IPUMS Data Training Exercise:

An introduction to IPUMS NHIS

(Exercise 1 for SPSS)

Learning goals

- Understand how IPUMS NHIS dataset is structured
- Create and download an NHIS data extract
- Decompress data file and read data into SPSS
- Analyze the health insurance coverage, educational attainment, and flu shot attainment of people in the United States using sample code

Summary

In this exercise, you will gain an understanding of how the NHIS dataset is structured and how it can be leveraged to explore your research interests. This exercise will use the NHIS dataset to explore basic frequencies of flu vaccination, health insurance coverage, educational attainment, and overall health status. You will create data extracts that include the variables HINOTCOVE, EDUCREC2, HEALTH, and VACFLUSH12M; then you will use the sample code to analyze these data.

SPSS Code to Review

Code	Purpose
compute	Creates a new variable
frequencies	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

Common Mistakes to Avoid

- Excluding cases you don't mean to. Avoid this by turning off weights and select cases after use or using the temporary command, otherwise changes in selected cases will apply to all subsequent analyses.
- 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 NHIS

Go to https://nhis.ipums.org/nhis/, click on Log in and login if you are a registered user. If you are a first time user, click on Create an account, enter an email address and password, and then submit your user information so you can create NHIS data extracts.

Make a Data Extract

- Return to the homepage and click on Browse and Select Data.

Select Samples

- Click the Select Samples button, and check the box for the 2010 sample. Click the submit sample selections button.

Select Variables

- The variable drop-down menus allow you to explore variables by topic. For example, you might expect to find variables about flu shots under the "Vaccinations" group.

- The search tool allows you to search for variables. Observe the options for limiting your search results by variable characteristics or variable type.

- You may add a variable to your cart by clicking on the plus sign in the "Add to Cart" column of the topical variable list, or list of search results.

- You may view information about the variable by clicking on the variable name, and navigating through the tabs that include a description of the variable, codes and value labels, the universe of persons asked the question, and information on the comparability of the variable among other pieces of information. If you are reviewing variable-specific information, you may click on the "Add to Cart" button near the top of the screen to add this variable to your data cart.

- Using the drop down menu or search feature, select the following variables and add them to your data cart using the plus symbol to the left of the variables:

  > HINOTCOVE: Health insurance status

  > EDUCREC2: Educational attainment

Review and submit your extract

- Click the green VIEW CART button under your data cart.

- Review variable selection. Note that additional variables are in your data cart. The data extract system automatically supplies variables that indicate the sample (YEAR), are needed for variance estimation (SERIAL, PERNUM), and are used for weighting the variables and years selected. Click the green Create Data Extract button.

- Review the 'Extract Request Summary' screen, describe your extract, and click Submit Extract.

- You will receive an email when the data is available to download.
- To access the page to download the data, follow the link in the email, or click on the Download or Revise Extracts link on the homepage.

Create two additional extracts

- Create an extract using the 1972, 1981, 1997, and 2010 samples and the HEALTH variable.
- Create an extract using the samples of years 1997 through 2010 and the VACFLUSH12M variable.

Getting the data into your statistics software

The following instructions are for SPSS. If you would like to use a different stats package, see: https://ipums.org/support/exercises

Download the data

- Go to https://nhis.ipums.org/nhis/ and click on Download or Revise Extracts.
- Right-click on the Data link next to the extract you created.
- Choose "Save Target As…" (or "Save Link As…").
- Save into "Documents" (Documents should pop up as the default location).
- Do the same for the SPSS link next to the Data link.

Decompress the data

- All IPUMS extracts are compressed. We recommend 7zip for Windows users. Macs can open these types of files without additional software.
- Find the "Documents" folder under the Start menu.
- Right click on the ".dat" file.
- Use your decompression software to extract the .dat files.
- Double-check that the Documents folder contains three files starting with "nhis_000…".

Read the data

- Double click on the ".sps" file, which should automatically have been named "nhis_000...".
- The first two lines should read:

```
cd ".".
data list file = "nhis_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\nhis_000...dat"/
```

- Under the "Run" menu, select "All"; 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.

Analyze the Sample

Part 1: Frequencies

These questions use the first data extract with the variables HINOTCOVE and EDUCREC2 for the 2010 sample.

1. On the website, find the universe page for the HINOTCOVE variable and write down the universe statement, which indicates who was asked this specific question.

   _____

   _____

2. How many people in the 2010 sample report being uninsured? _____

3. What proportion of the 2010 sample report being uninsured? _____

```
frequencies

    /variables=hinotcove.
```

Using person weights (PERWEIGHT)

4. To get a more accurate estimation of demographic patterns, you will have to utilize the person weight. Using weights:

   a. How many people were uninsured in 2010? _____

   b. What proportion of the population was uninsured in 2010? _____

```
weight by perweight.
frequencies
   /variables=hinotcove.
```

5. On the website, examine the variable description for EDUCREC2 and write down the universe statement. _____

6. Using weights, how many people had a 4 year college or Bachelor's degree as their highest educational attainment? _____

7. Using weights, what proportion of the population had a 4 year college or Bachelor's degree as their highest educational attainment? _____

```
weight by perweight.

frequencies

    /variables=educrec2.
```

Part 2: Relationships in the Data

These questions require the second data extract using the 1972, 1981, 1997, and 2010 samples and the HEALTH variable.

8. Using weights, determine the proportion of the population that reported excellent health status over time. Note: You'll want to exclude the unknown responses for HEALTH using the Select Cases function in SPSS to exclude them. On the website, check the codes for HEALTH.

```
select if (health<=5).

weight by perweight.

crosstabs

    /tables=year by health

    /cells=count row.
```

    1972: _____   1997: _____

    1981: _____   2010: _____

9. An initial glance may lead you to conclude that excellent health has declined since 1972. This interpretation is complicated by a change in the data collection during this time period. Using the website, navigate to the HEALTH variable description and find the year that this variable changed from a four-point scale to a five-point scale. _____

These questions require you to use the third data extract with the VACFLUSH12M variable for the samples of years 1997 through 2010.

10. Examine the documentation for the flu shot variable (*VACFLUSH12M*) and write down the universe statements from 1997 to 2010. _____

11. Suppose you want to examine trends in the proportion who reported Influenza vaccination during the past 12 months using the extracted data. Since this variable was only for a sample person we will use the sample weight (SAMPWEIGHT) instead of the person weight.

    Which survey years had the highest and lowest percentage receiving the vaccine within the past 12 months?

    Highest: _____   Lowest: _____

```
weight by sampweight.
crosstabs
    /tables=year by vacflush12m
    /cells = count row.
```

EXERCISE 1 FOR SPSS

Answers

Part 1: Frequencies

1. On the website, find the universe page for the HINOTCOVE variable and write down the universe statement, which indicates who was asked this specific question.
   <u>1988: Sample persons under age 18. 1998-2010: All persons.</u>

2. How many people in the 2010 sample report being uninsured? <u>16,029 individuals in the sample</u>

3. What proportion of the 2010 sample report being uninsured? <u>17.81% of the sample</u>

Using person weights (PERWEIGHT)

4. Using weights:
   a. How many people were uninsured in 2010? <u>48,311,184 individuals</u>

   b. What proportion of the population was uninsured in 2010? <u>15.9% of the population</u>

5. On the website, examine the variable description for EDUCREC2 and write down the universe statement. <u>1982-2010: Persons age 5+.</u>

6. Using weights, how many people had a 4 year college or Bachelor's degree as their highest educational attainment? <u>40,229,764</u>

7. Using weights, what proportion of the population had a 4 year college or Bachelor's degree as their highest educational attainment? <u>13.23%</u>

Part 2: Relationships in the Data

8. Determine the proportion of the population that reported excellent health status over time.

        1972: <u>51.8%</u>   1997: <u>38.3%</u>

        1981: <u>49.3%</u>   2010: <u>35.2%</u>

9. An initial glance may lead you to conclude that excellent health has declined since 1972. This interpretation is complicated by a change in the data collection during this time period. Using the website, navigate to the HEALTH variable description and find the year that this variable changed from a four-point scale to a five-point scale. <u>1982</u>

10. Examine the documentation for the flu shot variable (VACFLUSH12M) and write down the universe statements from 1997 to 2010. <u>1997-2004: Sample adults age 18+; 2005-2010: Sample adults age 18+ and sample children under age 18</u>.

11. Which survey years had the highest and lowest percentage receiving the vaccine within the past 12 months?

    Highest: <u>2010</u>

    Lowest: <u>2005</u>