



---

## IPUMS Training and Development: Requesting Data

---



### IPUMS PMA Exercise 2

**OBJECTIVE:** Gain an understanding of how IPUMS PMA service delivery point datasets are structured and how it can be leveraged to explore your research interests.

# 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

- 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 birth control pills have been out of stock

## Stata Code to Review

Code	Purpose
<code>PROC FREQ;</code>	Begins a frequency procedure
<code>PROC MEANS;</code>	Begins a means procedure, returns the mean value of a variable, can be used with <code>output out</code> to collapse data
<code>tables</code>	Required syntax to display frequencies
<code>where</code>	Selects only specified cases to include in a procedure
<code>weight;</code>	Provides a weight for each observation in the input data set
<code>data newdata;</code> <code>set olddata;</code>	A data step can be used to rename or classify data using logical operators, if/then statements, or renaming.
<code>PROC UNIVARIATE</code>	Use to create histograms
<code>PROC GCHART</code>	Can be used with <code>hbar</code> to generate a horizontal bar chart
<code>PROC SQL</code>	Can be used to output a value of a variable of interest
<code>merge</code>	Can be used in a data step to merge datasets

## Review Answer Key (page 9)

## Step 1 Make an Extract

### Common Mistakes to Avoid

- 1 Not fully decompressing the data
- 2 Giving the wrong filepath to indicate the dataset
- 3 Forget to close a procedure with "run;"
- 4 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.

- Go to Select Data
- Choose the Service Delivery Point 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 Facility Respondents
- Click the Submit Sample Selections box

### Sample Members

- Facility Respondents
- All Cases (Respondents and Non-respondents to Service Delivery Point Questionnaires)

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

FACILITYTYPEGEN: Type of facility

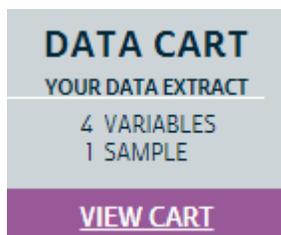
FACILITYADV: Advanced facility

PILLOBS: In or out of stock of birth control pills

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

## Step 2 Request the Data

- Click the purple VIEW CART button under your data cart



- Review variable selection. 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					Codebook
			Data	Command Files				
36	2018-10-15	<a href="#">Download STATA</a>	--	-	-	-	-	<a href="#">Basic</a> <a href="#">DDI</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\_ooo..."
- Free decompression software is available at <http://www.iris.net/soft/wingzip/>
  
- Open the "pma\_ooo##.sas" file
- In the "pma\_ooo##.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\_ooo##.dat;
- Choose Submit under the Run file menu.

## Step 3 Open the file in SAS

## Section 1 Explore the Data

### Analyze the sample - Part 1: Exploring Facility Types

A) 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;
```

---

B) According to the Universe tab, what facilities are included in the surveyed universe for FACILITYTYPEGEN? \_\_\_\_\_

C) 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

```
PROC FREQ;  
tables facilitytypegen*facilityadv / norow nocol nopercnt;  
run;
```

as “advanced facilities” in Kenya for 2016. \_\_\_\_\_

D) 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? \_\_\_\_\_

• • •

## Section 2 Analyze the Data

### Analyze the sample - Part 2: Descriptive Statistics

A) 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? \_\_\_\_\_

B) According to the Comparability tab, possible responses to PILLOBS may vary from sample to sample. How so? \_\_\_\_\_

## Section 2

### Analyze the Data

C) According to the Universe tab, what facilities are included in the surveyed universe for PILLOBS?

---

D) 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*pillob / nocol nopercnt;
where pillob <90;
run;
```

• • •

## Section 3

### Visualize the Data

---

#### Analyze the sample - Part 3: Data Visualization

A) 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.

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). Be sure to replace the #s with the

```
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;
```

## Section 3

### Visualize the Data

```
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;
```

name of your title file.

**B)** 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

```
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;
```

between urban and rural facilities of each type?

## Section 4

### Use the SDP and HHF together

#### Analyze the Sample – 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.

**A)** 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?

## Section 4

### Use the SDP and HHF together

```
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;
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;
```

**B)** 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*

```
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;
```

## Section 4

### Use the SDP and HHF together

*extract used in Exercise 1)*

C) 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

```
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;
```

in the woman's enumeration area are out of stock of pills and the woman's current use of the pill for family planning?

• • •

**Complete!**  
**Validate**  
**Your**  
**Answers**  
**Below**

## Answers: Section 1 Explore the Data

### ANSWERS - Analyze the sample - Part 1: Exploring Facility Types

A) Create a frequency table for FACILITYTYPEGEN showing the

```
PROC FREQ;  
tables facilitytypegen;  
run;
```

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:  $4/428 = 0.93\%$

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

C) 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

```
PROC FREQ;  
tables facilitytypegen*facilityadv / norow nocol nopercnt;  
run;
```

advanced, except for Pharmacy / Chemist / Drug Shop

D) 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.

## Answers: Section 2 Analyze the Data

### ANSWERS - Analyze the sample - Part 2: Descriptive Statistics

A) 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
- 2 - In-stock but not observed
- 3 - Out of stock
- 94 - Not interviewed (SDP questionnaire)
- 98 - No response or missing
- 99 - NIU (not in universe)

B) 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".

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

D) 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

```
PROC FREQ;  
tables facilitytypegen*pillobs / nocol nopercnt;  
where pillobs <90;  
run;
```

likely to be out of pills with 25% out of stock. \_

## Answers: Section 3 Visualize the Data

### ANSWERS - Analyze the sample - Part 3: Data Visualization

A) 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.

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

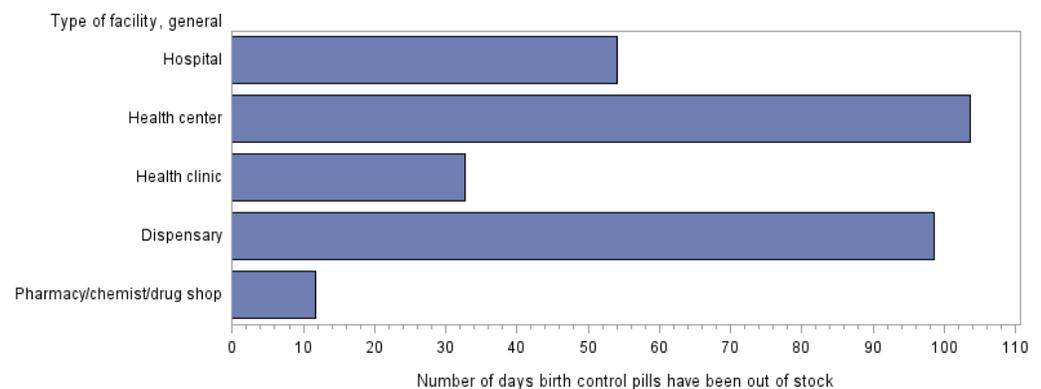
```
PROC MEANS data=IPUMS.pma_000##;  
var pilloutday;  
where pilloutday <9000;  
run;  
PROC MEANS data=IPUMS.pma_000##;  
var pilloutday;  
class facilitytypegen;  
where pilloutday <9000;  
run;  
PROC MEANS data=IPUMS.pma_000##;  
var pilloutday;  
where pilloutday <9000;  
run;  
PROC MEANS data=IPUMS.pma_000##;  
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;
```

## Answers: Section 3 Visualize the Data

SDPs in universe for PILLOUTDAY).

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

**Mean days out of stock by facility type**



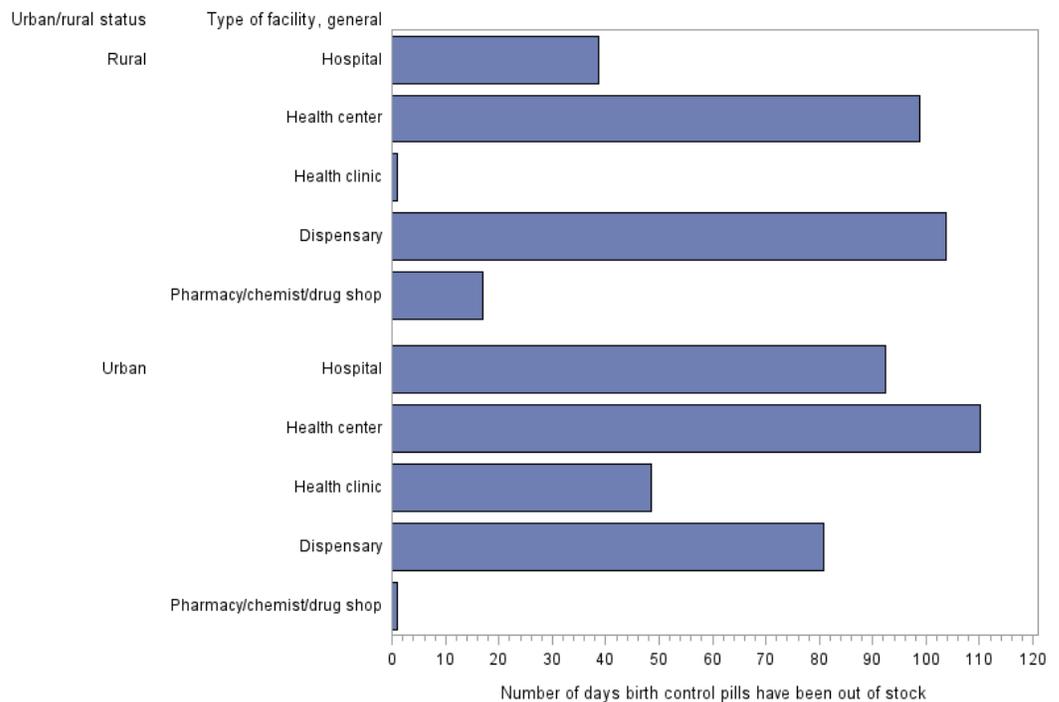
**B)** Suppose you suspect that the apparent difference between the facilities in A) is really a disparity between types of facilities that are most likely to be found in urban vs. rural areas. Create a pair of bar charts groups by URBAN to test if this is true. Are there differences

```
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;
```

between urban and rural facilities of each type?

## Answers: Section 3 Visualize the Data

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



• • •

## Answers: Section 4 Use the SDP and HHF together

### ANSWERS - Analyze the Sample – 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.

## Answers: Section 4 Use the SDP and HHF together

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.

A) 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

```
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;
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;
```

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? 43 enumeration areas

B) 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)*

**Answers:  
Section 4  
Use the  
SDP and  
HHF  
together**

```
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;
```

These are the first three EAIDs in the list:

EAID 4013 = 33 women  
EAID 4047 = 44 women  
EAID 4163 = 47 women

C) 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.

**Answers:**  
**Section 4**  
**Use the**  
**SDP and**  
**HHF**  
**together**

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

```
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;
```

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)

```
replace fpnowuspill = 0 if fpnowuspill > 90
replace pilloutday = 0 if pilloutday == .
replace fqweight = round(fqweight)
logistic fpnowuspill pilloutday [fweight = fqweight]
```