
Subject: Combining mother's info on child row in PR file

Posted by [lillo?S](#) on Mon, 05 Dec 2016 19:20:59 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hi, from the MLPR41FL file I am trying to paste the information of the mother on the child's row using a function that I've written in R. In other words, I find the mother in the household if available (idmother) and then copy the mother's age (agemother) for each member under 17 years old having a mother in the hh (idmother is not NA). I was checking the values of agemothers and I find children under 17 who have mothers who are 0 or 1 years old or even 89 or 90 years old, which does not make sense of course.

This is an extract of the dataset:

#1	302102	1	66	NA	1	NA	NA
#2	302105	1	74	NA	1	NA	NA
#3	302105	1	53	NA	2	NA	NA
#5	302105	1	14	2	4	2	53
#6	302105	1	10	0	5	NA	NA
#7	302105	1	0	3	6	3	0
#8	303	7	1	25	NA	1	NA
#9	303	11	1	40	NA	1	NA
#10	303	11	1	35	NA	2	NA
#11	303	11	1	12	2	3	2
#12	303	11	1	0	2	4	2
#13	303	13	4	75	NA	1	NA
#14	303	16	1	32	NA	1	NA
#15	303	16	1	16	NA	2	NA
#16	303	16	1	53	NA	3	NA

his/her mother (idmother) is 3, which is obs #4, who is 0 years old (hv105). Basically the function that I used to get the variable agemothers is doing its job, but there might be inconsistencies in the data.

Let's assume that women can have children from 10 to 60, there are 419 children who have mothers who are either below 10 or above 77.

I find the same in other surveys and for other years, thereby finding that a child has a mother who is 1 or 2 or even 89 or 90 years old.

Thanks.

Subject: Re: Combining mother's info on child row in PR file

Posted by [Trevor-DHS](#) on Tue, 06 Dec 2016 13:38:52 GMT

[View Forum Message](#) <> [Reply to Message](#)

This looks like you are mismatching the IDs for the cases. If you post your R code, we can take a look and try to resolve it.

Subject: Re: Combining mother's info on child row in PR file

Posted by [lillo?S](#) on Tue, 06 Dec 2016 14:21:15 GMT

[View Forum Message](#) <> [Reply to Message](#)

Sure and thanks. This is the code:

```
df <- df %>%
  group_by(hhid) %>%
  mutate(idmother = ifelse(hv112>0 & under17=="yes", hv112, NA))
df <- df %>%
  group_by(hhid) %>%
  mutate(agemother = ifelse(!is.na(idmother), hv105[idmother], NA))
```

Subject: Re: Combining mother's info on child row in PR file

Posted by [Trevor-DHS](#) on Wed, 07 Dec 2016 23:56:47 GMT

[View Forum Message](#) <> [Reply to Message](#)

When I looked at the data in detail, I do see the problem case that you pointed out, and I see some other cases with issues, but I do not see anything like as many as you do. I used (I think) a slightly tighter check than you did - Mothers aged less than 12 years older or more than 60 years older than the child - and I only found 37 cases (see attached file). I think there are three main types of problems:

- 1) Reporting of the correct person as the mother. If the line number of the mother is recorded incorrectly, then the age of the mother is going to look incorrect. This probably accounts for many of the problems, particularly those where the "Mother's" age is under 10, or the age difference is less than 10.
- 2) Some women are reported as much older than they are - there certainly appear to be a few of these cases, e.g. where the mother is reported as age 70.
- 3) Age of mother is reported too young, and the gap between the age of the mother and the child is too small. There are likely to be a few of these too.

I looked at a few other datasets, and I find a few cases of problems here and there, but not many. The Mali DHS 2001 has more than most, but I only identified 37 cases with my check.

Can you check your logic again and make sure you are not incorrectly identifying problem cases. Also, remember that 98 or 99 mean Don't know and Missing respectively for variable hv105, and should be excluded from the checking.

File Attachments

1) [Mothers_age.txt](#), downloaded 685 times

Subject: Re: Combining mother's info on child row in PR file

Posted by [lillo?S](#) on Thu, 08 Dec 2016 14:27:40 GMT

[View Forum Message](#) <> [Reply to Message](#)

I have checked using the age difference between the mother and the child. There are 82 cases in which this difference is 12 years and less and there are 5 cases in which this difference is 60 years and more (please see attached file). So, the 419 children that I mentioned in my first post are instead 87 -- my bad.

In the file you attached, the 37 cases include observations in which the age difference is 57 and 58 and 59, meaning that in the end the 'strange' cases are even less, i.e. 33. Moreover, in your file, it is specified that the Input Data used is: C:\Projects\Analysis\UserForum\Mothers age mismatch\temp\mlir41.dat, while I am using the mlpr41.dat. Why are you using that one?

Can I ask you which R code you used to copy/paste the mother's information? I struggle at understanding why I find more cases than you do.

File Attachments

1) [mothers_age.csv](#), downloaded 607 times

Subject: Re: Combining mother's info on child row in PR file

Posted by [lillo?S](#) on Thu, 08 Dec 2016 15:54:29 GMT

[View Forum Message](#) <> [Reply to Message](#)

I've just realised that I was not working on the full data because I had excluded observations for which the age was not available or missing before running the code to paste the mother's age. I have re-run the code using the full data and I get 49 cases for which the age difference is ≤ 12 and ≥ 60 (and not 37) and I get 39 cases for which the age difference is ≤ 11 and ≥ 57 (which I guess were the lower and upper limit that you applied to verify the number of 'strange' cases, the two additional cases are hhid: 145 15 4 and 182 60 1). Can you confirm these numbers?

I have two further questions. I would consider as 'strange cases' those cases for which the age difference is < 10 and > 60 and remove them from the data. In other words, I would be using only observations for which the age difference between the mother and the child is between 10 and 60 ($10 \leq \text{age difference} \leq 60$). Does this make sense or should I consider a different lower or upper limit?

If I had to do the same for the father, which would the upper and lower limit for the age difference between the father and the child be?

Thanks a lot.

Subject: Re: Combining mother's info on child row in PR file

Posted by [Trevor-DHS](#) on Tue, 03 Jan 2017 21:28:32 GMT

[View Forum Message](#) <> [Reply to Message](#)

I used a slightly different test. I first selected all children ($HV105 < 18$) with a mother listed in the

household (HV112 in 1:97), and then output any case with a mother's age of < 10 or ≥ 70 or a difference in age between mother and child of < 12 or > 60 years. I get 37 cases only. If I use ≤ 11 and ≥ 57 I match the 39 cases.

In terms of picking cut offs, that is really your choice, but I think those choices make some sense. For the father, I think you would have to go with a much higher cutoff at the top end as there are many cases of the being recorded as 70 years or more older than the child. However, at the low end you could probably use a minimum difference of up to about 15 years. The choice is really yours for your analysis.