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 SPSS
- 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

SPSS Code to Review

Code	Purpose
compute	Creates a new variable
freq	Displays a simple tabulation and frequency of one variable
crosstabs	Displays a cross-tabulation for up to 2 variables and a control
~=	Not equal to

Review Answer Key (page 7)

Common Mistakes to Avoid

1 Excluding cases you don't mean to. Avoid this by turning off weights and select cases after use, otherwise they will apply to all subsequent analyses

2 Terminating commands prematurely or forgetting to end commands with a period (.) Avoid this by carefully noting the use of periods in this exercise

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 samples for March, 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
AGE	☐	☐	☐	☑
HEALTH	☐	☑	☑	☑
DIFFHEAR	☐	☐	☐	☑
DIFFEYE	☐	☐	☐	☑

- 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 SPSS. 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 SPSS 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.irnis.net/soft/wingzip/

• • •

Step 3

Read in the Data

- Double click on the ".sps" file, which should automatically have been named "cps_000....."

- The first two lines should read:

    cd ".".

    data list file = 'cps_000...'/

- Change the first line to read: cd (location where you've been saving your files). For example:

    cd "C:\Documents".

- Change the second line to read:

    data list file = "C:\Documents\cps_000...dat"/

- Under the "Run" menu, select "All" and an output viewer window will open

- Use the Syntax Editor for the SPSS code below, highlight the code, and choose "Selection" under the Run menu

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.

> compute age_dif = age - age_sp.

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

For men aged 30 and under? _____   For 50 and over? _____

> weight by wtsupp.
>
> *(Data menu/Select Cases/If... type "sex = 1 and (marst = 1 or marst = 2)")*
>
> means age_dif.
>
> *(Data menu/Select Cases/If... also type "age <=30")*
>
> means age_dif.
>
> *(Data menu/Select Cases/If... replace other age condition with type "age >=50")*
>
> means age_dif.

Section 1

Analyze the Data

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? _____

```
select if (diffhear ~=0 and diffeye ~=0).
        freq diffhear.
        freq diffeye.
```

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

```
crosstabs
/tables = diffhear by marst
/cells = count row.
```

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

```
compute couple_deaf = 0.
if (diffhear = 2 and diffhear_sp = 2) couple_deaf = 1.

crosstabs
/tables = diffhear by couple_deaf
/cells = count row.
```

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? _____

```
crosstabs
/tables = health by health_mom
/cells = count row.


crosstabs
/tables = health by health_pop
/cells = count 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? _____

A) What are the names of the attached variables (can be found on extract request screen, or in the data)? <u>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</u>

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? <u>A missing code</u>

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

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

```
compute age_dif = age - age_sp.
```

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

For men aged 30 and under? <u>-.09</u> For 50 and over? <u>3.2</u>

```
weight by wtsupp.
 (Data menu/Select Cases/If… type "sex = 1 and (marst = 1 or marst = 2)")
means age_dif.
(Data menu/Select Cases/If… also type "age <=30")
means age_dif.
(Data menu/Select Cases/If… replace other age condition with  type "age >=50")
means age_dif.
```

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%

```
select if (diffhear ~=0 and diffeye ~=0).
        freq diffhear.
        freq diffeye.
```

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

```
crosstabs
/tables = diffhear by marst
/cells = count row.
```

D) What percent of the deaf population is married to a spouse who is also deaf? 7.4%

```
compute couple_deaf = 0.
if (diffhear = 2 and diffhear_sp = 2) couple_deaf = 1.

crosstabs
/tables = diffhear by couple_deaf
/cells = count row.
```

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). <u>Children under age 19</u>

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

```
crosstabs
/tables = health by health_mom
/cells = count row.

crosstabs
/tables = health by health_pop
/cells = count row.
```

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

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