IPUMS Data Training Exercise

An Introduction to IPUMS PMA

(Exercise 2 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 service delivery point datasets are structured and how it can be leveraged to explore your research interests. This exercise will use the PMA dataset to explore basic frequencies of births, facilities and rural or urban locations. You will create data extracts that include the variables: EAID, FACILITYTYPEGEN, FACILITYADV, PILLOSBS, PILLOUTDAY, and URBAN; then, you will use a 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, can be used with output out to collapse data
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
PROC GCHART	Can be used with hbar to generate a horizontal bar chart
PROC SQL	Can be used to output a value of a variable of interest
merge	Can be used in a data step to merge datasets

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

EXERCISE 2 FOR SAS

you receive the approval email, click "Log In" at the top of the page and use your email and password.

Select Samples

- Choose the Service Delivery Point unit of analysis

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

- Scroll to the bottom of the page and click the radio button option for All Cases. The default is Facility Respondents

Click the Submit Sample Selections box

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:

  EAID: Enumeration area (primary sampling unit)

  FACILITYTYPEGEN: Type of facility

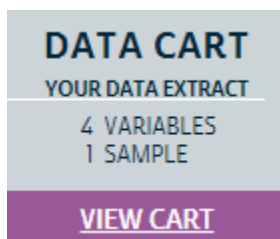 FACILITYADV: Advanced facility

  PILLOBS: Observed and in or out of stock of birth control pills

  PILLOUTDAY: Number of days that birth control pills have been out of stock

  URBAN: Urban or rural status

Review and submit your extract

Click the purple VIEW CART button under your data cart

DATA CART
YOUR DATA EXTRACT
4 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

Download the data

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

Extract Number	Date	Formatted Data	Fixed-width Text Files			
			Data	Command Files ⓘ		Codebook ⓘ
51	2018-10-26	--	Download .DAT	SPS SAS STATA R		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)

- 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 www.7-zip.org/download.html

- 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 ASCIIDAT", enter the full file location, ending with pma_000##.dat;

- Choose Submit under the Run file menu.

Analyze the Sample

Part 1: Exploring Facility Types

1. Create a frequency table for FACILITYTYPEGEN showing the proportion of each type of facility surveyed in Kenya 2016 Round 5. _____

```
PROC FREQ;

tables facilitytypegen;

run;
```

2. According to the Universe tab, what facilities are included in the surveyed universe for FACILITYTYPEGEN? _____

_____

3. Users should note that many variables in the service delivery point (SDP) survey have a universe defined by FACILITYADV, a country-specific designation of "advanced facility" types. Create a crosstab to see which types of facilities from the previous question were designated as "advanced facilities" in Kenya for 2016.

```
PROC FREQ;

tables facilitytypegen*facilityadv / norow nocol nopercent;
```

4. Consult the Comparability tab for FACILITYADV, taking care to note that advanced facility designations vary by country, and sometimes vary by survey round within a country. Locate the entry for Kenya, and determine whether its advanced facility designation matches what you found in Question 3. Is the designation consistent for all Kenyan survey rounds that included this variable? _____

_____

Part 2: Descriptive Statistics

5. Consider the variable PILLOBS, which describes whether the SDP had an observable stock of birth control pills on the day of the interview. According to the Codes tab, what are the possible responses for SDPs surveyed in Kenya 2016?

   _____

   _____

6. According to the Comparability tab, possible responses to PILLOBS may vary from sample to sample. How so? _____

   _____

7. According to the Universe tab, what facilities are included in the surveyed universe for PILLOBS? _____

   _____

8. Among facilities that usually provide birth control pills shown in PILLOBS, what type of facility was least likely to have supplies of birth control pills in-stock on the day of the interview? What proportion of facilities of this type were out of stock? (Restrict analysis only to completed interviews and in-universe cases). _____

   _____

```
PROC FREQ;

tables facilitytypegen*pillobs / nocol nopercent;

where pillobs <90;

run;
```

Part 3: Data Visualization

For facilities that were out of birth control pills, PILLOUTDAY shows the number of days that supplies had been unavailable. Because some SDPs had been out of stock for more than 90 days, NIU and missing value codes for PILLOUTDAY are coded as values 9994, 9997, and 9999 in order to exceed the range of valid responses.

9. Calculate the mean shortage of pills for *all* in-universe facilities in PILLOUTDAY, taking care to exclude any value above 9000. Then find the mean for *each facility type* in FACILITYTYPEGEN, and display the result as a bar chart. (Restrict analysis only to valid responses from SDPs in universe for PILLOUTDAY).

```
PROC MEANS data=IPUMS.pma_000##;

var pilloutday;

where pilloutday <9000;

run;

PROC MEANS data=IPUMS.pma_000##;

var pilloutday;

class facilitytypegen;

where pilloutday <9000;

run;

PROC GCHART data=IPUMS.pma_000##;

title1 "Mean days out of stock by facility type";

hbar facilitytypegen / discrete type=mean sumvar=pilloutday
nostats;

where pilloutday <9000;

run;
```

10. Suppose you suspect that the apparent difference between the facilities in 9 is really a disparity between types of facilities that are most likely to be found in urban vs. rural areas. Create a pair of bar charts groups by URBAN to test if this is true. Are there differences between urban and rural facilities of each type?

```
PROC GCHART data=IPUMS.pma_000##;

title1 "Mean days out of stock by facility type and urban/rural status";

hbar facilitytypegen / discrete type=mean sumvar=pilloutday nostats group=urban;

where pilloutday <9000;

run;
```

Part 4: Combining SDP and HHF Data

Users should note that PMA2020 surveyed facilities in the same sampling areas as households and females in the same survey round. These SDP data are *not meant to be nationally representative.* Instead, they are meant to portray the health provision environment of the surveyed households and women. Thus, there are no sampling weights for SDP variables.

The files do contain a weight for the sampling units EAWEIGHT, which is a probability weight representing the likelihood of the enumeration area (EA) being selected for sampling. The collectors of the original data do not recommend using EAWEIGHT to weight SDP variables. Rather, the best use of SDP variables is to calculate summary statistics at the EA level and attach them to the Household and Female (HHF) dataset using the EAID

variable as a source of contextual information for each woman's service delivery environment.

For example, one could use the variables PILLOBS and PILLOUTDAY to calculate whether any facility in each EAID was out of stock of birth control pills and the mean number of days the facility or facilities in each EAID were out of stock of pills, respectively. These summary statistics may be merged with the HHF dataset in order to show whether each female respondent had reliable local access to birth control pills.

11. Calculate statistics described above in the SDP data by EAID. With this transformation, PILLOUTDAY should reflect the mean of PILLOUTDAY for each EAID, and PILLOBS should indicate whether any facility in the EAID was out of pills. How many enumeration areas contain a facility that was out of birth control pills on the day of the interview?

```
data IPUMS.pma_000##;
    set IPUMS.pma_000##;
        if pilloutday >9000 then pilloutday= .;
        if pillobs =3 then pillobs = 1;
        else pillobs =0;
run;
PROC MEANS data=IPUMS.pma_000## nway;
class eaid;
var pillobs pilloutday;
```

EXERCISE 2 FOR SAS

```
output out=IPUMS.pma_000##_1 max(pillobs) = pillobs
mean(pilloutday) = pilloutday;

run;

PROC SQL;

select count(*) as N_obs

from IPUMS.pma_000##_1

where pillobs=1;

quit;
```

12. Merge PILLOBS and PILLOUTDAY to the Household and Female (HHF) dataset by matching EAIDs for all respondents. How many of the women aged 15-49 (ELIGIBLE == 1) in the Kenya 2016 Round 5 HHF sample resided in each of the enumeration areas where birth control pills were not available at all local facilities in the SDP survey? *(Note that you will need to replace ## with the number of your IPUMS PMA extract used in Exercise 1)*

```
PROC SORT data=IPUMS.pma_000##_1;
by eaid;
run;
PROC SORT data=IPUMS.pma_000##;
by eaid;
run;
data IPUMS.HHF_SDP;
 merge  IPUMS.pma_000##_1 IPUMS.pma_000##;
 by eaid;
 run;
PROC FREQ;
tables eaid;
where pillobs=1 & eligible=1;
run;
```

13. Run a logistic regression to predict the association between women currently using the pill (FPNOWUSPILL) and the mean shortage duration (PILLOUTDAY) for each enumeration area that was out of pills on the day of the SDP interview. Adjust your model to be representative of Kenyan women using FQWEIGHT. Consider the effect of the mean number of days out of stock as a continuous variable and the predictor of interest. Recode values for FPNOWUSPILL that are not in universe or missing to zero.

    Is there an association between the number of days that the facilities in the woman's enumeration area are out of stock of pills and the woman's current use of the pill for family planning?

```
data IPUMS.HHF_SDP;

set IPUMS.HHF_SDP;

if fpnowuspill >90 then fpnowuspill= 0;

if pilloutday =. then pilloutday = 0;

run;

PROC LOGISTIC data=IPUMS.HHF_SDP;

class fpnowuspill / param=ref;

model fpnowuspill = pilloutday;

weight fqweight;

run;
```

ANSWERS

Part 1: Exploring Facility Types

1. Create a frequency table for FACILITYTYPEGEN showing the proportion of each type of facility surveyed in Kenya 2016 Round 5.

Hospital: 79/428 = 18.46%

Health Center: 90/428 = 21.03%

Health Clinic: 16/428 = 3.74%

Other Health Facility: 1/428 = 0.23%

Dispensary: 190/428 = 44.39%

Pharmacy / Chemist / Drug Shop: 48/428 = 11.21%

Other: 4/428 = 0.93%

2. According to the Universe tab, what facilities are included in the surveyed universe for FACILITYTYPEGEN? All service delivery points

3. Users should note that many variables in the service delivery point (SDP) survey have a universe defined by FACILITYADV, a country-specific designation of "advanced facility" types. Create a crosstab to see which types of facilities from the previous question were designated as "advanced facilities" in Kenya for 2016. All are advanced, except for Pharmacy / Chemist / Drug Shop

4. Consult the Comparability tab for FACILITYADV, taking care to note that advanced facility designations vary by country, and sometimes vary by survey round within a country. Locate the entry for Kenya, and determine whether its advanced facility designation matches what you found in Question C. Is the designation consistent for all Kenyan survey rounds that included this variable? It does match, and all Kenyan rounds interviewed have the same designation.

Part 2: Descriptive Statistics

5. Consider the variable PILLOBS, which describes whether the SDP had an observable stock of birth control pills on the day of the interview. According to the Codes tab, what are the possible responses for SDPs surveyed in Kenya 2016?

    1 - In-stock and observed   94 - Not interviewed (SDP questionnaire)

    2 - In-stock but not observed   98 - No response or missing

    3 - Out of stock   99 - NIU (not in universe)

6. According to the Comparability tab, possible responses to PILLOBS may vary from sample to sample. How so? Some early samples include less detail, providing dichotomous responses based on whether the interviewer observed contraceptive pills in-stock. In these early samples, if contraceptive pills were not observed, they were assumed to be "out of stock". In later surveys, interviewers had the option of reporting that contraceptive pills were "in-stock but not observed".

7. According to the Universe tab, what facilities are included in the surveyed universe for PILLOBS? Service delivery points that provide contraceptive pills.

8. Among facilities that usually provide birth control pills shown in PILLOBS, what type of facility was least likely to have supplies of birth control pills in-stock on the day of the interview? What proportion of facilities of this type were out of stock? (Restrict analysis only to completed interviews and in-universe cases). Health clinics were most likely to be out of pills with 25% out of stock.
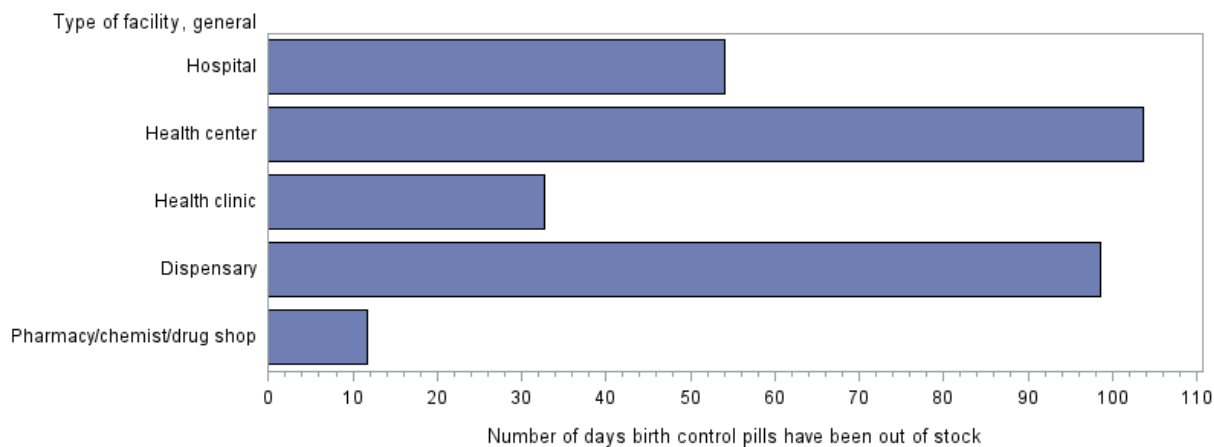
Part 3: Data Visualization

9. For facilities that were out of birth control pills, PILLOUTDAY shows the number of days that supplies had been unavailable...
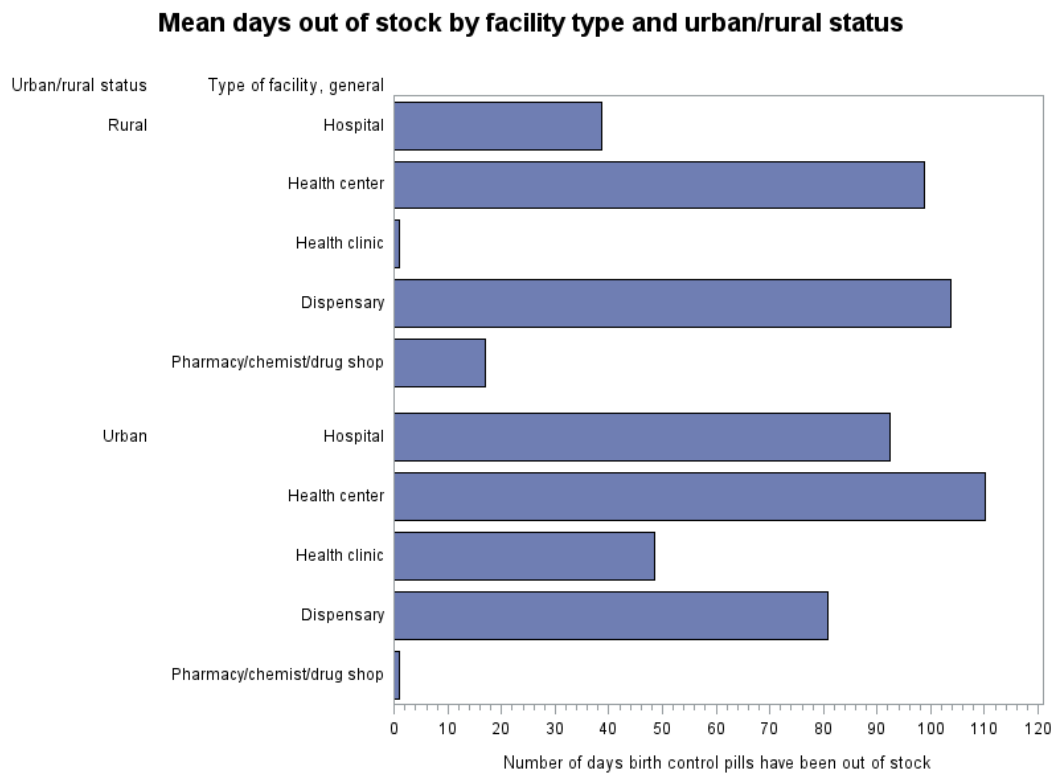
    Calculate the mean shortage of pills for *all* in-universe facilities in PILLOUTDAY, taking care to exclude any value above 9000. Then find the mean for *each facility type* in FACILITYTYPEGEN, and display the result as a bar chart. (Restrict analysis only to valid responses from SDPs in universe for PILLOUTDAY).

All facilities: 87	Health Clinic: 33
Hospital: 54	Dispensary: 99
Health Center: 104	Pharmacy/chemist/drug shop: 12

Mean days out of stock by facility type

10. Suppose you suspect that the apparent difference between the facilities in A) is really a disparity between types of facilities that are most likely to be found in urban vs. rural areas. Create a pair of bar charts groups by URBAN to test if this is true. Are there differences between urban and rural facilities of each type?

Mean days out of stock by facility type and urban/rural status

Part 4: Combining SDP and HHF Data

11. ...How many enumeration areas contain a facility that was out of birth control pills on the day of the interview? 43 enumeration areas

12. Merge PILLOBS and PILLOUTDAY to the Household and Female (HHF) dataset by matching EAIDs for all respondents. How many of the women aged 15-49 (ELIGIBLE == 1) in the Kenya 2016 Round 5 HHF sample resided in each of the enumeration areas where birth control pills were not available at all local facilities in

EXERCISE 2 FOR SAS

the SDP survey? *(Note that you will need to replace ## with the number of your IPUMS PMA extract used in Exercise 1).*

These are the first three EAIDs in the list:

- EAID 4013 = 33 women

- EAID 4047 = 44 women

- EAID 4163 = 47 women

13. Run a logistic regression to predict… Is there an association between the number of days that the facilities in the woman's enumeration area are out of stock of pills and the woman's current use of the pill for family planning?

The likelihood that a sampled woman uses birth control pills remains the same regardless of the average number of days that her local SDP had none available (odds ratio = 1.000). However, this finding is not statistically significant (p = 0.6164).

FPNOWUSPILL	Odds Ratio	P value	95% CI
PILLOUTDAY	1.001	0.6164	(0.998, 1.003)