
Subject: Append IPUMS_Merge to DHS Data
Posted by [Yawo](#) on Wed, 05 Jun 2019 23:50:28 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello, Thanks for your recent updates re IPUMS. I have a question about linking merged data produced by IMPUMS to similar dataset from the DHS (cross-posted on IPUMS forum on IDHS site).

I am examining HIV risk behaviors in sub-Saharan Africa, and will be using data from the following countries:

Angola
Benin
Burkina Faso
Burundi
Cameroon
Chad
Congo
Congo Democratic Republic
Cote d'Ivoire
Eswatini
Ethiopia
Gabon
Gambia
Ghana
Guinea
Kenya
Lesotho
Liberia
Malawi
Mali
Mozambique
Namibia
Niger
Rwanda
Senegal
Sierra Leone
South Africa
Tanzania
Togo
Uganda
Zambia
Zimbabwe

I plan to use IPUMS to create a pooled dataset containing selected/harmonized variables of interest. However, IPUMS currently does not have data for the following countries (unless I am mistaken):

Chad
Gabon
Liberia
Sierra Leone
Togo
Swaziland/Eswatini

If this is the case, my plan is to go ahead generate the pooled dataset using IPUMS. I have access to the data for these subset of six countries. How do I append them to the IPUMS harmonized version and ensure that it is truly harmonized - with variable and value/labels?

thanks - Yawo

Subject: Re: Append IPUMS_Merge to DHS Data
Posted by [boyle014](#) on Thu, 06 Jun 2019 23:37:07 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Yawo,

Great question. There are multiple ways to do this, but this is my recommendation. I would use the IPUMS DHS data first. Once you're happy with your variables and analysis, then add in The DHS Program samples.

To add in the other samples, first find The DHS Program variable names for the variables you're using. These appear at the top of all the documentation pages at IPUMS DHS. (You get to the documentation pages by clicking on a variable's name). For example, the IPUMS DHS variable "aidgetbite" is V754JP in the original DHS Program files. Download the six additional surveys from The DHS Program website and retain the relevant variables.

In Excel, create a sheet for each variable. In the first column, include the IPUMS DHS codes and value labels; put the IPUMS DHS variable name at the top. In the second column, include the Chad codes and labels with The DHS Program variable name at the top. In the third column, include the Gabon codes and labels, and so forth. Use the Excel sheets to recode the DHS variables to match the IPUMS DHS variables. Do this within each file. Then append the revised files to your IPUMS DHS file.

This should work. Just let us know if you have any more questions.

Subject: Re: Append IPUMS_Merge to DHS Data
Posted by [Yawo](#) on Mon, 10 Jun 2019 13:50:42 GMT

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

Dr. Boyle: good morning. Thanks very much for your response. I understand the process, and will execute it as advised.

I have a related question: do i have to go through de-normalizing the weights since I am pooling the data for recent surveys only?

And if so, how will these work?

Do I denormalize after I've merged both IPUMS and the DHS data or before/separately, and then merge them.

Thanks - Yawo

Subject: Re: Append IPUMS_Merge to DHS Data
Posted by [boyle014](#) on Wed, 19 Jun 2019 19:14:14 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Yawo,

Apologies for my delay. The question about denormalizing the samples depends on your research question.

For descriptive statistics aggregated to the country level, use the appropriate IPUMS DHS weighting variable (perweight for women; dvweight for the domestic violence module, etc.). If you want to create confidence intervals, you'll want to use the svyset command in Stata. There is no reason to denormalize weights for this type of research.

If you are running a regression analysis on the pooled data, you will want to use a multilevel or mixed model, or possibly use country or region fixed effects. Weights are less essential since you're controlling for the characteristics (like urban/rural) that went into the construction of the weights.

If you want to calculate the total number of women in all of these countries who are experiencing something, then you need to apply the regular weights and inflate the sample sizes so that they reflect the number of women aged 15-49 in each country. In other words, this is when denormalization is necessary.

Let me know if you have more questions. Feel free to share your research question if you want more specific guidance.

Subject: Re: Append IPUMS_Merge to DHS Data

Posted by [Yawo](#) on Sun, 23 Jun 2019 20:50:26 GMT

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

Thanks very much, Dr. Boyle: I appreciate your assistance. I will keep you updated on my progress.

best wishes - yawo
