



Anthony Damico &lt;ajdamico@gmail.com&gt;

---

## different record counts in tutorial #3

---

PSID Help &lt;psidhelp@umich.edu&gt;

Thu, May 23, 2013 at 3:36 PM

To: Anthony Damico &lt;ajdamico@gmail.com&gt;

I have answered your questions below each paragraph in red

thanks Noura :) ..so what you're telling me is that any tutorial created after 2007 is not reproducible? because it appears that all of the tutorials that use the panel were created before then? i am trying to write a blog post that explains to others how to work with the psid for longitudinal analysis.. is there any way you can move "fixing the tutorials" up on your priority list?? waiting another six months for everything to be fixed seems unreasonable.

There are differences between the tutorials that use individual level vs. family level variables. Those that use family-level variable do not have added case counts the way the individual level does. Also, as weights may be updated or variables change, we do update the tutorials. Tutorial #1, as an example, was last updated in June 2012.

in the meanwhile, is there any way to identify who has been added to the case counts? just throwing out 300 records for no reason other than "the spreadsheet said so" is not a feasible solution. maybe there's a pattern or identifier that we can use to get 1,633 observations that doesn't involve merging to the "answers" excel sheet?

The purpose of the tutorials is to get an idea of how to work with the data in different ways (cross-section, longitudinal, inter-generational). When you merge the data sets by using the first 2 columns of the 'answer sheet' (namely ER30001 and ER30002), you will go through the process of creating the balanced panel and doing the analysis that follows. Once you see that you have been able to replicate the results using the 1,633 rows, you can then follow the same steps with confidence, knowing that the results are accurate for the 1,924.

Also, merging the sets will actually show you the exact individuals who make up that ~300. It is not that you would not use all of the individuals in the sample for data analysis for personal research, but for the purpose of the tutorial cutting down to the 1,633 will allow you to check you answers.

alternatively, while we're waiting for the tutorials to be fixed, could you please reproduce this attached table using the correct numbers? that would give me something from umich researchers that shows my analysis code is working properly.. i am not asking for you to update the whole tutorial, just to provide me with that single table that i'll then be able to replicate.

the purpose of all of this is to reproduce your work with code in the R language, so just merging records on from the "answer sheet" does not work - i need to show other R users how they can get the correct answer by only using R code..

Though we use different programs in the tutorials (some Excel, some SAS), your personal replication can be shown if you go through the process of the tutorial with the 1,633, then the 1,924. You can first use Excel, then R.

Though it is true that there have been added rows to the excel sheet, the idea remains the same and the steps to follow in the tutorial are sound. The tutorials are meant to be a tool to show users the different ways in which the data can be used (by using the different identifiers and techniques).

I will certainly ask when we can expect Tutorial #3, or others, to be updated, or if someone will be able to replicate that table. I will let you know when I hear back.

Enjoy your evening,

Noura