IPUMS Training and Development: Online Extraction

IPUMS CPS Exercise 1

OBJECTIVE: Gain an understanding of how the IPUMS CPS dataset is structured and how it can be leveraged to explore your research interests. This exercise will use the IPUMS CPS dataset to create basic frequencies of food stamp usage and to explore the relationship between health and work status.

IPUMS CPS — Training and Development

Research Question
What is the frequency of food stamp recipiency in the US? Are health and work statuses related?

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

IPUMS Variables
- PERNUM: Person number in sample unit
- FOODSTMP: Food stamp receipt
- AGE: Age
- EMPSTAT: Employment status
- AHRSWORKT: Hours worked last week
- HEALTH: Health status

SAS Code to Review

CODE	PURPOSE
PROC FREQ;	Returns frequencies for a given variable
PROC MEANS;	Returns average values for variable observations
TABLES;	
WEIGHT;	

Review Answer Key (page 8)

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 ";"

Section 1

Step 1

Register with IPUMS

• • •

Step 2

Make an Extract

• • •

Step 3

Request the Data

Creating a Data Extract

Go to http://cps.ipums.org, click on Register with IPUMS and apply for access. On login screen, enter email address and password and submit it!

▪ Go back to homepage and go to Select Data

▪ Click the Select Samples box, deselect All Default samples box, check the box for the 2011 ASEC sample, Click the Submit Sample Selections box

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

PERNUM:	Person number in sample unit
FOODSTMP:	Food stamp receipt
AGE:	Age
EMPSTAT:	Employment status
AHRSWORKT:	Hours worked last week
HEALTH:	Health status

▪ Click the 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 Download and Revise Extracts link on the homepage

Section 2

Step 1

Download the Data

Getting the Data into Your Statistics Software

The following instructions are for SAS. If you would like to use a different stats package, see:
http://cps.ipums.org/cps/extract_instructions.shtml

- Go to http://cps.ipums.org and click on Download or Revise Extracts

- Right-click on the data link next to the 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

• • •

Step 2

Decompress the Data

- Find the "Documents" folder under the Start menu

- Double-click on the ".dat" file

- In the window that comes up, press the Extract button

- Double-check that the Documents folder contains three files starting "cps_ooo..."

- Free decompression software is available at https://www.7-zip.org

• • •

Step 3

Read in the Data

- Open the "cps_ooo##.sas" file

- In the sas file window, change the first line from "libname IPUMS '.'" to "libname IPUMS \\Documents...;" using the file directory where you saved your data files

- After "filename ASCIIDAT", enter the full file location, ending with cps_ooo##.dat";

- Choose Submit under the Run file menu

Section 3

Step 1

Analyze the Data

Analyze the Data – Part I Frequencies of FOODSTMP

A) On the website, find the codes page for the FOODSTMP variable and write down the code value, and what category each code represents. _____

B) What is the universe for FOODSTMP in 2011 (under the Universe tab on the website)? _____

C) How many people received food stamps in 2011? _____

```
proc freq;
      tables foodstmp;
      weight asecwt;
run;
```

D) What proportion of the population received food stamps in 2011? _____

• • •

Step 2

Weighting the Data

Using household weights (ASECWTH)

Suppose you were interested not in the number of people living in homes that received food stamps, but in the number of households that had food stamp participants. To get this statistic you would need to use the household weight.

To use household weight, you should be careful to select only one person from each household to represent that household's characteristics. You will need to apply the household weight (ASECWTH).

A) How many households received food stamps in 2011? _____

B) What proportion of households received food stamps in 2011? _____

Analyze the Data – Part II Relationships in the Data

Section 4
Analyze the Data

A) What is the universe for EMPSTAT in 2011? _____

B) What are the possible responses and codes for the self-reported HEALTH variable? _____

C) What percent of people with 'poor' self-reported health are at work? _____

D) What percent of people with 'very good' self-reported health are at work? _____

E) In the EMPSTAT universe, what percent of people:

```
proc freq;
      tables empstat;
      where health in (5);
      weight asecwt;
run;
```

i. self-report 'poor' health and are at work? _____

ii. self-report 'very good' health and are at work? _____

```
proc freq;
      tables empstat;
      where health in (2);
      weight asecwt;
run:
```

```
proc freq;
      tables empstat;
      where health in (5) and age >=15;
      weight asecwt;
run;
```

```
proc freq;
      tables empstat;
      where health in (2) and age >=15;
      weight asecwt;
run;
```

Analyze the Data – Part III Relationships in the Data

A) What is the universe for AHRSWORKT? _____

B) What are the average hours of work for each self-reported health

```
proc means;
      class health;
      var ahrsworkt;
      where empstat = 10 and age >=15 and ahrsworkt < 999;
      weight asecwt;
run;
```

category? _____

• • •

ANSWERS

Analyze the Data – Part I Frequencies of FOODSTMP

A) On the website, find the codes page for the FOODSTMP variable and write down the code value, and what category each code represents. **o NIU; 1 No; 2 Yes**

B) What is the universe for FOODSTMP in 2011 (under the Universe tab on the website)? **All interviewed households and group quarters. Note the NIU on the codes page, this is a household variable and the NIU cases are the vacant households.**

C) How many people received food stamps in 2011? **39,187,348**

D) What proportion of the population received food stamps in 2011? **12.8%**

Using household weights (ASECWTH)

Suppose you were interested not in the number of people living in homes that received food stamps, but in the number of households that had food stamp participants. To get this statistic you would need to use the household weight.

To use household weight, you should be careful to select only one person from each household to represent that household's characteristics. You will need to apply the household weight (ASECWTH)

A) How many households received food stamps in 2011? **12,855,238 households**

B) What proportion of households received food stamps in 2011? **10.7% of households**

Section 4

Analyze the Data

ANSWERS

Analyze the Data – Part II Relationships in the Data

A) What is the universe for EMPSTAT in 2011? **1967 – 1987 age 14+; 1988 – 2011 age 15+**

B) What are the possible responses and codes for the self-reported HEALTH variable?   **Excellent; 2 Very Good; 3 Good; 4 Fair; 5 Poor**

C) What percent of people with 'poor' self-reported health are at work? **11.6%**

D) What percent of people with 'very good' self-reported health are at work? **51.6%**

E) In the EMPSTAT universe, what percent of people:

i. self-report 'poor' health and are at work? **11.8%**

ii. self-report 'very good' health and are at work? **64.0%**

Note: both age>=15 and empstat!=00 will ensure that you are in the EMPSTAT universe.

Section 5

Analyze the Data

ANSWERS

Analyze the Data – Part III Relationships in the Data

A) What is the universe for AHRSWORKT? **Civilians age 15+, at work last week**

B) What are the average hours of work for each self-reported health category? **Excellent – 38.4; Very Good – 38.66; Good – 37.78; Fair – 35.67; Poor – 32.41**