WEBSITE EXERCISES USING IPUMS-DHS DATA FOR UGANDA

These exercises are designed to illustrate the many tools available in IPUMS-DHS that facilitate research with DHS data.

Part I. Exploring IPUMS-DHS

Getting started

On the IPUMS-DHS website, at **www.idhsdata.org**, click "Source Documents" in the left-hand column. Click on the link to the Uganda 2011 report and download the Final Report for that year. The Final Reports are an important source of information and provide a way for researchers to check that they are handling the data appropriately. Go to Table 3.1 on page 29 of the report.

Now return to the IPUMS-DHS home page and click on "Select Data" at the top of the page.

Chose "Women" as the unit of analysis. Click on "Select samples," check the box next to Uganda, and click on "submit sample selections." Now you will only see variables and documentation relevant to the Ugandan data.

Data discovery, variable names, composite coding, frequencies, weights

Find the RELIGION variable, using either the "Search" function or the drop-down Topics menu. (Hint: RELIGION is treated as a "Demographic" variable.)
   a. Click on "Original DHS variable names" at the top of the screen. What is the original DHS name for RELIGION? Now click back to IDHS variable names.

Click on the name of the variable, RELIGION, to pull up its documentation. There are a series of tabs across the top of the screen.

   b. On the "Codes" tab, you will see the Codes and Frequencies for the RELIGION variable. An "X" means that a given value appears in a sample; a "." means a value is not found in a sample. Identify a category of religion that never appears in the Ugandan samples. What is a category of religion that always appears?
   c. We have created a composite code for RELIGION, which facilitates recoding while preserving all information. What is the first digit for all Christian religions?
   d. Click on the "Survey text" tab. How are Seventh Day Adventists referred to differently across the surveys?
   e. Now return to the "Codes" tab and choose "Case Count view." Compare the unweighted frequencies for this variable on the IPUMS-DHS website to the unweighted frequencies (in the fourth column, headed Women unweighted frequencies) shown on Table 3.1 of the Final Report that you downloaded earlier. Do they match? Compare the unweighted numbers in column 4 to the weighted numbers in column 3. Briefly, how does weighting affect the results?

f. Click on "Select Data" at the top of the page to return to the drop-down list of topics. Select the "Technical" variables. What is the name of the variable you would use to weight the RELIGION variable appropriately, to produce the weighted frequencies shown in Table 3.1? This variable will automatically be included in any dataset you create in IPUMS-DHS.

Comparability across samples

Now we are going to explore the URBAN variable. Click on the "Change Samples" box and add the samples for Zambia (along with the Ugandan samples). Now find the URBAN variable (Click on Select Data at the top of the page, then use either the drop-down menu of Geography variables or the Search tool). Click on URBAN to access the documentation for this variable.

a. Look at the ratio of urban residents to rural residents in the (unweighted) frequencies for Uganda 2011 and Zambia 2013. Which country appears to be more urbanized?

b. Now click on the Comparability tab for URBAN and read how "urban" is defined for Uganda versus Zambia. Does the difference in how "urban" is defined for Uganda and Zambia suggest that the unweighted frequencies for URBAN understate or overstate the difference in urbanization in the two countries? Briefly justify your answer. Then return to sample selection, uncheck Zambia, and limit your samples to just those from Uganda.

Finding the most appropriate variable for your analyses

The figures on Education in Table 3.1 of the 2011 DHS Final Report are based on the EDUCLVL variable in IPUMS-DHS. Bring up the variable description for EDUCLVL. (Hint: Education variables are considered indicators of Socioeconomic Status.)

a. Consult the variable description to answer the question, what does EDUCLVL report? Is it an indicator of the highest level of education the person **attended** or the highest level she **completed**? For use in a later part of these exercises, please write down the (unweighted) number of women with no education, according to the EDUCLVL variable, in the 2011 Ugandan DHS sample.

b. From the drop-down list of variables relating to Education, find an alternative variable that you could use to determine how many women *completed* primary school. Which variable would you use to study completion rates within levels of schooling?

c. Sometimes analysts want to conduct a multivariate analysis, to study how, for example, each additional year of schooling by the mother affects a child's chance of survival or height-for-age. Which IPUMS-DHS education variable could you use like this in a regression equation?

Learn how complex variables were constructed

Read the variable description for WEALTHQ (Hint, found in "Household characteristics"}. What is the wealth quintile variable based on: the household's take-home wage and salary income or on something else? Explain briefly in your own words.

Explore IPUMS-DHS

Select the samples that interest you and explore which variables are available for them. Develop a research question that involves one or more dependent variables and several independent variables. Write out your ideas and explain why your question is important. You will use this material in the next exercise.

Part II. Registering to use DHS data in research projects.

The DHS data are free, but they require permission to use. If you already have permission, then you are all set. If you do not already have permission or want to extend your permissions to more samples, here are the instructions.

1. Go to http://dhsprogram.com/data/new-user-registration.cfm.

2. Enter your personal information. Please use the same email address that you used when registering for the IPUMS-DHS training.

3. For **Title of proposed study**, enter the topic and location of your research question, for example, "Family Size Preferences in 21st Century Uganda."

4. For **Brief Description of this Study**, describe your analysis. You might consider using the following template, filling in the brackets with your personal information and research.

> *Through IPUMS-DHS training at Makerere University, I have become interested in using DHS data for a research project. I am a [DESCRIBE YOUR POSITION AND AFFILIATION], and my previous research has addressed [TOPIC]. Using DHS data from [COUNTRIES], I plan to study [RESEARCH QUESTION].*

If you haven't reached 300 characters after filling in the brackets, add a sentence: "I plan to use [LIST VARIABLES] to answer my research question."

5. Under the section, **Please select a region to display the countries for which you want to request datasets**, from the drop-down menu, select the relevant region, which is likely "Sub-Saharan Africa." (Besides 19 countries in Sub-Saharan Africa, IPUMS-DHS also includes Egypt, which is available under "North Africa/West Asia/Europe," and India, which is available under "South & Southeast Asia"). Check the countries that interest you. Click "Save selections" at the bottom of the screen and move to another region or click "Submit Registration and Dataset Requests." It usually takes a day or two to acquire access. Keep trying at the IPUMS-DHS website. If you haven't received access within a week, email Bridgette Wellington, Bridgette.Wellington@icfi.com, to see if your application got overlooked.

Part III. Antenatal care in Uganda: exploring universe and sampling differences using IPUMS-DHS

The One.org blog recently reported, "Another Victory for vaccines: Uganda eliminates tetanus." The story continues, "According to the news release from UNICEF, 'Between 2002 and 2009, 25 high-risk districts in Uganda were targeted for intervention, and close to two million women of child bearing age received three doses of tetanus vaccines in those areas ... In 2010, Uganda reported it had eliminated the disease - and this year, a validation survey has taken place, confirming Uganda's elimination campaign has been successful."

Our next exercises focus on variables that can be used to study improvements in antenatal care in Uganda over time. The antenatal care variables for women in IPUMS-DHS currently relate only to the woman's most recent birth. Let's switch to children as the unit of analysis to see the information available for all recent births.

If you are already in the Select Data browsing area, you should see a gray bar above the topics list, saying "Currently browsing [unit of analysis]," with a link to "change." Click on "change" and then click on the blue box saying "Children: Each record will be a child under age 5." Now the drop down topics menu will show variables relating to young children (or their mother's characteristics). The gray box should read, Currently browsing "Children."

Click on the "Select Samples" box to limit the information to Uganda only for this new unit of analysis.

From the drop-down list of topics, select "Child antenatal and delivery care" and then "Child antenatal care."

      a. What variable could you use to track progress in vaccinating pregnant women against tetanus, from 1995 forward?

      b. How does the universe for ANTETNUSNO vary across years? (Check the "Universe" tab in the documentation for this variable). Do you think that difference matters in studying change over time in this case? Why or why not?

      c. Let's examine another indicator of prenatal care, ANCARECOM, about whether the mother was told about pregnancy complications. Does the ratio of "yes" to "no" responses in the (unweighted) frequencies for ANCARECOM indicate improvements in this dimension of antenatal care in Uganda? Explain.

      d. Access to medical care, including antenatal care, often varies by socioeconomic status. Perhaps you'd like to see whether children born to mothers with no education are less like to have gotten antenatal care. Keeping children as your unit of analysis, find the variable EDUCLVL (Highest education level of the mother). For Uganda 2011, the (unweighted) codes and frequencies show 1,427 women with no education when children are the unit of analysis, compared to 1332 women with no education when women are the unit of analysis (the number you retrieved earlier). Why is there a difference in the numbers? Which women (mothers) of childbearing age are not included in the figures on women's education when children are the unit of analysis? Which women are counted more than once when children are the unit of analysis?

Part IV. Making a data extract

Researchers have used CONUSMAN1 (V761, Condom used during woman's most recent sexual intercourse) as the dependent variable in studies about predictors of condom use by Ugandan women, based on DHS surveys taken before 2011. For a class exercise, we are going to make a customized data file with material from the 2011 Ugandan DHS, with variables that might predict recent condom use for Ugandan women. We will use some of the variables used in the following study: "Condom Use in Uganda and Zimbabwe: Exploring the Influence of Gendered Access to Resources and Couple-Level Dynamics," by Zubia Mumtz, Emma Slaymaker, and Sarah Stalway, in A Focus on Gender: Collected Papers on Gender Using DHS Data, published by USAID, 2005 and edited by Sunita Kishor. Time doesn't allow us to fully reproduce the results of the study, but by creating a data file that includes some of their dependent and independent variables, we are set up to conduct a similar analysis.

Step 1: Log in
Click on the Login link at the top of the IPUMS-DHS home page.

You should use your own DHS account after you receive e-mail confirmation of your own approval to download DHS data using your individual email and password.

Step 2. Choose the unit of analysis
Choose the unit of analysis that best fits your research topic. For this workshop exercise, we want to know what characteristics are associated with condom use by women during their last sexual encounter. Thus, we should pick Women as the unit of analysis. Each line of data in your customized (extract) file will represent one woman, and the variables displayed will relate to women of childbearing age.

Step 3. Select samples
Once you have logged in and chosen the unit of analysis, go to "Sample selections." All the African samples in IPUMS-DHS should now show up as green, meaning you can make a data extract (a customized data file for analysis) from any of those samples.

Check the box next to Uganda 2011 only, and submit your sample selection. Now you'll see information on only the variables available for the Uganda 2011 sample.

Step 4. Select variables of interest

Making a customized data file from IPUMS-DHS is somewhat like shopping online, except you are choosing samples and variables--and everything is free! Now we'll start adding variables to our "data cart," to make a data extract.

> **Dependent variable**: *CONUSMAN1* (Condom used during woman's most recent intercourse, in the "Condom Use" group of variables). Click on the yellow circle to the left of that variable

name, to add the variable to your data cart, or click on the "Add to cart" box within the variable description.

Independent variables:
For education: *EDUCLVL* (Highest educational level, in the "Socioeconomic status/education" group of variables).

For employment: *CURRWORK* (Woman currently working, in the "Socioeconomic status/education" group of variables)

For exposure to sources of information: *NEWSAMT* (Frequency of reading newspaper or magazine, in the" Socioeconomic status/media exposure" group of variables)

For HIV-related knowledge, in the "AIDS Knowledge" variable group:
AIDSHEARD (Heard of AIDS)
AID1PARLOWRYN (Thinks having only 1 sex partner reduces AIDS risk, yes/no)
AIDCONLOWRYN (Thinks always using a condom reduces AIDS risk, yes/no)

For Marital status, to distinguish between "married" and "living together":
MARSTAT (Woman's current marital or union status, in "Demographic/Marriage and cohabitation" variable group)

For patterns of decision-making, in the "Decision making" variable group:
DECBIGHH (Final say on making large household purchases)
DECFAMVISIT (Final say on visits to family or relatives)
DECFEMEARN (Final say on spending woman's earnings)
DECHUSEARN (Final say on spending husband's earnings)

For attitudes toward wife-beating (in "Domestic violence/attitudes" variable group)
DVAARGUE (Attitude toward wife-beating: Justified in woman argues with him)
DVABURNFOOD (Attitude toward wife-beating: Justified if woman burns food)
DVAGOOUT (Attitude toward wife-beating: Justified if woman goes out without telling him)
DVAIFNOSEX (Attitude toward wife-beating: Justified if woman refused to have sex)
DVANEGKID (Attitude toward wife-beating: Justified if woman neglects the children)

For acceptability of woman's refusal to have sex:
NOSEXOTHWF (When okay to refuse sex: Husband has other women, in Sexual attitude variable group)

When you have added all these variables to your data cart, you should have 18 variables and 1 sample in your data cart. Click on the green "view cart" box.

Step 5. Review and/or modify your data cart

After you click the "view cart" box, you will see the list of variables you added to your data cart. You will also see other variables that the system automatically adds to your extract to handle weighting (e.g., PERWEIGHT), sample identification (e.g., SAMPLE), variables used for linking (e.g., HHID), and variables used for variance estimation (e.g., PSU).

When you have finished looking at the list of variables and are satisfied, click on the green "create data extract" box. If you had decided to add another variable at this point, such as "AGE5YEAR" for woman's age, you could click on "Add more variables" and modify your extract request.

Step 6. Submit data extract request

After you click on "Create data extract," you are brought to another page, which allows you to specify the type of file you want. For example, if you want an SPSS file, click on "Change" next to "Data format" and then click on the box in front of SPSS.

Next, click on "Submit." Then fill in a description in the box headed "Describe your extract," include YOUR NAME, so you can identify your own extract.

When finished, click on the green "Submit extract" box.

Step 7. Check for your data extract (and download the data, if time permits)

Within a very few minutes, your data file will be available for downloading to your computer. Keep refreshing the page until it appears in the "Formatted Data" column. Click on the link when it appears to begin your data download. IPUMS-DHS files tend to be much smaller than the original DHS files, which makes downloading and file management much easier.

You can return to the Download or Revise Extracts page by clicking on "My Data Extracts" link on the IPUMS-DHS home page. To add or delete samples or variables from your tailored data file, return to your data extracts page and click on "Revise." Then click "Change" next to Samples, Variables, or Data Format.

Thank you!

Thank you very much for participating in the IPUMS-DHS Makerere Workshop. You have learned how to navigate the IPUMS-DHS website and how to create your own personalized data file. We hope that you will find these tools useful in your teaching and research, and that you will let others know about them.