IPUMS Data Training Exercise

An Introduction to IPUMS PMA

(Exercise 1 for SPSS)

Learning goals

- Create and download an IPUMS PMA data extract
- Decompress data file and read data in SPSS
- Analyze the data using sample code

Summary

In this exercise, you will gain an understanding of how IPUMS PMA household and female datasets are structured and how it can be leveraged to explore your research interests. This exercise will use the IPUMS PMA dataset to explore the relationship between family planning discontinuation and wealth. You will create data extracts that will include the variables: LINENO, WEALTHQ, FPSTOPWHY, FPAGE1STUSE, FP1STMETHOD, FQWEIGHT, POPWT, HQ WEIGHT, and FPNOWUSPILL; then, you will use sample code to analyze these data.

SPSS Code to Review

Code	Purpose
compute	Creates a new variable
freq	Displays a simple tabulation and frequency of one variable
crosstabs	Displays a cross-tabulation for up to 2 variables and a control
recode	Recodes variables to new specifications
filter	Filters by a variable with values of 0 and 1

Common Mistakes to Avoid

- Excluding cases you don't mean to. Avoid this by turning off weights and select cases after use, otherwise they will apply to all subsequent analyses
- Terminating commands prematurely or forgetting to end commands with a period (.) Avoid this by carefully noting the use of periods in this exercise

Registering with IPUMS

Go to http://pma.ipums.org/, click on Register to Use IPUMS PMA on the left hand side of the screen. Click the Register for IPUMS PMA button and fill out the form to apply for access. You will have to wait for your account to be approved to access the data. Once you receive the approval email, click "Log In" at the top of the page and use your email and password.

Select Samples

- Go to Select Data
- Choose the Person unit of analysis

CHOOSE THE UNIT OF ANALYSIS FOR DATA BROWSING

PERSON	EACH RECORD WILL BE A PERSON
	DESCRIPTION
SERVICE DELIVERY POINT	EACH RECORD WILL BE A SERVICE DELIVERY POINT
	DESCRIPTION

- Click the Select Samples box, check the box for the Kenya 2016 R5

| ☐ Kenya | ☑ 2016 R5 | ☐ 2015b R4 | ☐ 2014b R2 |
| | | ☐ 2015a R3 | ☐ 2014a R1 |

- Scroll to the bottom of the page and click the radio button option for All Cases. The default is Female Respondents
- Click the Submit Sample Selections box

Sample Members

- ○ Female Respondents
- ○ Female Respondents and Household Members
- ○ Female Respondents and Female Non-respondents
- ● All Cases (Respondents and Non-respondents to Household and Female Questionnaires)

SUBMIT SAMPLE SELECTIONS

Select Variables

- The search tool allows you to search for variables. Observe the options for limiting your search results by variable characteristics or variable type.
- You may 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.
- You may view 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.

- Using the drop down menu or search feature, select the following variables:

  LINENO: Person number in sample unit

  WEALTHQ: Wealth quintile

  FPSTOPWHY: Reason for discontinuing family planning method

  FPAGE1STUSE: Age first used family planning

  FP1STMETHOD: First method of family planning used
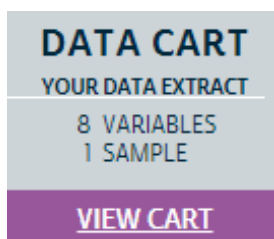
  FQWEIGHT: Female level weighting variable

  POPWT: Population expansion factor

  HQWEIGHT: Household level weighting variable

  FPNOWUSPILL: Currently using the pill

Review and submit your extract

- Click the purple VIEW CART button under your data cart

DATA CART
YOUR DATA EXTRACT
8 VARIABLES
1 SAMPLE

VIEW CART

- Review variable selection. Note that certain variables appear in your data cart even if you did not select them, and they are not included in the constantly updated count of variables in your data cart. The preselected variables are needed for weighting, for variance estimation, or to identify the year, country, and round of a sample.

- Click the Create Data Extract button
- Review the 'Extract Request Summary' screen, describe your extract and click Submit Extract
- You will get an email when the data is available to download.
- To get to the page to download the data, follow the link in the email, or follow the My Data Extracts link on the homepage.

Getting the data into your statistics software

The following instructions are for SPSS. If you would like to use a different stats package, see: https://ipums.org/support/exercises

Download the data

- Go to https://pma.ipums.org/ and click on Download or Revise Extracts.
- Right-click on the Data link next to the extract you created.
- Chpoose "Save Target As..." (or "Save Link As...").
- Save into "Documents" (Documents should pop up as the default location).
- Do the same for the SPSS link next to the Data link.

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.
- Right click on the ".dat" file.
- Use your decompression software to extract the .dat files.
- Double-check that the Documents folder contains three files starting with "pma_000...".

Read the data

- Double click on the ".sps" file, which should automatically have been named "pma_000...".
- The first two lines should read:

```
cd ".".
data list file = "pma_000..."/
```

- Change the first line to read: cd (location where you've been saving your files). For example:

```
cd "C:\Documents".
```

- Change the second line to read:

```
data list file = "C: \Documents\pma_000...dat"/
```

- Under the "Run" menu, select "All"; an output viewer window will open.
- Use the Syntax Editor for the SPSS code below, highlight the code, and choose "Selection" under the Run menu.

Analyze the sample

Part 1: Household Survey frequencies of WEALTHQ

1. According to the Description tab on the website, what does the variable WEALTHQ measure?

   _____

   _____

2. According to the Codes tab, what value labels and numeric codes apply to each wealth quintile? Include labels and codes that explain any "missing" values.

   _____

   _____

3. According to the Universe tab, who is in universe for WEALTHQ in the Kenya 2016 Round 5 sample?

   _____

   _____

4. What percentage of all *household members* are listed as living in a household in the lowest income quintile? _____

```
freq wealthq.
```

5. What percentage of *households* are listed in the lowest income quintile? _____

```
recode lineno (2 thru 27=0)(1=1).
filter by lineno.
freq wealthq.
filter off.
```

EXERCISE 1 FOR SPSS

Part 2: Female Survey frequencies of FPSTOPWHY

6. According to the Description tab on the website, what does the variable FPSTOPWHY measure? It can be found in the Discontinuation of Family Planning variable group. _____

   _____

7. According to the Universe tab, who is in universe for FPSTOPWHY in the Kenya 2016 Round 5 sample? _____

   _____

8. Create a frequency table for the FPSTOPWHY variable for the Kenya 2016 Round 5 sample. What are the top two most common responses? What proportion of surveys provide these responses? _____

   _____

```
freq fpstopwhy.
```

9. Given the universe stated in 7, only a small fraction of women aged 15 to 49 provided answers to FPSTOPWHY. Not in universe cases also include men and women outside of childbearing age. Suppose you wanted to know what proportion of *those who provided an answer* to FPSTOPWHY chose one of the two top answers. How does this answer differ to your response for 8? _____

   _____

   _____

```
recode fpstopwhy (95 thru 99=SYSMIS).
filter by fpstopwhy.
freq fpstopwhy.
```

10. Of the women in the sample who stopped using a family planning method in the last year because of cost, what proportion lived in households in either of the lowest 2 income quintiles? _____

_____

```
crosstabs fpstopwhy by wealthq.

filter off.
```

Part 3: Using household, female, and population weights (HQWEIGHT, FQWEIGHT, and POPWT)

IPUMS PMA data require the use of weights to ensure that each sample is representative of the population from which it was drawn. However, because variables alternately reference either the population of households or the population of women aged 15-49 in each country, users must select weights that correspond to their intended unit of analysis.

For example: the frequency table for FPSTOPWHY from Part 2D indicates that, among the women aged 15-49 in Kenya 2016 Round 5 *who answered this question*, 30.85% stopped using family planning methods in the past year because they wanted to become pregnant. Because the relevant unit of analysis for FPSTOPWHY includes individual women aged 15-49 (rather than households), FQWEIGHT should be used to make inferences about that particular *population*.

11. Use FQWEIGHT to estimate the proportion of all Kenyan women aged 15-49 who stopped using family planning methods in order to become pregnant in 2016. _____

_____

```
weight by fqweight.

freq fpstopwhy.

weight off.
```

While FQWEIGHT may be used to estimate proportions, it should not be used to generate representative population counts. Instead, POPWT combines information from FQWEIGHT with national population estimates from the United Nations Population Division to create an expansion factor appropriate for this purpose (see https://pma.ipums.org/pma/population_weights.shtml for details).

12. Use POPWT to estimate the total number of all Kenyan women aged 15-49 who stopped using family planning methods in order to become pregnant in 2016. _____

```
weight by popwt.

freq fpstopwhy.

weight off.
```

By contrast, variables pertaining to household characteristics should be weighted with HQWEIGHT, which accounts for each household's probability of being sampled and for household non-response. Because each member of the same household shares the same value for HQWEIGHT, users should also ensure that each household is only represented once in their analysis.

For example: Part 1 revealed that 18.85% of households in the Kenya 2016 Round 5 sample fall into the lowest household income quintile. Because the variable WEALTHQ reflects a household characteristic, HQWEIGHT should be used to make inferences about the population of households in Kenya in 2016.

13. Use HQWEIGHT to estimate the proportion of households in the lowest household income quintile in Kenya for 2016. Does the sample over- or under-represent the households in the lowest income quintile compared to the true population of households? _____

```
filter by lineno.

weight by hqweight.

freq wealthq.

filter off.

weight off.
```

Users should note that, much like FQWEIGHT, HQWEIGHT should only be used to generate proportion estimates for the population of households. There is currently no expansion factor for the purpose of estimating population frequency counts.

Part 4: Graphing

14. Examine a frequency table for the variable FPAGE1STUSE, which reflects the age a respondent first began use of any family planning method. Create a histogram for PAGE1STUSE, excluding any missing values.

```
recode fpage1stuse (995 thru 999=SYSMIS).
GGRAPH
/GRAPHDATASET Name="graphdataset" Variables=FPAGE1STUSE
MISSING=LISTWISE REPORTMISSING=NO
/GRAPHSPEC SOURCE=INLINE.
begin gpl
   SOURCE: s=userSource(id("graphdataset"))
   DATA: FPAGE1STUSE=col(source(s), name("FPAGE1STUSE"))
   GUIDE: axis(dim(1), label("Age at first use of FP"))
   GUIDE: axis(dim(2), label("Frequency"))
   ELEMENT:
interval(position(summary.count(bin.rect(FPAGE1STUSE,
binCount(35)))), shape.interior(shape.square))
end gpl.
```

15. Now consider the variable FP1STMETHOD, which classifies the first family planning method ever used by a respondent as either "modern" (values 100-199) or "traditional" (values 200-299). Create a new variable that divides the method type into modern and traditional methods based on the codes and frequencies presented on the PMA website. Recreate the histogram above twice, restricting to only modern or only traditional methods. Specify endpoints between 10 and 50 and bin widths of 1.14.

```
recode FP1STMETHOD (100 thru 199=1) (200 thru 299=0) into
mod_method.

filter by mod_method.

GGRAPH

/GRAPHDATASET Name="graphdataset" Variables=FPAGE1STUSE
MISSING=LISTWISE REPORTMISSING=NO

/GRAPHSPEC SOURCE=INLINE.

begin gpl

    SOURCE: s=userSource(id("graphdataset"))

    DATA: FPAGE1STUSE=col(source(s), name("FPAGE1STUSE"))

    GUIDE: axis(dim(1), label("Age at first use of FP - modern"))

    GUIDE: axis(dim(2), label("Frequency"))

    ELEMENT:
interval(position(summary.count(bin.rect(FPAGE1STUSE,
binCount(35)))), shape.interior(shape.square))

end gpl.

filter off.

recode mod_method (1=0)(0=1).

recode FP1STMETHOD (100 thru 199=1) (200 thru 299=0) into
mod_method.

filter by mod_method.

GGRAPH
```

```
/GRAPHDATASET Name="graphdataset" Variables=FPAGE1STUSE
MISSING=LISTWISE REPORTMISSING=NO

/GRAPHSPEC SOURCE=INLINE.

begin gpl

    SOURCE: s=userSource(id("graphdataset"))

    DATA: FPAGE1STUSE=col(source(s), name("FPAGE1STUSE"))

    GUIDE: axis(dim(1), label("Age at first use of FP -
traditional"))

    GUIDE: axis(dim(2), label("Frequency"))

    ELEMENT:
interval(position(summary.count(bin.rect(FPAGE1STUSE,
binCount(35)))), shape.interior(shape.square))

end gpl.

filter off.
```

ANSWERS

Part 1: Household Survey frequencies of WEALTHQ

1. According to the Description tab on the website, what does the variable WEALTHQ measure? <u>For all households with complete survey results, WEALTHQ refers to the relative wealth of the household where the woman lives, divided into quintiles from the poorest (Lowest quintile) to the richest (Highest quintile).</u>

2. According to the Codes tab, what value labels and numeric codes apply to each wealth quintile? Include labels and codes that explain any "missing" values.

   <u>01 - Lowest quintile</u>   <u>05 - Highest quintile</u>

   <u>02 - Lower quintile</u>   <u>96 - Not interviewed (household questionnaire)</u>

   <u>03 - Middle quintile</u>

   <u>04 - Higher quintile</u>   <u>98 - No response or missing</u>

3. According to the Universe tab, who is in universe for *WEALTHQ* in the Kenya 2016 Round 5 sample? <u>All persons.</u>

4. What percentage of all *household members* are listed as living in a household in the lowest income quintile? <u>22.68%</u>

5. What percentage of *households* are listed in the lowest income quintile? <u>19.35%</u>

Part 2: Female Survey frequencies of FPSTOPWHY

6. According to the Description tab on the website, what does the variable FPSTOPWHY measure? It can be found in the Discontinuation of Family Planning variable group. <u>For women who recently used a method of family planning to delay or avoid pregnancy but are no longer using it, FPSTOPWHY reports the reason why they stopped.</u>

7. According to the Universe tab, who is in universe for FPSTOPWHY in the Kenya 2016 Round 5 sample?   Women age 15-49 who used a family planning method to delay or avoid pregnancy in the last 12 months, but who are not currently using.

8. Create a frequency table for the FPSTOPWHY variable for the Kenya 2016 Round 5 sample. What are the top two most common responses? What proportion of surveys provide these responses? NIU (not in universe) (97.23%); Wanted to become pregnant (0.46%)

9. Given the universe stated in 7, only a small fraction of women aged 15 to 49 provided answers to FPSTOPWHY. Not in universe cases also include men and women outside of childbearing age. Suppose you wanted to know what proportion of *those who provided an answer* to FPSTOPWHY chose one of the two top answers. How does this answer differ to your response for 8? Wanted to become pregnant (31.25 %) Infrequent sex / husband away (22.92%). These proportions now have a more appropriate denominator for analysis._

10. Of the women in the sample who stopped using a family planning method in the last year because of cost, what proportion lived in households in either of the lowest 2 income quintiles? 2 of 3, or 66%

## Part 3: Using household, female, and population weights (HQWEIGHT, FQWEIGHT, and POPWT)

11. Use FQWEIGHT to estimate the proportion of all Kenyan women aged 15-49 who stopped using family planning methods in order to become pregnant in 2016. 2.19%

12. Use POPWT to estimate the total number of all Kenyan women aged 15-49 who stopped using family planning methods in order to become pregnant in 2016. 273,016

13. Use HQWEIGHT to estimate the proportion of households in the lowest household income quintile in Kenya for 2016. Does the sample over- or under-represent the households in the lowest income quintile compared to the true population of households? 18.86%. The lowest income quintile is slightly over-represented in the sample, compared to the true population of households.

Mean = 23
Std. Dev. = 5.754
N = 3,544
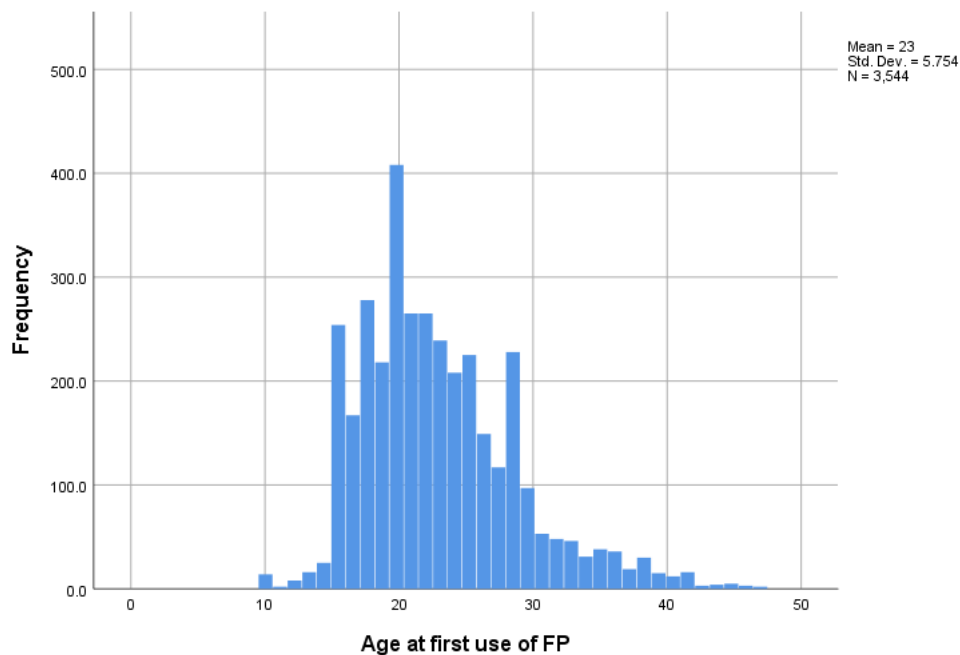
Part 4: Graphing

14. Examine a frequency table for the variable FPAGE1STUSE...Create a histogram for FPAGE1STUSE, excluding any missing values.

15. Now consider the variable FP1STMETHOD... Recreate the histogram above twice, restricting to only modern or only traditional methods. Specify bins endpoints from 10 to 50 and binwidths of 1.14.

Age at first use of FP - modern

Mean = 23.09
Std. Dev. = 5.754
N = 3,392

Age at first use of FP - traditional

Mean = 20.67
Std. Dev. = 5.296
N = 142