IHIS Extraction and Analysis

Exercise 1

OBJECTIVE: Gain an understanding of how the IHIS dataset is structured and how it can be leveraged to explore your research interests. This exercise will use the IHIS dataset to explore basic frequencies of flu vaccination, health insurance coverage, and educational attainment, and the relationship between overall health status and employment status.

Research Questions

What is the distribution of insurance coverage and educational attainment in the United States? How many people in the US receive a flu shot every year?

Objectives

- Create and download an IHIS data extract
- Decompress data file and read data into SAS
- Analyze the data using sample code
- Validate data analysis work using answer key

IHIS Variables

- HINOTCOVE: Health Insurance Status
- EDUCREC2: Education attainment
- EMPSTAT: Employment status
- HEALTH : Self-reported health status
- SHOTFLUYR: Flu vaccination within the past 12 months

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 - Forgetting to close a procedure with "run;".

4 - Forgetting to terminate a command with a semicolon ";".

Registering with IHIS

Go to http://www.ihis.us, click on User Registration and Login and Apply for access. Log in if you are a registered user. If you are a first time user, enter an email address and password, then submit your user information so you can create IHIS data extracts.

Step 1

Make an Extract

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

▪ Click the Select Samples box, and check the box for the 2010 sample. Click the Submit sample selections box.

▪ 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 variable:

> HINOTCOVE: Health Insurance Status
>
> EDUCREC2: Education attainment

• • •

Step 2

Request the Data

▪ 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.

▪ Note: there are three different data extracts required to complete the exercises included in this tutorial; you may create the data extracts as you go, or may want to look ahead and create all three extracts before beginning the exercises.

▪ 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.

Getting the data into your statistics software

The following instructions are for SAS. If you would like to use a different stats package, see: https://www.ihis.us/ihis/extract_instructions.shtml

Step 1

Download the Data

- Go to http://www.ihis.us/ 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 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 the .dat file.
- Double-check that the Documents folder contains three files starting "ihis_000…".
- Free decompression software is available at http://www.irnis.net/soft/wingzip/

• • •

Step 3

Read in the Data

- Open the "ihis_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 "ihis_000##.dat";
- Choose Submit under the Run file menu.

Analyze the Sample – Part I Frequencies

A) 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. _____

B) How many people in 2010 sample report being uninsured? _____

C) What proportion of the 2010 sample report being uninsured? _____

```
proc freq;
        tables hinotcove;
run;
```

Using person weights (PERWEIGHT)

To get a more accurate estimation of demographic patterns within a country from the sample, you will have to utilize the person weight.

A) Using weights:

i. How many people were uninsured in 2010? _____

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

```
proc freq;
        tables hinotcove;
        weight perweight;
run;
```

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

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

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

```
proc freq;
        tables educrec2;
        weight perweight;
run;
```

Analyze the Sample – Part II Relationships in the Data

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

Section 1

Analyze the Data

A) Determine the proportion of the population that reported excellent health status over time. Note: You'll want to exclude the unknown responses for HEALTH, so use a conditional command in SAS to exclude them. On the website, check the codes for HEALTH.

```
proc freq;
        tables health;
        by year;
        where health <7;
        weight perweight;
run;
```

1972: _____

1981: _____

1997: _____

2010: _____

• • •

Section 2

Thinking Critically

B) 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. _____

Analyze the Sample – Part III Relationships in the Data

These questions require you to create a third extract using samples of years 1997 through 2010, and the SHOTFLUYR variable.

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

B) 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. Also, exclude respondents who were not in the universe (ie they were not asked the question). The code "where SHOTFLUYR = 1 or SHOTFLUYR=2;" tells SAS to only consider cases with a No or a Yes response.

```
proc freq;
        where shotfluyr = 1 or shotfluyr = 2;
        tables shotfluyr;
        weight sampweight;
        by year;
run;
```

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

Highest: _____

Lowest: _____

Section 1

Analyze the Data

A) 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>

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

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

```
proc freq;
        tables hinotcove;
run;
```

• • •

Section 2

Weighting the Data

Using person weights (PERWEIGHT)

To get a more accurate estimation of demographic patterns within a country from the sample, you will have to utilize the person weight.

A) Using weights:

i. How many people were uninsured in 2010? <u>48,311,184 individuals</u>

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

```
proc freq;
        tables hinotcove;
        weight perweight;
run;
```

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

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

D) 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>

```
proc freq;
        tables hinotcove;
        weight perweight;
run;
```

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

A) Determine the proportion of the population that reported excellent health status over time. Note: You'll want to exclude the unknown responses for HEALTH, so use a conditional command in SAS to exclude them. On the website, check the codes for HEALTH.

1972: 51.8%

1981: 49.3%

1997: 38.3%

2010: 35.2%

```
proc freq;
        tables health;
        by year;
        where health < 7;
        weight perweight;
run;
```

B) 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. 1982

These questions require you to create a third extract using samples of years 1997 through 2010, and the SHOTFLUYR variable.

<table>
<tr>
<td>

Section 1

Analyze the Data

</td>
<td>

A) Examine the documentation for the flu shot variable (SHOTFLUYR) 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>

B) 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. Also exclude respondents who did not answer either yes or no to the question using the code "where SHOTFLUYR = 1 or SHOTFLUYR=2;".

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

</td>
</tr>
</table>

```
proc freq;
        where shotfluyr = 1 or shotfluyr = 2;
        tables shotfluyr;
        weight sampweight;
        by year;
run;
```

Highest: <u>2010</u>

Lowest: <u>2005</u>