IPUMS Exercise

IPUMS International Data Extract and Analysis
Data Quality, European Doctoral School of Demography 2024-2025
INED – Paris (France)
Instructor: Mariona Lozano, Centre d'Estudis Demogràfics

Question and Variables

In this exercise you will gain basic familiarity with the IPUMS International data exploration and extract system to answer the following question: *"What are the differences in urbanization, literacy, and occupational participation in Uganda and Mexico?"*
You will create a data extract that includes the following variables: **URBAN, SEX, EMPSTAT, OCCISCO, LIT, AGE.**

Register as an IPUMS User

Instructions sent by email on Wednesday March, 26

Make a data extract

- Navigate to the IPUMS International homepage and click on "Browse Data"

Select sample
- Click on the "Select Samples" button to choose the census samples to include in your extract
- Check the boxes for the 2000 sample for Mexico and 2002 for Uganda
- Submit your sample selections by clicking Submit sample selection box
- Note that by selecting samples first you will now only see variables available for either Mexico or Uganda
    - If you would prefer to see all variables, regardless of their availability in your selected samples, click on "Display Options" from the main variable browsing, page, and choose to display variables that are not available in your selected samples.

Select variables
- The variable drop-down menus allow you to explore variables by topic. For example, you might find variables about occupational participation under the "Work" group.
- The search tool allows you to search for variables. Observe the options for limiting your search results by variable characteristics or variable type.
- 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 reviewing 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 drop down menu or search feature to add these variables to your cart.
    - URBAN: household location
    - SEX
    - EMPSTAT: Employment status
    - OCCISCO: Employment category
    - LIT: Literacy
    - AGE

Review data and request the extract
- Click on the "View Cart" button underneath your data cart.
- Review your variable and sample selection to ensure your extract will be complete.
    - 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.
- Add additional variables or samples if they are missing from your extract or click on the "Create Data Extract" button.
- Review the Extract Request screen that summarizes your extract. Add a description of your extract (e.g., "Data Quality Exercise 2025") and click "Submit Extract".
- You will receive an email when your data is available to download.

Getting the Data into your Statistic Software

The IPUMS International extract builder provides raw ASCII data files and the command files necessary for reading the raw data into a stats package.

Download the data
- Follow the link in the email notifying you that your extract is ready, or by clicking the "Download and Revise Extracts" link on the left-hand side of the IPUMS International homepage.
- Right-click on the data link next to the extract you created.
- Choose "Save Link/Target As"
- Save into your working directory.
- Do the same thing for the DDI link next to the extract.
- You do not need to decompress the data to use it in R.

Install the IPUMSR package
- In R, install the following command: `install.packages ("ipumsr")`

Read in the data

- Set your working directory to where you saved the data: `setwd("~/")  # "~/"`
- Run the following command from the console, adapting it so it refers to the extract you just created:

```
library(ipumsr)
ddi <- read_ipums_ddi("ipums_00001.xml")
data <- read_ipums_micro(ddi)
```

Analyse the Data

Part 1. Variable documentation

For each variable below, search through the tabbed sections of the variable description to answer each question.

1. Under the "Household" dropdown menu, find the "Geography" subcategory and click on the variable URBAN. What constitutes an urban area in each country?
   a. Mexico 2000:
   b. Uganda 2002:
2. What are the codes for URBAN?

3. Find the variable EMPSTAT. Is the reference period of work the same for Mexico and Uganda?

4. What is the universe for EMPSTAT in:
   a. Mexico 2000?
   b. Uganda 2002?

Part 2. Frequencies

5. Find codes page for the SAMPLE variable. What are the codes for:
   a. Mexico 2000?
   b. Uganda 2002?
6. How many individuals are in the Mexico 2000 sample extract?
7. How many individuals are in the Uganda 2002 sample extract?
8. What proportion of individuals in the sample lived in urban areas in each country?
   a. Mexico 2000:
   b. Uganda 2002:

Part 3. Weighted frequencies

To get a more accurate estimate of the actual proportion of individuals living in urban areas, you will have to use the person weight.

9. Using weights, what is the total population of each country?
   a. Mexico 2000:

       b. Uganda 2002:

10. Using weights, how many individuals lived in urban areas in each country?
       a. Mexico 2000:
       b. Uganda 2002:

11. Using weights, what proportion of individuals lived in urban areas in each country?
       a. Mexico: 2000:
       b. Uganda 2002:

Part 4. Trends

12. Using weights, which occupational category has the highest percentage of workers?
       a. In Mexico 2000:
       b. In Uganda 2002:

13. Which occupation category has the highest percentage of female workers?
       a. In Mexico 2000:
       b. In Uganda 2002:

14. What is the labour force participation distribution by gender in each country?
       a. Mexico 2000:
       b. Uganda 2002:

15. What percentage of women within the labour force is working:
       a. In agriculture in Mexico 2000:
       b. In agriculture in Uganda 2002:
       c. In service in Mexico 2000:
       d. In service in Uganda 2002:

Part 5: Graphical Analysis

16. What percentage of the population is literate in each country?
       a. Mexico 2000:
       b. Uganda 2002:

17. (*OPTIONAL*) Create a graph to visualize differences in the percentage of literacy by AGE and SEX in both countries.