Stata Homework

# Download DHS Data

For this exercise you will need to download the “individual recode” dataset from the 2006 DHS survey in Mali. Due to DHS data sharing policies, we can’t directly share this dataset with you, but requesting and downloading this dataset if straightforward. Please note that it takes about a day for DHS to approve requests for access to data. If you are unable to access this data, you can download a sample dataset from DHS [here](https://dhsprogram.com/data/Download-Sample-Datasets.cfm) but answers you obtain from this dataset will obviously differ from that of the real dataset. DHS is one of the most useful sources of data in international development so it’s useful to create an account for future access. To request this dataset, complete the following steps:

## Register for DHS and request data access

1. Go to [www.dhsprogram.com](http://www.dhsprogram.com)
2. Click on “login” and then “register for dataset access”
3. Fill out the required information. Note that you must write a description of what you will use the data for that is at least 300 words. It is acceptable to state that you will use the data to generate descriptive statistics and that part of the reason for downloading the data is to learn how to use Stata to analyze DHS data.
4. Select “Sub-Saharan Africa” for region
5. Click “select all” to request access for all of these countries. (Note: If you are only interested in the Mali data, you can just click Mali but it is useful to request access to other countries’ data at the same time. You might even request access to datasets from other regions.)

## Download Mali 2006 Individual Data

Once your access to the DHS dataset has been approved, do the following to download the data:

1. Login to www.dhsprogram.com
2. Select the project for which you wrote a description when you requested access
3. Select “Mali” for the country and click “view surveys”
4. Click on “Mali 2006”
5. Click on “Data available” under the heading **Survey Datasets**
6. Click on [mlir53dt.zip](https://dhsprogram.com/customcf/legacy/data/download_dataset.cfm?Filename=MLIR53DT.ZIP&Tp=1&Ctry_Code=ML&surv_id=276) to download the individual recode file in Stata format

# What’s in this Dataset?

This dataset contains responses from a survey of women between the ages of 15 to 49 conducted in Mali in 2006. To obtain the final report from this survey, which includes a copy of the complete questionnaire, click [here](http://dhsprogram.com/publications/publication-fr199-dhs-final-reports.cfm).

# Open the Dataset

Start by opening the dataset in Stata. Depending on the version of Stata that you are using you may get an error if you try to open the dataset by double clicking on it. The reason for the error is that some versions of Stata have a limit on the number of variables that you can have in a dataset. If you receive an error, open the dataset using the following command (after replacing **PATH** with the path to the file on your computer):

use caseid v\* using "**PATH**\MLIR53FL.DTA"

This will load only the variables caseid and all variables starting with the letter “v” into memory. For the purpose of this assignment, these variables will be sufficient. (Note that you will only rarely have to use this command.)

# Questions

Use the dataset that you have opened to answer the following questions. If you want to check your answers later, either create a do file with the Stata code (recommended) or cut and paste the Stata commands into a word document.

1. How many respondents in the dataset have had 10 or more children (v201)?
   1. Commands to use: count

Answer:



1. Generate a variable for whether the respondent has 5 years of education (v133) or more and (also) has a television (v121). How many such households are there in the Mopti region?
   1. Commands to use: generate, replace, count
   2. Hints: You will need to use “if” conditions multiple times to do this. Use the variable v024 for region.

Answer:



1. Create table showing usage of family planning (v302) by wealth quintile (v190). (Output should look like the table below.)
   1. Commands to use: tabulate

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Never used** | **Used on folkloric** | **Used only trad.** | **Used modern method** | **Total** |
| Poorest |  |  |  |  | 100% |
| Poorer |  |  |  |  | 100% |
| Middle |  |  |  |  | 100% |
| Richer |  |  |  |  | 100% |
| Richest |  |  |  |  | 100% |

Answer:



1. Show the average and median number of children (v201) by wealth quintile (v190) for women using a modern means of family planning. (Output should look like the table below.)
   1. Commands to use: tabstat
   2. Hint: don’t forget to use an “if” condition so that the stats are only for women using a modern means of family planning

|  |  |  |
| --- | --- | --- |
|  | **Mean # children** | **Median # children** |
| Poorest |  |  |
| Poorer |  |  |
| Middle |  |  |
| Richer |  |  |
| Richest |  |  |

Answer:



1. Create a “do” file which does the following:
   1. Opens the Mali DHS data
   2. Generates a new variable, “agecategory” with the following values:
   3. Creates a table which displays usage of contraception method by value of agecategory. The table should look like this

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Never used** | **Used only folkloric** | **Used only traditional** | **Used modern method** | **Total** |
| agecategory == 0 |  |  |  |  | 100% |
| agecategory == 1 |  |  |  |  | 100% |
| agecategory == 2 |  |  |  |  | 100% |
| agecategory == 3 |  |  |  |  | 100% |
| agecategory == 4 |  |  |  |  | 100% |

Answer (not including line to open dataset):



1. Optional: Revise the answers above to take into account the sampling weight for the dataset so that the results are representative of the Mali population as a whole.[[1]](#footnote-1)
   1. Hint: The commands will be almost identical. You should just use the “weight” option for each command.

1. Note that the estimates you obtain after taking into account the weights will be representative of the total population but the standard errors and confidence intervals for these estimates will still be very slightly off. To obtain accurate confidence intervals you must take into account the complete sampling design including stratification. [↑](#footnote-ref-1)