

# 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 SAS
- 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

## 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 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 ";"

## Registering with IPUMS

Go to <http://cps.ipums.org>, click on CPS Registration 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 green VIEW CART button under your data cart
- Review variable selection. Click the green 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 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

- 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

•••

### 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 <http://www.irisnet.net/soft/wingzip/>

•••

### Step 3

#### Read in the Data

- 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 ASCIIIDAT", enter the full file location, ending with cps\_000##.dat";
- Choose Submit under the Run file menu

## 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.

```
data ipums.cps_000##;
    set ipums.cps_000##;
    agedif = _null_;
    if sex = 1 and (marst = 1 or marst = 2) then
        agedif = age - age_sp;
run;
```

*Note: Enter the filename with the extract number after "data" and "set". ie. replace the # symbols with the number of your data extract.*

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

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

```
proc means;
    var agedif;
    weight wtsupp;
run;
proc means;
    var agedif;
    where age<=30;
    weight wtsupp;
run;
proc means;
    var agedif;
    where age>=50;
    weight wtsupp;
run;
```

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

```
proc freq;
    where diffhear > 0;
    tables diffhear;
    weight wtsupp;
run;
proc freq;
    where diffeye > 0;
    tables diffeye;
    weight wtsupp;
run;
```

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

```
proc freq;
    where diffhear in (2);
    tables marst;
    weight wtsupp;
run;
```

*Exercise Continues Below...*

## Analyze the Sample – Part II Continued

### Section 1

### Analyze the Data

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

```
data ipums.cps_000##;
  set ipums.cps_000##;
  couple_deaf = _null_;
  if diffhear = 2 and diffhear_sp = 2 then couple_deaf = 1;
  if diffhear = 2 and diffhear_sp ^=2 then couple_deaf = 0;
run;

proc freq;
  tables couple_deaf;
  weight wtsupp;
run;
```

## Section 1

### Analyze the Data

...

## Section 2

Complete!  
Check  
your  
Answers!

### Analyze the Sample – Part III Relationships in 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? \_\_\_\_\_

```
proc freq;
  tables health*health_mom;
  weight wtsupp;
run;

proc freq;
  tables health*health_pop;
  weight wtsupp;
run;
```

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? \_\_\_\_\_



## 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.

```
data ipums.cps_000##;  
    set ipums.cps_000##;  
    agedif = _null_;  
    if sex = 1 and (marst = 1 or marst = 2) then  
        agedif = age - age_sp;  
run;
```

E) What is the mean age difference between married men and their spouses? **2.3** For men 30 and under? **-.09** For 50 and over? **3.2**

## Section 1

### Analyze the Data

```
proc means;
    var agedif;
    weight wtsupp;
run;
proc means;
    var agedif;
    where age<=30;
    weight wtsupp;
run;
proc means;
    var agedif;
    where age>=50;
    weight wtsupp;
run;
```

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

- A) What is the universe for DIFFEYE and DIFFHEAR? What is the Code for NIU (Not in Universe)? **Persons age 15+, 0**
- B) What percent of the population (in the universe) is deaf or has

```
proc freq;
    where diffhear > 0;
    tables diffhear;
    weight wtsupp;
run;
proc freq;
    where diffeye > 0;
    tables diffeye;
    weight wtsupp;
run;
```

## ANSWERS: Analyze the Sample – Part II Continued

### Section 1

#### Analyze the Data

a serious hearing difficulty? 3.1%

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

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

```
proc freq;
    where diffhear in (2);
    tables marst;
    weight wtsupp;
run;
```

D) What percent of the deaf population is married to a spouse who is also deaf? 7.37%

## Section 1

### Analyze the Data

```
data ipums.cps_000##;  
  set ipums.cps_000##;  
  couple_deaf = _null_;  
  if diffhear = 2 and diffhear_sp = 2 then couple_deaf = 1;  
  if diffhear = 2 and diffhear_sp ^=2 then couple_deaf = 0;  
run;  
  
proc freq;  
  tables couple_deaf;  
  weight wtsupp;  
run;
```

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

```
proc freq;
    tables health*health_mom;
    weight wtsupp;
run;

proc freq;
    tables health*health_pop;
    weight wtsupp;
run;
```

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? **Yes, parents are reporting children's health**