

How do I create a woman-level file from a pregnancy level file?

Perhaps your data are structured so there is at least three pregnancies per woman:

<u>momid</u>	<u>pregid</u>	<u>outcome</u>	<u>weeks</u>
1	1	1	40
1	2	2	4
2	1	1	40
2	2	1	40
2	3	1	40
3	1	3	12
3	2	2	5
3	3	1	40

You can see that the first mother had 2 pregnancies, the second mother had three, etc.. What we want to do is have an individual record for each mother with all the data included for each pregnancy.

```
data mother(keep=momid  pregid1-pregid3  outcomel-outcome3
weeks1- weeks3 totalb);
set child;
by momid;
retain  pregid1-pregid3  outcomel-outcome3  weeks1-weeks3
totalb i ;
array p( 3 ) pregid1-pregid3 ;
array o( 3 ) outcomel-outcome3;
array w( 3 ) weeks1-weeks3;
if first.momid then do ;
totalb= 0 ;
do j= 1 to 3 ;
p(j)= . ;
o(j)= . ;
w(j)= . ;
end ;
i= 1 ;
end ;
p(i)=pregid;
o(i)=outcome;
w(i)=weeks;
i=i+ 1 ;
if outcome= 1 then totalb=totalb+ 1 ;
if last.momid then output ;
run ;
```

- Keep states what variables you want in you mom level dataset. If you do not list the variables here it will keep some extra ones, such as the index variables i and j or the original outcome and weeks variables which no longer function the same way.
- Retain holds the values over throughout the iterations of the data step.

- Array statements list the three arrays to be used for each of the pregnancies. The index number for the arrays has to be the highest number of possible pregnancies; in our examples it is '3'. (In a real dataset you would just run a frequency of the pregnancy id to find out.)
- If first.momid statement tells SAS to do the following as soon as you reach a new and unique mom id.
- Totalb is set to equal zero so we can begin a total count of each pregnancy.
- do j=1 to 3 tells SAS to initialize each of the new variables we are creating with the arrays p, o, and w to missing to start.
- p(i)=pregid tells SAS to set the first pregnancy to the value of the first array.
- i=i+1 moves the index forward.
- if outcome= 1 then totalb=totalb+ 1 creates a count of the total number of births.
- if last.momid tells SAS to output once it reaches the last pregnancy.

Here is what the output looks like

momid	pregid1	pregid2	pregid3	outcome1	outcome2	outcome3	weeks1	weeks2	weeks3	totalb
1	1	2		1	2		40	4		1
2	1	2	3	1	1	1	40	40	40	3
3	1	2	3	3	2	1	12	5	40	1