
Subject: IPUMS-DHS Tip #1. Applying weights to IPUMS-DHS data files in Stata
Posted by [boyle014](#) on Wed, 16 May 2018 19:40:02 GMT

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

Applying weights to IPUMS-DHS data files in Stata

Nearly all DHS use a multistage cluster survey design, which calls for more than the simple application of probability weights. Here, we explain the basic process for IPUMS-DHS users to weight their analyses. These steps will work for either a single sample or pooled samples.

For standard weighting of full samples of women, children, or births, make sure your IPUMS-DHS extract includes the following variables:

PERWEIGHT, which is V005 in the standard DHS files, but does NOT need to be divided by 1,000,000. PERWEIGHT is automatically added to every IPUMS-DHS data extract.

IDHSPSU, which corresponds to V021 in the standard DHS but is unique for each sample, is primary sampling units or PSUs.

IDHSSTRATA, a cross-sample version of V022, is the sampling strata.

Because IDHSPSU and IDHSSTRATA are unique cross-sample identifiers in IPUMS-DHS, they will work with single samples or pooled samples.

You will find IDHSPSU and IDHSSTRATA in the drop-down menu under TECHNICAL -> IDENTIFIERS.

Before working with DHS data in any form, we highly recommend that you review the excellent videos that ICF has created, keeping in mind the unique IPUMS-DHS variable names above (PERWEIGHT, IDHSPSU, and IDHSSTRATA):

Part 1: Introduction to DHS Sampling Procedures

Part II: Introduction to the Principles of DHS Sampling Weights

Part III: Demonstration of How to Weight DHS Data in Stata

Part IV: How to Weight DHS Data in SPSS and SAS

To weight IPUMS-DHS data in Stata, the command is:

```
svyset [pw=perweight], psu(idhspsu) strata(idhsstrata)
```

This establishes the weights in Stata; they are then applied to relevant commands by putting "svy:" at the beginning, such as:

```
svy: regress y x  
svy: mean(y), over(x)
```

This brief introduction to weighting with IPUMS-DHS will work for most standard analyses. As you become more familiar with DHS data, we encourage you to review other User Forum threads on more complex weighting topics, such as how to handle missing DHS strata, working with DHS subsamples (e.g., the domestic violence module) or your own subsample (e.g., children under 12 months old), denormalization, the limitations of the svyset command in Stata, etc.

From time to time, we will post more detailed information about these topics. Always feel free to post your own tips, suggestions, or questions about using IPUMS-DHS here as well.
