Creating Cohabiting Couples File Using the PUMS

The PUMS is a hierarchical file—there are lines of data for household information and lines of data for person information. There are identifiers for the household and the individual. Plus, there is a variable which indicates the relationship of each person living in the household to the person identified as the household head. Perhaps all we are interested in is obtaining information on different-sex cohabiting couples. To do this, we need to identify those people that are living in a household as the 'unmarried' partner of the household head. It is also important to determine whether the household head and the unmarried partner are not the same sex.

In this example, we use the 2000 five-percent PUMS file for Ohio. Below is the code used to input the raw data.

```
(1)   data pums;
(2)   infile 'C:\DATA\pums\pums5_39.txt'  ;
(3)   input rectype $ 1@;
(4)   if rectype = 'P' then do;
(5)   input          hhid 2-8          pid 9-10
      pweight 13-16  relate 17-18      sex 23        marstat 44;
(6)   end;
(7)   if relate in (1,19) then output; run;
```

Lines 1 through 7 are inputting the raw data.

Line 1 creates a temporary dataset called 'pums'.

Line 2 points to the location of the raw data file called pums5_39.txt'. Line 4 conditions the input so that we only get the person level records.

Line 5 provides the column locations.

Line 6 tells SAS to stop inputting once it reaches the last variable.

Line 7 tells SAS to output only those cases in which the respondent is either the household head (relate=1) or the unmarried partner (relate=19).

We need to count up those individuals living in a household who are considered the unmarried partner. To do this we need to create a new household level file called 'hh' that has one record for each household.

```
(8)   proc sort data=pums ;by hhid;
(9)   data hh (keep=hhid cohab);
(10)  set pums;
(11)  by hhid;
(12)  retain cohab;
(13)  if first.hhid then do;
      cohab=0;end;
(14)  if relate =19 then cohab=cohab+1;
(15)  if last.hhid then output;
```

Lines 8 through 15 are creating a household level file.

Line 8 Sort the data by the household id number (hhid)

Line 9 Create a new temporary dataset called 'hh' and only keep the household id along with a new variable which we will create called 'cohab'.

Line 10-11 sets the data set equal to the 'pums' data by Household ID.

Line 12 retains the variable 'cohab' across each person in the household.

Line 13 Tells SAS to look at the first household id and then execute the creation of the variable 'cohab'.
Line 14 Tells SAS to watch out for any case that has the value '19' on the variable 'relate', when the condition is met, 1 is added to the value of 'cohab'. Think of it this way, it is counting up all the people in the household that have the value 19 on the variable 'relate'.
Line 15 Tells SAS to output the information when the last household record is reached.

Now we need to combine the information located on both the 'pums' and the 'hh' datafiles to create a file that only contains the householder and their cohabiting partner.

```
(16)  data couple_pums;
(17)  merge pums hh;
(18)  by hhid;
(19)  sex=sex-1;
(20)  if cohab =0 then delete; run;
```

Lines 16 through 20 merge two files together.
Line 16 creates a new dataset called 'couple_pums'.
Line 17 merges the two data sets together.
Line 18 specifies the variable that the two data sets have in common.
Line 19 in the original data, the variable 'sex' is coded 1 for men, and 2 for women. We need it to be a dummy variable. So now, sex=1 for women, and 0 for men.
Line 20 Tells SAS if the variable 'cohab' equals 0 then delete if from the file. This gets rid of any people not living as a householder with an unmarried partner.

But what if there are same-sex couples in our file? To sort this out we need to use a PROC MEANS to create a dataset.

```
(21)  proc means noprint data=couple_pums;
(22)  by hhid;
(23)  var  sex;
(24)  output out=sumdat1 sum(sex)=sexcount; run;
```

Line 21 Tells SAS we want something from the data file 'couple_pums' but we don't want it printed in the output window. (we don't need it).
Line 22 The 'by' statement tells SAS that we want it to calculate the means for each household—that's why we don't want to print it—too much space.
Line 23 The variable that we want to focus on is 'sex'.
Line 24 This looks complicated but it isn't. The 'output' statement tells SAS that we want the information from the PROC MEANS to be put into a special little dataset called 'sumdat1'. And further, what we really want is the SUM of the variable sex and we want to call if 'sexcount'.

Think of it this way, for each household we are adding the values of sex. That is, if sex =1 for the household head and sex=0 for the unmarried partner then sexcount=1. If however, the value of sexcount=0 or 2 that means that both the household head and the unmarried partner are the same sex.

```
(25)  data final;
(26)  merge couple_pums sumdat1 (keep=hhid sexcount);
(27)  by hhid;
(28)  if sexcount =1 then output; run;
```

Line 25 tells SAS to create a new dataset called 'final'.

Line 26 tells SAS to merge together the dataset 'couple_pums' with the file called 'sumdat', and further to only keep the variables for household ID along with the variable 'sexcount' from the sumdat1 file.
Line 27 means that we are matching on household id (hhid).
Line 28 tells SAS to include only cases where 'sexcount'=1. That is, only those cases that are part of a different-sex cohabiting couple.