IPUMS – CPS Extraction and Analysis

Exercise 1

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

Research Questions

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 into Stata
- 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

Stata Code to Review

Code	Purpose
generate	Creates a new variable, "replace" specifies a value according to cases
mean	Displays a simple tabulation and frequency of one variable
tabulate	Displays a cross-tabulation for up to 2 variables
!=	Not equal to

Review Answer Key (page 7)

Common Mistakes to Avoid

1 Not changing the working directory to the folder where your data is stored

2 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.

3 Forgetting to put [fweight=*weightvar*] into square brackets

Note: In this exercise, for simplicity we will use "fweights". For any real analysis using CPS data, please *svyset* the data and use the Stata *svy* commands (with the accompanying *subpop* option).

Registering with IPUMS

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!

Step 1

Make an Extract

- Go to http://cps.ipums.org, click on Register with IPUMS
- On login screen, enter email address and password and submit
- Go back to homepage and go to Select Data
- Click the Select 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

● ● ●

Step 2

Request the Data

- 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

Getting the data into your statistics software

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

Step 1

Download the Data

- Go to http://cps.ipums.org and click on Download or Revise Extracts
- 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 Stata link next to the extract

• • •

Step 2

Decompress the Data

- 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 three files starting "cps_000..."
- Free decompression software is available at

https://www.7-zip.org

• • •

Step 3

Read in the Data

- Open Stata from the Start menu
- In "File" menu, choose "Change working directory..."
    Select "Documents", click "OK"
- In "File" menu, choose "Do..."
    Select the *.do file
- You will see "end of do-file" when Stata has finished reading in the data

Analyze the Sample – 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? _____

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

```
replace asecwt = round(asecwt)
tabulate foodstmp  [fweight=asecwt]
```

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 were food stamp participants. To get this statistic you would need to use the household weight.

In order to use household weight, you should be careful to select only one person from each household to represent that household's characteristics. This is accomplished using the "if pernum==1" qualifier in the example code below. 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?

_____

```
replace asecwth = round(asecwth)
tabulate foodstmp if pernum==1  [weight=asecwth]
```

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? _____

```
tab empstat  if health==5 [fweight=asecwt]
```

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

```
tab empstat  if health==2 [fweight=asecwt]
```

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

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

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

```
tab empstat  if health==5 & age>=15 [fweight=asecwt]
tab empstat  if health==2 & empstat!=00  [weight=asecwt]
```

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

A) What is the universe for AHRSWORKT? _____

B) What are the average hours of work (even if they did not work last week) for each self-reported health category?

_____

```
mean ahrsworkt if age>=15 & ahrsworkt <999
[weight=asecwt], over(health)
```

A) On the website, find the codes page for the FOODSTMP variable and write down the code value, and what category each code represents. 0 NIU; 1 No; 2 Yes

Section 1

Analyze the Data

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? 39,187,407

D) What proportion of the population received food stamps in 2011? 12.78%

```
replace asecwt = round(asecwt)
tabulate foodstmp  [fweight=asecwt]
```

• • •

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 were food stamp participants. To get this statistic you would need to use the household weight.

In order to use household weight, you should be careful to select only one person from each household to represent that household's characteristics. This is accomplished using the "if pernum==1" qualifier in the code below. You will need to apply the household weight (ASECWTH).

A) How many households received food stamps in 2011? 12,855,335 households

B) What proportion of households received food stamps in 2011? 10.71% of households

```
replace asecwth = round(asecwth)
tabulate foodstmp if pernum==1  [fweight=asecwth]
```

A) What is the universe for EMPSTAT in 2011?

age 15+

B) What are the possible responses and codes for the self-reported HEALTH variable? 1 Excellent; 2 Very good; 3 Good; 4 Fair; 5 Poor

C) What percent of people with 'poor' self-reported health are at work? 11.61%

```
tab empstat  if health==5 [fweight=asecwt]
```

D) What percent of people with 'very good' self-reported health are at work? 51.62%

```
tab empstat  if health==2 [fweight=asecwt]
```

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

i. self-report 'poor' health and are at work? 11.81%

```
tab empstat  if health==5 & age>=15 [weight=asecwt]
tab empstat  if health==2 & empstat!=00  [weight=asecwt]
```

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.

A) What is the universe for AHRSWORKT?

Civilians age 15+, at work last week

B) What are the average hours of work (even if they did not work last week) for each self-reported health category? Excellent – 38.4; Very Good – 38.66; Good – 37.78;   Fair – 35.67; Poor – 32.41

```
mean ahrsworkt if age>=15 & ahrsworkt <999
[weight=asecwt], over(health)
```