IPUMS Data Training Exercise:

An introduction to IPUMS DHS

(Exercise 2 for Stata)

Learning goals

- Understand how IPUMS DHS dataset is structured and select between IPUMS DHS variables and units of analysis to explore research questions
- Create and download an IPUMS data extract
- Decompress the data file and read the data into a statistical package

Summary

In this exercise, you will look at various indicators of child demographics and children health, and gain basic familiarity with the IPUMS DHS data exploration and extract system. You will create a data extract that includes the variables KIDSEX, KIDALIVE, KIDTWIN, KIDLIVESWITH, CHDESIRE, IDEALKIDGRP, URBAN, DIARRECENT, DIATRENONE, and DIAGIVORS; then you will use the sample code to analyze the data. While the maternal

health indicators (women as unit of analysis) in IPUMS DHS relate only to recent last births (_01), the child data relate to all children born in the last 3-5 years before the survey.

Stata Code to Review

Code	Purpose
generate	Creates a new variable, "replace" specifies a value according to cases
mean	Displays a simple tabulation and frequency of one variable
tabulate	Displays a cross-tabulation for up to 2 variables
!=	Not equal to

Common Mistakes to Avoid

- Not changing the working directory to the folder where your data is stored.

- Mixing up = and ==; to assign a value in generating a variable, use "=". Use "=="
  to specify a case when a variable is a desired value using an if statement.

- Not using svy suite of commands for appropriate variance estimation

Register as a DHS Program User

Whether you make your own extract or use a pre-made file, you will need to log in using an email and password approved by the Demographic and Health Surveys Program to access data from Benin and India.

- Go to https://dhsprogram.com/, click on Login. On the login screen, click on
  Register for Dataset Access. Complete and submit the registration form. Please note
  that The DHS Program reviews user applications, and may take several business
  days.

Make a Data Extract

If you make your own data extract, which requires a statistical package and software to unzip a file on your own computer, follow the directions below. If you use a premade data

file, just note the variables included and you may begin work on the exercises. Answers are at the end of the handout.

- Navigate to the IPUMS DHS homepage (https://www.idhsdata.org/idhs/), login, and click on "Select Data." Select women as the unit of analysis.

Select Samples

- Click on the "Select Samples" box and check the following boxes:
  - o Egypt 2008
  - o India 2005
  - o Nigeria 2013
- Submit your sample selections by clicking the Submit sample selections box.
- Note that by selecting samples first, you will now only see variables available for Egypt 2008, India 2005, and Nigeria 2013.

Select Variables

- The variable drop-down menus allow you to explore variables by topic. For example, you might find variables about education under the "Socioeconomic status" topic.
- The search tool allows you to search for variables. Observe the options for limiting your search results by variable characteristics.
- Add a variable to your cart by clicking on the plus sign in the "Add to Cart" column of the topical variable list, or list of search results.
- View more information about the variable by clicking on the variable name, and navigating through the tabs that include a description of the variable, codes, and value labels, the universe of persons asked the question, and information on the comparability of the variable, among other pieces of information. If you are viewing variable-specific information, you may click on the "Add to Cart" button near the top of the screen to add this variable to your data cart.
- Use the Search tool to identify the following variables. Click on the plus sign in the "Add to Cart" column to the left of the following variables to add them to your data cart:

KIDSEX (Sex of child)

KIDALIVE (Child is alive)

KIDTWIN (Child is twin or single birth)

KIDLIVESWITH (Child lives with female respondent or others)

CHDESIRE (Whether and when this child's pregnancy was wanted)

IDEALKIDGRP (Mother's ideal number of children, with results grouped)

URBAN (Urban-rural status)

DIARRECENT (Child had diarrhea recently)

DIATRENONE (Whether no advice or treatment sought for child's diarrhea)

DIAGIVORS (Child given oral rehydration for diarrhea)

Review and submit your extract

- Click on the "View Cart" button underneath your data cart.
- Review your variable and sample selection to ensure your extract is complete.
  - Your data cart should show 9 variables and 3 samples. If you use a ready-made data file, these 3 samples and variables are included.
  - You may notice a number of additional variables you did not select are in your cart; IPUMS preselects a number of key technical variables, which are automatically included in your data extract. These variables indicate the sample (SAMPLE), are needed for variance estimation (PSU, STRATA), allow you to link to other DHS files (CASEID, HHNUM, DISTRICT), and are used for weighting the variables (PERWEIGHT, DVWEIGHT).
- Add additional variables or samples if they are missing from your extract, or click the "Create Data Extract" button.
- Review the Extract Request screen that summarizes your extract. For data format, choose Stata. Add a description of your extract (e.g., "Child health and demographics in Egypt (2008), India (2005), and Nigeria (2013))" and click "Submit Extract."

- The email account used to log in is sent a message when the data extract is ready to download.

Getting the Data Into Your Statistics Software

The IPUMS DHS extract builder provides raw ASCII data files and the command files necessary for reading the raw data into a stats package. Note that these instructions are for Stata.

Download the data

- Follow the link in the email notifying you that your extract is ready, or by clicking on the "My Data Extracts" link on the left-hand side of the IPUMS DHS homepage.
- Right-click on the data link next to the extract you created.
- Choose "Save Link As..."
- Save into your preferred working directory. This tutorial assumes you will save the file into "Documents" (which should pop up as the default location).

Decompress the data

- All IPUMS extracts are compressed. We recommend 7zip for Windows users. Macs can open these types of files without additional software.
- Find the "Documents" folder under the Start menu.
- Double click on the ".dta.gz" file.
- In the window that pops up, press the "Extract" button.

Read in the data

- Open the idhs_0001.dta folder.
- Double click on the file to open it in Stata.

Analyze the Data

Analyze characteristics of Egypt 2008, Nigeria 2013, and India 2005 child health and demographics.

Weighted frequencies

We will be comparing children under age 5 in the India 2005, Nigeria 2013, and Egypt 2008 samples (with the Egyptian sample limited to children born to ever-married women) and adding weights..

Note that the weight PERWEIGHT supplies numbers equal to the total number of young children in the survey, not the total number of young children in the population. (Ignore the numbers after the decimal point in the weighted results).

1. Is it rarer for children in these samples to be twins/non-singulate births or for surviving children to live apart from their (living) mothers?

    a. India 2005: _____

    b. Nigeria 2013: _____

    c. Egypt 2008: _____

```
tab kidtwin sample [aw=perweight], col
```

```
tab kidliveswith sample [aw=perweight] if kidalive==1, col
```

2. How would the results on young children's co-residence apart from their mothers differ if the data were based on the living arrangements of children under 5 according to the results of a national population census?

    _____

    _____

    _____

3. Does a preference for sons or a low status of females result in a lower proportion of surviving girls than surviving boys in these samples?

    a. India 2005: _____

b. Nigeria 2013: _____

c. Egypt 2008: _____

For survival of male children: _____

```
tab kidalive sample [aw=perweight] if kidsex==1, col
```

For survival of female children: _____

```
tab kidalive sample [aw=perweight] if kidsex==2, col
```

4. What proportion of these young children live/lived in urban areas?

a. India 2005: _____

b. Nigeria 2013: _____

c. Egypt 2008: _____

```
tab urban sample [aw=perweight], col
```

5. Does the term urban mean the same thing in all three countries? Consult the Comparability tab of the URBAN variable description for the answer.

_____

_____

6. What proportion of children experienced diarrheal disease in the past 2 weeks (excluding dead children)?

a. India 2005: _____

b. Nigeria 2013: _____

c. Egypt 2008: _____

```
tab diarrecent sample [aw=perweight] if kidalive==1, col
```

7. For what proportion of children with diarrheal disease in the past two weeks did their mothers seek some treatment or advice about the child's illness? What percent of the

children were treated with oral rehydration solution (a sugar-salt-water mixture, sometimes containing other electrolytes and minerals, which is a cheap and effective treatment against dehydration during diarrheal disease?

a. India 2005: _____

b. Nigeria 2013: _____

c. Egypt 2008: _____

```
tab diatrenone sample [aw=perweight] if diarrecent==23, col
```

```
tab diagivors sample [aw=perweight] if diarrecent==23, col
```

8. For what percent of children did the mother not want any more births when she became pregnant?

a. India 2005: _____

b. Nigeria 2013: _____

c. Egypt 2008: _____

```
tab chdesire sample [aw=perweight], col
```

9. Does a lower proportion of unwanted pregnancies in these samples suggest more effective use of family planning or larger family size desires? Consider the distribution of reported ideal family sizes reported by the children's mothers for the Egypt, India, and Nigeria samples. What is the modal number for ideal family size for each sample? (Note: Total family size is top-coded at 6 or more children).

a. India 2005: _____

b. Nigeria 2013: _____

c. Egypt 2008: _____

```
tab idealkidgrp sample [aw=perweight], co
```

10. How would you expect the numbers for IDEALKIDGRP to be different if you had chosen women of childbearing age, rather than their young children, as the unit of analysis?

Answers

Weighted frequencies

1. Is it rarer for children in these samples to be twins/non-singulate births or for surviving children to live apart from their (living) mothers?

   a. India 2005: Singulate (non-multiple) births: 98.41%; Percent children living with their mothers: 99.22%

   b. Nigeria 2013: Singulate (non-multiple) births: 96.54%; Percent children living with their mothers: 96.89%

   c. Egypt 2008: Singulate (non-multiple) births: 96.09%; Percent children living with their mothers: 99.83%

      For (surviving) children under 5, it is rarer to live apart from their mothers than to be a twin or triplet.

2. How would the results on young children's co-residence apart from their mothers differ if the data were based on the living arrangements of children under 5 according to the results of a national population census?

   If the results were based on data from a national population census, the percentage of children living apart from their mothers would be higher. For a young child to be part of the DHS sample, it had to be reported by their living mother, so none of these children are maternal orphans at the time of the survey. In the total population, some young children enumerated would have outlived their mothers and thus could not reside with their mothers.

3. Does a preference for sons or a low status of females result in a lower proportion of surviving girls than surviving boys in these samples?

   a. India 2005: Share of male children who had died: 6.08%; Share of female children who had died: 6.59%

   b. Nigeria 2013: Share of male children who had died: 9.64%; Share of female children who had died: 8.44%

c. Egypt 2008: <u>Share of male children who had died: 2.81%; Share of female children who had died: 2.14%</u>

<u>Female children appear to be slightly less likely to survive than their male counterparts in India, but to have a slight survival advantage in Nigeria and Egypt. Ideally, a researcher should carry out statistical tests to see whether these differences result from sampling variability only.</u>

4. What proportion of these children live/lived in urban areas?

   a. India 2005: <u>25.34%</u>

   b. Nigeria 2013: <u>34.96%</u>

   c. Egypt 2008: <u>37.05%</u>

5. Does the term urban mean the same thing in all three countries?

   <u>No, each country's definition of what is urban varies, which limits the comparability between countries of the number of children who live/lived in urban areas.</u>

   <u>India defines as urban that are either statutory areas or have at least 5,000 inhabitants, at least 75% of the adult male population engaged in non-agricultural pursuits, and a population density of at least 400 per square kilometer or 1000 per square mile.</u>

   <u>Nigeria defines places as urban that have 20,000 or more inhabitants whose main occupations are primarily non-agricultural.</u>

   <u>Egypt defines places as urban based on capital city or urban governorate status.</u>

6. What proportion of children experienced diarrheal disease in the past 2 weeks (excluding dead children)?

   a. India 2005: <u>8.99%</u>

   b. Nigeria 2013: <u>10.25%</u>

   c. Egypt 2008: <u>8.46%</u>

7. For what proportion of children with diarrheal disease in the past two weeks did their mothers seek some treatment or advice about the child's illness? What percent of the children were treated with oral rehydration solution?

   a. India 2005: 69.08% sought advice or treatment. The percent treated with oral rehydration solution (which could be prepared at home) was much lower, at 26.04%

   b. Nigeria 2013: 70.92% sought advice or treatment. The percent treated with oral rehydration was 33.71%

   c. Egypt 2008: 62.65% sought advice or treatment. The percent treated with oral rehydration was 28.41%

8. For what percent of children did the mother not want any more births when she became pregnant?

   a. India 2005: 10.98%

   b. Nigeria 2013: 1.58%

   c. Egypt 2008: 9.05%

9. Does a lower proportion of unwanted pregnancies in these samples suggest more effective use of family planning or larger family size desires?

   Mothers' reported ideal family size, using children as the unit of analysis, was much higher in Nigeria than in India or Egypt, which could explain the lower proportion of unwanted births for Nigeria.

   What is the modal number for ideal family size for each sample?

   a. India 2005: 78.75% of mothers wanted an ideal family size of two or three children.

   b. Nigeria 2013: 59.91% of mothers wanted an ideal family size of 6 or more children. Only 4.11% wanted an ideal family size of two or three children.

   c. Egypt 2008: 70.21% of mothers wanted an ideal family size of two or three children.

10. How would you expect the numbers for IDEALKIDGRP to be different if you had chosen women of childbearing age, rather than their young children, as the unit of analysis?

Ideal family size norms would be smaller if women rather than children were the unit of analysis, for several reasons. First, women who choose to have zero children are not included in the sample when children are the unit of analysis. Second, mothers who have large families are overrepresented when children are the unit of analysis. For example, every woman gets equal weight in the distribution of family size preferences when women of childbearing age are the unit of analysis; however, women who had two children in the past five years are double counted when children under 5 are the unit of analysis.

This result reflects a broader difference in the measured family size of women versus the family sizes of their children. Mothers who gave birth to one child and mothers who gave birth to eight children have equal weight in the first measure (family sizes of women), but the latter are counted 8 times more for the second measure (family sizes of children).