

# IPUMS – CPS Extraction and Analysis

## Exercise 2

**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 parent and child health, and analyzing relationships between disability variables and marital status.

## Research Questions

Is there an association between parent and child health? What are the trends in disabilities and marital status?

## 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

- AGE: Age
- SEX: Sex
- MARST: Marital status
- HEALTH: Health status
- DIFFHEAR: Hearing difficulty
- DIFFEYE: Vision difficulty

## Stata Code to Review

Code	Purpose
<code>generate</code>	Creates a new variable, "replace" specifies a value according to cases
<code>mean</code>	Displays a simple tabulation and frequency of one variable
<code>tabulate</code>	Displays a cross-tabulation for up to 2 variables
<code>!=</code>	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 [weight=*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 State *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 the homepage and go to Select Data
- Click the Select/Change Samples box, check the box for the 2010 and 2011 ASEC samples, then click Submit Sample Selections
- Using the drop down menu or search feature, select the following variables:
  - AGE: Age
  - SEX: Sex
  - MARST: Marital status
  - HEALTH: Health status
  - DIFFHEAR: Hearing difficulty
  - DIFFEYE: Vision difficulty

...

### Step 2

#### Request the Data

- Click the VIEW CART button under your data cart
- Review variable selection. Click the Create Data Extract button
- Click on 'Attach Characteristics'
- The following screen will allow you to select who you would like to attach variables for, it should look like this:

Attached Variable Selection				
Variable	Head	Father	Mother	Spouse
<a href="#">AGE</a>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<a href="#">HEALTH</a>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<a href="#">DIFFHEAR</a>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<a href="#">DIFFEYE</a>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

- Describe your extract and click Submit Extract
- 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
- Use the steps in the next section to get the data into your statistical package

## 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](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 Stata 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 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 Creating New Variables

### Section 1

### Analyze the Data

A) What are the names of the attached variables (can be found on extract request screen or in the data)?  
\_\_\_\_\_

B) On the website, find the FAQ entry for attaching characteristics. What value will the respondents without a parent or spouse present have for the attached variables? \_\_\_\_\_

C) What are the MARST codes for married respondents?  
\_\_\_\_\_

D) Create a variable for married men equal to the difference in spouses' age.

```
gen age_dif = age - age_sp if sex==1&(marst==1 | marst==2)
```

E) What is the mean age difference between married men and their spouses? \_\_\_\_\_

For men aged 30 and under? \_\_\_\_\_ For 50 and over? \_\_\_\_\_

```
mean age_dif [pweight=asecwt]  
mean age_dif if age<=30 [pweight=asecwt]  
mean age_dif if age>=50 [pweight=asecwt]
```

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

### Section 1

#### Analyze the Data

A) What is the universe for DIFFEYE and DIFFHEAR? What is the Code for NIU (Not in Universe)? \_\_\_\_\_

B) What percent of the population (in the universe) is deaf or has a serious hearing difficulty? \_\_\_\_\_

What percent of the population (in the universe) is blind or has serious sight difficulties? \_\_\_\_\_

```
replace asewt = round(asewt)
tab diffhear if diffhear>0 [weight=asewt]
tab diffeye if diffeye>0 [weight=asewt]
```

C) What percent of the deaf population is married with a spouse present? \_\_\_\_\_

```
tab marst if diffhear==2 [weight=asewt]
```

D) What percent of the deaf population is married to a spouse who is also deaf? \_\_\_\_\_

```
gen couple_deaf = 1 if diffhear==2&diffhear_sp==2
replace couple_deaf = 0 if diffhear==2&diffhear_sp!=2

tab couple_deaf [fw= asewt]
```

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

### Section 1

#### Analyze the Data

A) What ages of respondents have their parents identified through the attach characteristics? (hint: see variable descriptions for MOMLOC and POPLOC). \_\_\_\_\_

B) Does there seem to be a relationship between parents and children's health? \_\_\_\_\_

```
tab health health_mom [weight=asecwt], row  
tab health health_pop [weight=asecwt], row
```

C) What other tests could you do to examine this relationship?  
\_\_\_\_\_

D) Could there be a sampling issue affecting the relationship between children and parent's health? \_\_\_\_\_

...

### Section 2

Complete!  
Check  
your  
Answers!



## ANSWERS: Analyze the Sample – Part I Creating New Variables

### Section 1

#### Analyze the Data

A) What are the names of the attached variables (can be found on extract request screen, or in the data)? AGE\_SP, age of spouse; HEALTH\_MOM, health of mother; HEALTH\_POP, health of father; HEALTH\_SP, health of spouse; DIFFHEAR\_SP, hearing disability of spouse; DIFFEYE\_SP, vision disability of spouse

B) On the website, find the FAQ entry for attaching characteristics. What value will the respondents without a parent or spouse present have for the attached variables? A missing code

C) What are the MARST codes for married respondents? 1 Married, spouse present; 2 Married, spouse absent

D) Create a variable for married men equal to the difference in spouses' age.

```
gen age_dif = age - age_sp if sex==1&(marst==1|marst==2)
```

E) What is the mean age difference between married men and their spouses? 2.3

For men aged 30 and under? -.16 For 50 and over? 3.2

```
replace asewt = round(asewt)
mean age_dif [fweight=asewt]
mean age_dif if age<=30 [fweight=asewt]
mean age_dif if age>=50 [fweight=asewt]
```



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

### Section 1

#### Analyze the Data

A) What is the universe for DIFFEYE and DIFFHEAR? What is the Code for NIU (Not in Universe)?

Persons age 15+, 0 is the NIU code.

B) What percent of the population (in the universe) is deaf or has a serious hearing difficulty? 3.1%

What percent of the population (in the universe) is blind or has serious sight difficulties? 1.7%

```
tab diffhear if diffhear>0 [fweight=asecwt]
tab diffeye if diffeye>0 [fweight=asecwt]
```

C) What percent of the deaf population is married with a spouse present? 49.7%

```
tab marst if diffhear==2 [fweight=asecwt]
```

D) What percent of the deaf population is married to a spouse who is also deaf? 7.7%

```
gen couple_deaf = 1 if diffhear==2&diffhear_sp==2
replace couple_deaf = 0 if diffhear==2&diffhear_sp!=2

tab couple_deaf [fweight=asecwt]
```

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

### Section 1

#### Analyze the Data

A) What ages of respondents have their parents identified through the attach characteristics? (hint: see variable descriptions for MOMLOC and POPLOC). Children under age 19

B) Does there seem to be a relationship between parents and children's health? Parent's health and children's health seem to be directly correlated.

```
tab health health_mom [fweight=asecwt], row  
tab health health_pop [fweight=asecwt], row
```

C) What other tests could you do to examine this relationship? Correlation matrix, covariance analysis, regression analysis

D) Could there be a sampling issue affecting the relationship between children and parent's health? Parents are reporting children's health