

## IPUMS Data Training Exercise: An Introduction to IMPUS PMA (Exercise 1 for SAS)



### Learning Goals

- Create and download an IPUMS PMA data extract
- Decompress data file and read data in SAS
- 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.

## SAS Code to Review

Code	Purpose
PROC FREQ;	Begins a frequency procedure
PROC MEANS;	Begins a means procedure, returns the mean value of a variable
tables;	Required syntax to display frequencies
where;	Selects only specified cases to include in a procedure
weight;	Provides a weight for each observation in the input data set
data newdata; set olddata;	A data step can be used to rename or classify data using logical operators, if/then statements, or renaming.
PROC UNIVARIATE;	Use to create histograms

## Common Mistakes to Avoid

- Not fully decompressing the data
- Giving the wrong filepath to indicate the dataset
- Forget to close a procedure with "run;"
- Forget to terminate a command with a semicolon ";"

## 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

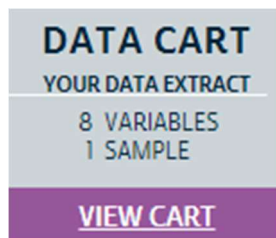
POPWT: Population expansion factor

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



- 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

- Go to <http://pma.ipums.org/> and click on My Data Extracts

Extract Number	Date	Formatted Data	Fixed-width Text Files			
			Data	Command Files 		
51	2018-10-26	--	<a href="#">Download .DAT</a>	<a href="#">SPS</a>	<a href="#">SAS</a>	<a href="#">STATA</a> <a href="#">R</a>

- Right-click on the data link next to extract you created
- Choose "Save Target As..." (or "Save Link As...")
- Save into "Documents" (that should pop up as the default location)
- Do the same thing for the SAS link next to the extract
- Find the "Documents" folder under the Start menu.
- Right click on the ".dat" file
- Use your decompression software to extract here
- Double-check that the Documents folder contains your files starting "pma\_000..."
- Free decompression software is available at <http://www.iris.net/software/wingzip/>
- Open the "pma\_000##.sas" file
- In the "pma\_000##.sas" file window, change the first line from "libname IPUMS '.'" to "libname IPUMS 'C:\Users\%Documents%';" using the file directory where you saved your data files.
- After "filename ASCIIIDAT", enter the full file location, ending with pma\_000##.dat;
- Choose Submit under the Run file 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? \_\_\_\_\_

```
PROC FREQ;  
  
tables wealthq;  
  
run;
```

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

```
PROC FREQ;  
  
tables wealthq;  
  
where lineno in(1);  
  
run;
```



## 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? \_\_\_\_\_  
\_\_\_\_\_

```
PROC FREQ;  
tables fpstopwhy;  
run;
```

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? \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_



```
PROC FREQ;

tables fpstopwhy;

where fpstopwhy <>96 and fpstopwhy<>95 and fpstopwhy<>99;

run;
```

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? \_\_\_\_\_

```
PROC FREQ;

tables fpstopwhy*wealthq / nocol norow nopercnt;

run;
```

### 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. \_\_\_\_\_
- 

```
PROC FREQ;  
  
tables fpstopwhy;  
  
weight fqweight;  
  
run;
```

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](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. \_\_\_\_\_
- 

```
PROC FREQ;  
  
tables fpstopwhy;  
  
weight popwt;  
  
run;
```

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? \_\_\_\_\_

```
PROC FREQ;  
  
tables wealthq;  
  
where lineno in(1);  
  
weight hqweight;  
  
run;
```

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.

```
PROC FREQ;  
  
tables FPAGE1STUSE;  
  
run;
```



```

PROC UNIVARIATE data=IPUMS.pma_00030;
var FPAGE1STUSE;
where FPAGE1STUSE<900;
histogram / endpoints=(10 to 50 by 1.14);
run;

```

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.

```

data IPUMS.pma_00030;
set IPUMS.pma_00030;
    if fp1stmethod <200 then method_mod = 1;
    else if fp1stmethod >200 & fp1stmethod <900 then method_mod
= 0;
    else method_mod = .;
run;

PROC UNIVARIATE data=IPUMS.pma_00030;
var FPAGE1STUSE;
where FPAGE1STUSE <50 & method_mod=1;
histogram / endpoints=(10 to 50 by 1.14);
run;

PROC UNIVARIATE data=IPUMS.pma_00030;
var FPAGE1STUSE;
where FPAGE1STUSE <50 & method_mod=0;
histogram / endpoints=(10 to 50 by 1.14);
run;

```



# ANSWERS

## Part 1: Household Survey frequencies of WEALTHQ

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

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.

01 - Lowest quintile

05 - Highest quintile

02 - Lower quintile

96 - Not interviewed (household questionnaire)

03 - Middle quintile

98 - No response or missing

04 - Higher quintile

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

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

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

## 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. 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.



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%



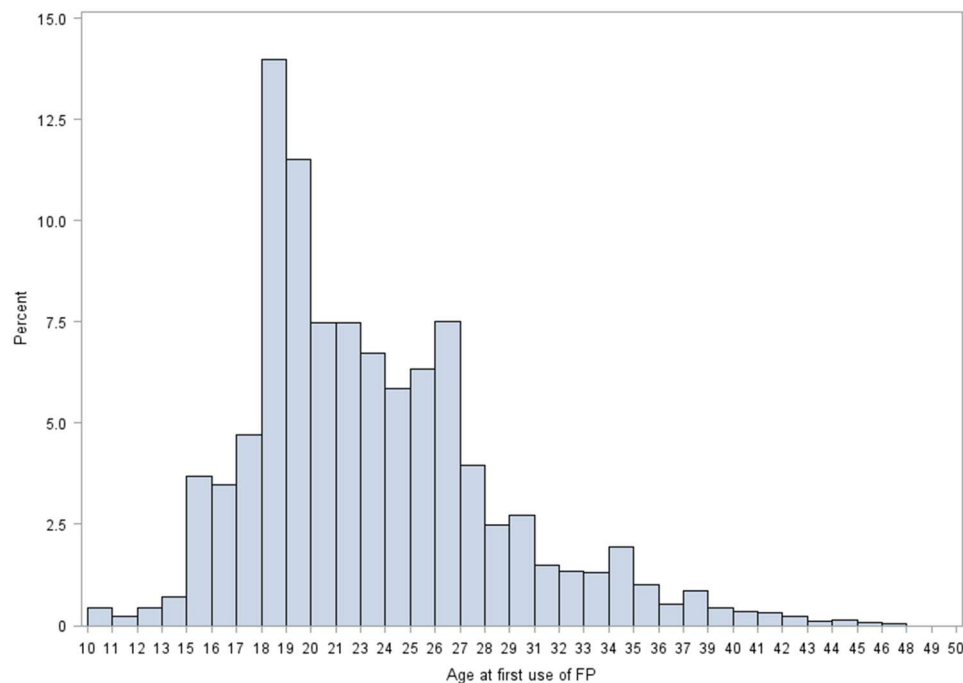
12. Use POPWT to estimate the total number of all Kenyan women aged 15-49 who stopped using family planning methods in order to become pregnant in 2016.

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.

## 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.

