age	spouse	pob
1	35	2	72
1	35	1	57
2	32	3	57
2	26	1	379
6	44	2	72
6	36	1	57

Here is what 'd2' looks like:

hhid	spouseage	lineno	spousepob
1	35	2	72
1	35	1	57
2	32	3	57
2	26	1	379
6	44	2	72
6	36	1	57

Lines 31-33 create a new 'spouse' file.

```
(31)      data spouse;
(32)      merge cps03 d2 ;
(33)      by hhid lineno;
(34)      atrrib spouseage label="Age of Spouse";
(35)      attrib spousepob label="Spouse POB"; run;
```

Line 31 creates a new dataset we call 'spouse'.

Line 32 merges together the original cps file 'cps03' and the one we created called 'd2'.

Try it yourself! The raw data and the SAS program are located on the CFDR server:\Public\Practice Data\Spouse CPS

Line 33 indicates the variables we merge on.
Lines 34 and 35 simply label the spouse variables.

Here is what the data look like:

food-stamps	hh id	poverty	lineno	age	marital	spouse	sex	pob	spouse age	spousepob
2	1	2	1	35	1	2	2	72	35	57
2	1	2	2	35	1	1	1	57	35	72
2	2	4	1	32	1	3	1	57	26	379
2	2	4	3	26	1	1	2	379	32	57
2	6	4	1	44	1	2	1	72	36	57
2	6	4	2	36	1	1	2	57	44	72