



IPUMS Training and Development: Requesting Data



IPUMS PMA Exercise 1

OBJECTIVE: 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.

IPUMS PMA— Training and Development

Objectives

- Create and download an IPUMS PMA data extract
- Decompress data file and read data in Stata
- Analyze the data using sample code
- Validate data analysis work using answer key

IPUMS 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

Stata Code to Review

Code	Purpose
<code>tabulate</code>	Displays a cross-tabulation for up to 2 variables
<code>generate</code>	Creates a new variable, "replace" specifies a value according to cases
<code>if ()</code>	Conditional operator that can be used in combination with most functions
<code>[fweight=var]</code>	Weights an analysis using a frequency weight

Review Answer Key (page 9)

Common Mistakes to Avoid

- 1 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.
- 2 Forgetting to put [fweight=weightvar] into square brackets

Step 1 Make an Extract

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.

- 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

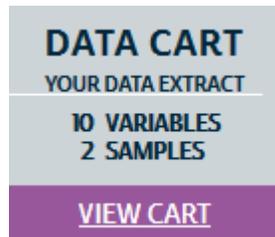
- Using the drop down menu or search feature, select the following variables:

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
POPWT: Population expansion factor
HQWEIGHT: Household level weighting variable

Step 2 Request the Data

FPNOWUSPILL: Currently using the pill

- Click the purple VIEW CART button under your data cart



- Review variable selection. Click the Create Data Extract button
- Across from "Data Format", click Change

DATA FORMAT: .dat (fixed-width text) [Change](#)

- Select Stata (.dta), and then click Submit
- 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.

Step 3 Open the file in Stata

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				Codebook ¹
			Data	Command Files ¹			
36	2018-10-15	Download STATA	--	-	-	-	Basic DDI

- 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)
- Find the "Documents" folder under the Start menu.
- Double-click on the ".dta.gz" file
- In the window that comes up, press the Extract button
- Free decompression software is available at <http://www.irisnet.net/soft/wingzip/>
- Open the "pma_000##.dta" file

Section 1

Analyze the Data

Analyze the sample - Part 1: Household Survey frequencies of WEALTHQ

A) According to the Description tab on the website, what does the variable WEALTHQ measure? _____

B) 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. _____

C) According to the Universe tab, who is in universe for WEALTHQ in the Kenya 2016 Round 5 sample? _____

```
tab wealthq
```

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

```
tab wealthq if lineno == 1
```

E) What percentage of *households* are listed in the lowest income quintile?

Analyze the sample - Part 2: Female Survey frequencies of FPSTOPWHY

A) 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. _____

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

C) Create a frequency table for the FPSTOPWHY variable for the Kenya 2016 Round 5 sample. What are the top two most common responses?

```
tab fpstopwhy
```

What proportion of surveys provide these responses? _____

Section 1 Analyze the Data

D) Given the universe stated in B), 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

```
tab fpstopwhy if (fpstopwhy != 96 & fpstopwhy != 95 &
fpstopwhy != 99)
```

FPSTOPWHY chose one of the two top answers. How does this answer differ to your response for C)?

```
tab fpstopwhy wealthq
```

E) 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? _____

• • •

Section 2 Weight the Data

Analyze the sample - 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*.

Section 2

Weight the Data

```
replace fqweight= round(fqweight)
tab fpstopwhy [fweight=fqweight]
```

A) 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. _____

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

```
tab fpstopwhy [fweight=popwt]
```

https://pma.ipums.org/pma/population_weights.shtml for details).

B) 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. _____

By contrast, variables pertaining to household characteristics should be weighted with HQWEIGHT, which accounts for each household's the 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,

```
replace hqweight= round(hqweight)
tab wealthq if (lineno ==1) [fweight=hqweight]
```

HQWEIGHT should be used to make inferences about the population of households in Kenya in 2016.

C) Use HQWEIGHT to estimate the proportion of households in the lowest household income quintile in Kenya for 2016. Does the sample

Section 3

Visualize the data

over- or under-represent the households in the lowest income quintile compared to the true population of households?

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

```
tabulate fpage1stuse  
hist(fpage1stuse) if fpage1stuse <50, bin(50)
```

estimating population frequency counts.

Analyze the Sample – Part 4: Graphing

A) 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 FPAGE1STUSE, excluding any missing values.

B) Now consider the variable FP1STMETHOD, which classifies the first family planning method ever used by a respondent as either “modern”

```
gen method_mod =1 if fp1stmethod <200  
replace method_mod=0 if fp1stmethod>200 & fp1stmethod<900  
hist (fpage1stuse) if method_mod==1 & fpage1stuse <50, bin(35)  
hist (fpage1stuse) if method_mod==0 & fpage1stuse <50, bin(35)
```

(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 35 bins for each using the bin(35) option.

• • •

Complete!
Validate
Your
Answers
Below

Answers: Section 1 Analyze the Data

ANSWERS -- Analyze the sample - Part 1: Household Survey frequencies of WEALTHQ

A) 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).

B) 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
02 - Lower quintile
03 - Middle quintile
04 - Higher quintile
05 - Highest quintile
96 - Not interviewed (household questionnaire)
98 - No response or missing

```
tab wealthq
```

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

```
tab wealthq if lineno == 1
```

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

E) What percentage of *households* are listed in the lowest income quintile? 19.35%

Analyze the sample - Part 2: Female Survey frequencies of FPSTOPWHY

A) According to the Description tab on the website, what does the variable FPSTOPWHY measure? It can be found in the Discontinuation

Answers: Section 1 Analyze the Data

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.

B) 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.

```
tab fpstopwhy
```

C) 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%)

D) Given the universe stated in B), 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

```
tab fpstopwhy if (fpstopwhy != 95 & fpstopwhy != 96 & fpstopwhy != 97 & fpstopwhy != 98 & fpstopwhy != 99)
```

FPSTOPWHY chose one of the two top answers. How does this answer differ to your response for C)? Wanted to become pregnant (31.25%) Infrequent sex / husband away (22.92%). These proportions now have

```
tab fpstopwhy wealthq
```

a more appropriate denominator for analysis.

E) 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%

Analyze the sample - Part 3: Using household, female, and population weights (HWEIGHT, FWEIGHT, 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,

• • •

Answers: Section 2 Weight the Data

Answers: Section 2 Weight the Data

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*.

```
replace fqweight= round(fqweight)
tab fpstopwhy [fweight=fqweight]
```

A) 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%

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).

```
tab fpstopwhy [fweight=popwt]
```

B) 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

By contrast, variables pertaining to household characteristics should be weighted with HQWEIGHT, which accounts for each household's the 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 19.35% 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.

C) Use HQWEIGHT to estimate the proportion of households in the lowest household

```
replace hqweight= round(hqweight)
tab wealthq if (lino ==1) [fweight=hqweight]
```

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.

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.

ANSWERS -- Analyze the Sample – Part 4: Graphing

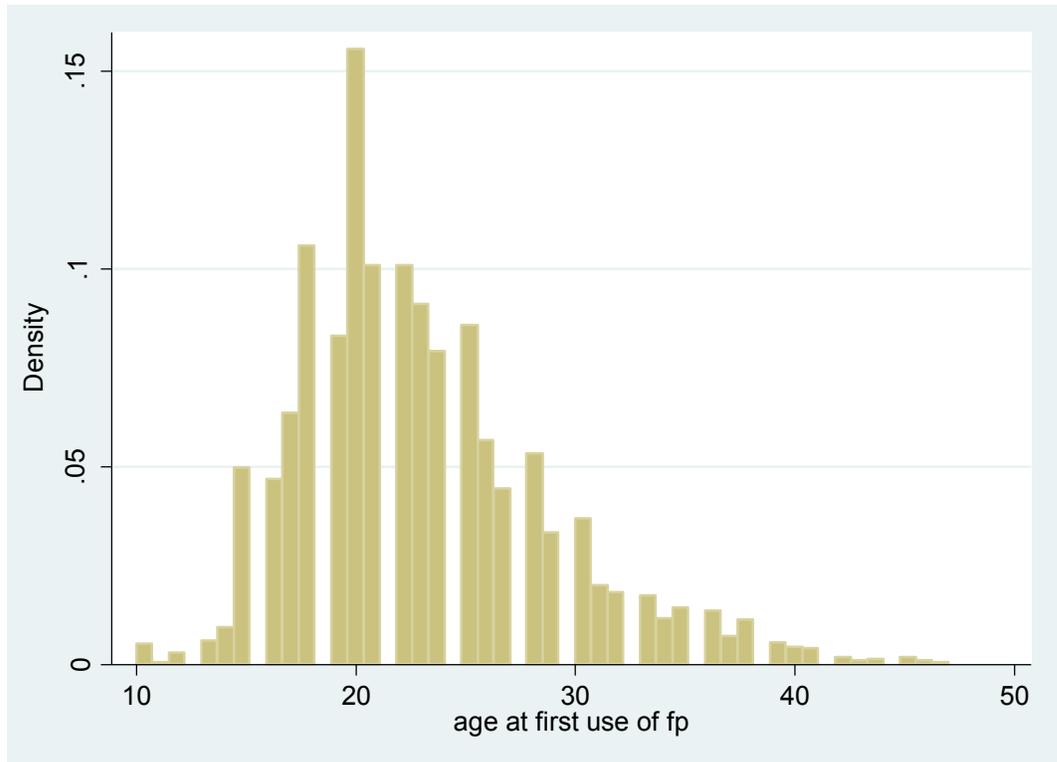
A) Examine a frequency table for the variable FPAGE1STUSE, which reflects the age a

```
tabulate fpage1stuse
hist(fpage1stuse) if fpage1stuse <50, bin(50)
```

respondent first began use of any family planning method. Create a histogram for FPAGE1STUSE, excluding any missing values.

**Answers:
Section 3
Visualize
the data**

```
gen method_mod =1 if fp1stmethod <200  
replace method_mod=0 if fp1stmethod>200 & fp1stmethod<900  
hist (fpage1stuse) if method_mod==1 & fpage1stuse <50, bin(35)  
hist (fpage1stuse) if method_mod==0 & fpage1stuse <50, bin(35)
```



B) 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 35 bins for each using the bin(35) option.

