

Guidance on Analyzing DHS Data

This task requires a basic ability/familiarity with analyzing large datasets, using standard software such as Stata, SPSS or SAS. The Demographic and Health Surveys (DHS) Program website offers help in [getting started](#) with downloading and analyzing the data. Registration is required in order to access the datasets. This guidance assumes that the dataset has been downloaded, and that Stata is the statistical software of choice.

Once the Stata files have been [downloaded](#) from the DHS website, you can begin analyzing trends in the use of the private health sector for the country of interest*. Depending on the scope of the assessment, there are several different variables that might be of interest. Below are the key variables used for a private health sector assessment:

Demographics

- Region: v024
- Urban/Rural: v025
- Wealth Index: v190
- Marital Status: v501

Reproductive Health/Family Planning

- Source of Contraceptives for Current Users: v326

HIV/AIDS

- Source of Last HIV Test: v829
- Place where HIV Test Taken as part of ANC: v842

Maternal and Child Health

- Source of Antenatal Care: m57
- Source of Delivery: m15
- Source of Postnatal Care/1st Infant Check-Up: m73
- Source of Treatment for Fever/Cough (children under 5): h32
- Source of Treatment for Diarrhea (children under 5): h12

You may find it easier to generate new variables that allow you to aggregate the data. For instance, the MCH variables are all disaggregated by child. So for instance, you might find m15_1, m15_2, m15_3, etc. in the variable list. These correspond to the source of delivery for the respondent's most recent birth, second most recent birth, and so on. Some variables, the source of antenatal care, are even more disaggregated. In your variable list, you will see m57a_1, m57a_2, m57b_1, m57b_2, etc. Again, the numbers refer to the birth order. The letters refer to the facility type. In this instance, creating a new variable using the gen command will make your analysis easier. You can do this as follows:

1. `gen variablename=.` ← This will create a new variable with the name “variablename”. All entries for this variable will be blank.
2. `replace variablename=1 if m57a_1==1|m57b==1` ← This will recode “variablename” so that all entries in which the respondents gave a positive answer for m57a and m57b will now be coded as 1. In this way, you can group all of the women who attended a public facility for their most recent birth into one code. You can repeat this as many times as you want to, replacing “variablename=1” with “variablename=2” or “=3” or any number as necessary.
3. `label var variablename “Variable Name”` ← This command names the variable. For example, when recoding m57, you would put Source of Antenatal Care in the parenthesis.
4. `Label define variablename 1 “XXX” 2 “YYY” 3 “VVV”` ← This will define the variable that you just created. So if you have grouped all of your public facilities into the “variablename=1” group that you made in step 2, you would write 1 “Public Facilities”.
5. `Label values variablename variablename` ← This command associates variablename with the label variablename. This will help when you ran your analysis later, so that the definitions you came up with Step 4 show up instead of just the numbers that you entered.

Once you have finished recoding your data, it is time to run your analysis. You can first create some simple descriptive statistics that summarize the data using the “tab” command. This will tell you the frequencies for specific variables. You can also run survey analysis using the “svy” command.

Examples:

`tab v829` ← tabulates how respondents answered the question about source of last HIV test

`tab v829 v024` ← tabulates how respondents answered the question about source of last HIV test, broken down by region

`svy: proportion v829` ← produces the proportion of respondents who gave the various responses about the source of last HIV test

`svy: proportion v829, over(v024)` ← produces the proportion of respondents who gave the various responses about the source of last HIV test, disaggregated by region

TIP > If you are performing survey analysis, make sure to weight the data so that you get more accurate results. First, generate a new variable for the weight (`gen weight=v005/1000000`). Then use the `svyset` command to tell STATA the primary sampling unit (generally v021), the weight variable (weight), and the stratified sampling unit (generally v022) by entering the command: `svyset v021 [pweight = zweight], strata(v022)`

You can use these simple commands to conduct a wide range of analysis that will strengthen your understanding of the country’s health system and the public-private mix for priority services.

*Note DHS data is available for many, but not all developing countries.