

# 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 the relationship between health and work status and to create basic frequencies of food stamp usage.

## Research Questions

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

## Objectives

- Create and download an IPUMS data extract
- Decompress data file and read data into 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;	Begins a frequency procedure
proc means;	Begins a means procedure, returns the mean value of a variable
tables	Required syntax to display frequencies
where	Selects only specified cases to include in a procedure

## Review Answer Key (page 7)

### 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://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 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 SAS. If you would like to use a different stats package, see: [http://cps.ipums.org/cps/extract\\_instructions.shtml](http://cps.ipums.org/cps/extract_instructions.shtml)

### Step 1

#### Download the Data

...

### Step 2

#### Decompress the Data

...

### Step 3

#### Read in 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 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 three files starting "cps\_000..."
  - Free decompression software is available at <https://www.7-zip.org>
- 
- Open the "cps\_000##.sas" file.
  - In the do 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\_000##.dat";
  - Choose Submit under the Run file menu.

## Analyze the Sample – Part I Frequencies of FOODSTMP

### Section 1

#### Analyze the Data

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? \_\_\_\_\_

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

•••

### Section 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.

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?  
\_\_\_\_\_

```
proc freq;
    tables foodstmp;
    where pernum in (1);
    weight aseawth;
run;
```

## Analyze the Sample – Part II Relationships in the Data

### Section 1

### 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? \_\_\_\_\_

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

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

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

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? \_\_\_\_\_

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

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

## Analyze the Sample – Part III Relationships in the Data

### Section 1

#### Analyze the Data

...

### Complete!

#### Validate Your Answers

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? \_\_\_\_\_

```
proc means;
    class health;
    var ahrsworkt;
    where age >=15 and ahrsworkt < 999;
    weight aseawt;

run;
```

## ANSWERS: Analyze the Sample – Part I Frequencies of FOODSTMP

### Section 1

#### Analyze the Data

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**

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%**

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

### *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.

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?  
**12,855,283 households**

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

•••

### Section 2

#### Weighting the Data



## ANSWERS: Analyze the Sample – Part II Relationships in the Data

### Section 1

#### Analyze 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%**

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

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

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

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%**

```
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;
```

## ANSWERS: Analyze the Sample – Part III Relationships in the Data

### Section 1

#### Analyze the Data

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**

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