

Using Continuous Month Format in the NLSY97

In the NLSY97, many events are reported on a monthly basis. For example, information is collected on a respondent's school enrollment, labor force participation and program participation. Rather than having actual dates in the data file, the NLSY uses a continuous month format. The continuous month format can be thought of as simply a numbering system for dates. In the CM format, January 1980 is considered month 1, February 1980 is month 2, etc.. Say a respondent's birth date is September 1981—in the CM format this would be month 21. Other information in the data provides you with the respondent's age in months at the time of each interview, as well as the CM for the first interview. This information can be used to create a birthdate:

Birthday in CM= CM of first interview-age in total months at time of first interview

So a respondent, who is first interviewed in Continuous Month 211 (July 1997) and is 190 months old at the time of first interview, has a Continuous Month birthday of 26 (September 1981).

The event history file also has start and stop dates in continuous month format for employment, schooling and welfare use. You may want to create an array that identifies the appropriate month that the respondent turns *an exact age*. That is, you might want to sum up welfare use or employment history from the first month the child turns 15 to the last month the child turns 19.

In the following example, we create a series of variables that give the respondents age in months for each month in the survey. This example uses waves 1 through 7 event history file which has information on the first month of 1997 through the seventh month of 2004, for a total of 91 months.

```
(1)  array mage [*] agem1-agem91;
(2)  agem1= r1193900;
(3)  do i=1 to 90;
(4)  mage[i+1] = agem1 + i ; end;
```

Lines 1 through 4 create a series of age variables.

Line 1 creates an array called 'mage' which refers to a series of variables that we will be creating called agem1, agem2, agem3, etc. The [*] tells SAS how many variables there are going to be. The asterisk is an easy way to represent the number of variables in the array.

Line 2 makes the variable 'agem1' equal to the value of the variable 'r1193900'. This variable is the age in months of the respondent the very first time of interview. So for our respondent above, the variable agem1=190.

Line 3 tells SAS to do something 90 times (that is what the 1 through 90 means), beginning with the number 1 and ending with the number 90.

Line 4 tells SAS when 'i' is equal to 1, the new variable created is 'mage2' or mage (1+1) and it is equal to 'agem1 + 1'. The 'end' tells SAS to stop when it goes all the way from i=1 to i=90. So for our respondent the variable agem1=190.

The new variable created in the do loop is `agem2` which equals $190 + 1$; when `i=2` the new variable `agem3= 190 + 2`, and so on. Our respondent now has a series of variables that tell us how old he is at each month he is potentially in the survey.

But what we really want to know is what happens by an exact age—the last month that he is 19. The following example shows how we would find out if he had ever enrolled in college by the last month of his 19th year.

```
(5)  do i=1 to 91 until (mage[i] = 240);  
(6)  m240=i ; end;  
(7)  array enroll [*] enrollstatus1 – enrollstatus91 ;  
(8)  collegemonths=0;  
(9)  do i=1 to m240;  
(10) if enroll[i]=1 then collegemonths=collegemonths+1; end;
```

Lines 5 through 10 find the exact age and determine months of college enrollment.

Line 5 uses a 'DO UNTIL' statement. It looks tricky but it really isn't. The 'do i=1 to 91' tells SAS that the full size of the array is 1 through 91. The 'until' statement then says, 'Do i=1 to 91' until your case matches the condition in the parentheses. So the condition (mage [i] ge 240) is telling SAS to keep going until the age in months of the respondent is 240—which is the age of someone who is in their last month of their 19th year.

Line 6 says the variable 'm240' is equal to i. So when SAS is going through the numbers 1 through 91 and it gets to the condition where the respondent is in the last month of their 19th year it says take whatever index number it is on (i) and set m240 equal to that. For example, our respondent hits 240 during his 51st month in the survey, so m240=51.

Line 7 is an array that references a series of variables that indicate the respondent's enrollment status for each month of the survey. When the variable `enrollstat1=3`, the respondent is enrolled in college during that month.

Line 8 creates a new variable and sets it equal to zero.

Line 9 is another DO loop. This one tells SAS to do the following from 1 until the value for m245 is reached. So for our respondent, SAS is supposed to do this from 1 to 51.

Line 10 is creating a tally variable called 'collegemonths' which sums up all the months spent attending college from the time the respondent first enters the survey to the last month they are 19. If the variable `enrollstatus1 = 1` then `collegemonths= 0 + 1`; if the variable `enrollstatus2 = 1` then `collegemonths= 1+1`; and so on. If the variables referenced in the 'enroll' array are any value but 1, nothing is added to the value of 'collegemonths'.