Minnesota Population Center
Training and Development

IHIS Extraction and Analysis

Exercise 2

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 associations among BMI, poverty, health status, sleep, and frequency of exercise.

1/30/2014

Research Questions

Is there a relationship between hours of sleep and health status? What is the relationship between health and poverty status, controlling for physiology?

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

- AGE: Age
- SEX: Sex
- POORYN: Above or Below Poverty threshold
- HEALTH: Health status
- BMI: Body Mass Index
- HRSLEEP: Usual hours of sleep per day
- VIG10FWK: Frequency of vigorous activity (10+min) per week

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 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 menus or search feature, select the following variables and add them to your data cart using the plus symbol to the left of the variable:

UNINSURED: Health Insurance Status

EDUCREC2: Education attainment

EMPSTAT: Employment status

HEALTH: Self-reported health status

SHOTFLUYR: Flu vaccination within the past 12 months

•••

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.
- 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.

Section 1

Analyze the Data

Analyze the Sample – Part I Group Means

A) On the website, find the codes page for the HRSLEEP and HEALTH variables. What code values for HRSLEEP should be excluded to avoid skewing the average number of hours slept? How would you restrict the code values for HEALTH to eliminate unknown responses?

_____

B) Suppose you wanted to study the relationship between hours of sleep and health status. Determine the average reported hours of sleep per night by health status. On average, how many hours does an individual with excellent health in this sample sleep per night? _____

```
proc means;
        where health < 6 and hrsleep > 0 and hrsleep <25;
        var hrsleep;
        class health;
run;
```

C) Is there a noticeable trend between health status and hours of sleep using this sample? _____

D) Does the trend change for people under 60 in this sample?

_____

```
proc means;
        where health < 6 and hrsleep > 0 and hrsleep <25 and age <60;
        var hrsleep;
        class health;
run;
```

Analyze the Sample – Part II Weighting the Data

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

Section 1

Weight the Data

A) Without weights, what proportion of people in this sample was below the poverty threshold in 2010? _____

```
proc freq;
        tables pooryn*year;
run;
```

B) Using weights, what proportion of the population was below the poverty threshold in 2010? _____

```
proc freq;
        tables pooryn*year;
        weight perweight;
run;
```

C) Using the household weight (and you must exclude all but one individual from a household), what proportion of households was below the poverty threshold in 2010? _____

```
proc freq;
        where pernum = 1;
        tables pooryn*year;
        weight hhweight;
run;
```

Analyze the Sample – Part III Generating Variables

Generate a variable that is 0 when an individual exercises less than 3 times a week, and 1 when an individual exercises 3 or more times a week.

Section 1

Analyze the Data

A) Check the output of the do file in the Log window to find the codes for VIG10FWK. Which code means "Never"? _____

Note: You'll have to exclude codes above 28 when defining when exer3 is greater than 3 times a week.

B) What is the average difference in BMI for an individual in this sample who exercises at least 3 times a week compared to someone who exercises fewer than 3 times per week? _____

Remember to restrict the codes for BMI so unknown and missing codes are excluded.

```
data ipums.ihis_000##;
        set ipums.ihis_000##;
        exer3 = 0;
        if vig10fwk >=3 & vig10fwk<=28 then exer3 = 1;
run;

proc means;
        where bmi >0 and bmi<99;
        var bmi;
        class exer3;
run;
```

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

C) What percent of more frequent exercisers report excellent health? _____ Less frequent exercisers? _____

```
proc freq;
        where health < 6;
        tables health*exer3;
run;
```

Analyze the Sample – Part IV Graphing

Create a graph to show the mean BMI over age for males and females.

Section 1

Analyze the Data

A) How does the universe for BMI appear on this graph?

_____

B) Approximately at what age does BMI peak for women? For men?

_____

```
proc means;
where bmi > 0 and bmi < 99;
var bmi;
class sex age;
output out=sumstat mean=AverageBmi;
run;

symbol Interpol=join
        value=dot;
proc gplot data=sumstat;
where sex ne .;
plot AverageBmi*age=sex;
run;
```

...

Section 2

Include a New Variable

Note: SAS graph procedures do not allow for WEIGHT options, so graph analyses are at the sample level

C) Create a graph to show how an associated effect of poverty status on BMI differs with gender, controlling for frequent exercise. Comment on three apparent trends.

```
proc sgpanel data=ipums.ihis_000##;
        where bmi > 0 and bmi < 99 and pooryn^=9;
        panelby sex exer3;
        vbar pooryn/ stat = mean
        response=bmi;
run;
```

...

Complete!

Validate Your Answers

ANSWERS: Analyze the Sample – Part I Group Means

Section 1

Analyze the Data

A) On the website, find the codes page for the HRSLEEP and HEALTH variables. What code values for HRSLEEP should be excluded to avoid skewing the average number of hours slept? How would you restrict the code values for HEALTH to eliminate unknown responses?

HRSLEEP: 00 NIU; 25 Less than 1 hour; 97 Unknown-refused; 98 Unknown-not ascertained; 99 Unknown-don't know HEALTH: 7 Unknown-refused; 8 Unknown-not ascertained; 9 Unknown-don't know.

B) Suppose you wanted to study the relationship between hours of sleep and health status. Determine the average reported hours of sleep per night by health status. On average, how many hours does an individual with excellent health in this sample sleep per night? 7.2 hours

```
proc means;
        where health < 6 and hrsleep > 0 and hrsleep <25;
        var hrsleep;
        class health;
run;
```

C) Is there a noticeable trend between health status and hours of sleep using this sample? There seems to be no trend at all, except perhaps Excellent and Poor health have slightly higher averages, which could indicate people in poor health sleep to improve and people with excellent health are associated with getting more sleep.

```
proc means;
        where health < 6 and hrsleep > 0 and hrsleep <25 and age <60;
        var hrsleep;
        class health;
run;
```

D) Does the trend change for people under 60 in this sample? When excluding the older population (perhaps with a higher incidence of poor health), better health is associated with more hours of sleep, though the differences between averages is small.

ANSWERS: Analyze the Sample – Part II Weighting the Data

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

Section 1

Weight the Data

A) Without weights, what proportion of people in this sample was below the poverty threshold in 2010? 16.48% of the sample

```
proc freq;
        tables pooryn*year;
run;
```

B) Using weights, what proportion of the population was below the poverty threshold in 2010? 13.76% of the sample

```
proc freq;
        tables pooryn*year;
        weight perweight;
run;
```

C) Using the household weight (and you must exclude all but one individual from a household), what proportion of households was below the poverty threshold in 2010? 12.91% of the sample

```
proc freq;
        where pernum = 1;
        tables pooryn*year;
        weight hhweight;
run;
```

ANSWERS: Analyze the Sample – Part III Generating Variables

Generate a variable that is 0 when an individual exercises less than 3 times a week, and 1 when an individual exercises 3 or more times a week.

Section 1

Analyze the Data

A) Check the output of the do file in the Log window to find the codes for VIG10FWK. Which code means "Never"? 95 Never

Note: You'll have to exclude codes above 28 when defining when exer3 is greater than 3 times a week.

B) What is the average difference in BMI for an individual in this sample who exercises at least 3 times a week compared to someone who exercises fewer than 3 times per week? 1.2 BMI (27.7-26.5)

```
data ipums.ihis_000##;
        set ipums.ihis_000##;
        exer3 = 0;
        if vig10fwk >=3 & vig10fwk<=28 then exer3 = 1;
run;

proc means;
        where bmi >0 and bmi<99;
        var bmi;
        class exer3;
run;
```

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

C) What percent of more frequent exercisers report excellent health? Exercise 3+ per week: 41.37% excellent health

Less frequent exercisers? Fewer than 3 per week: 34.19% excellent health.

```
proc freq;
        where health < 6;
        tables health*exer3;
run;
```

ANSWERS: Analyze the Sample – Part IV Graphing

Create a graph to show the mean BMI over age for males and females.

Section 1

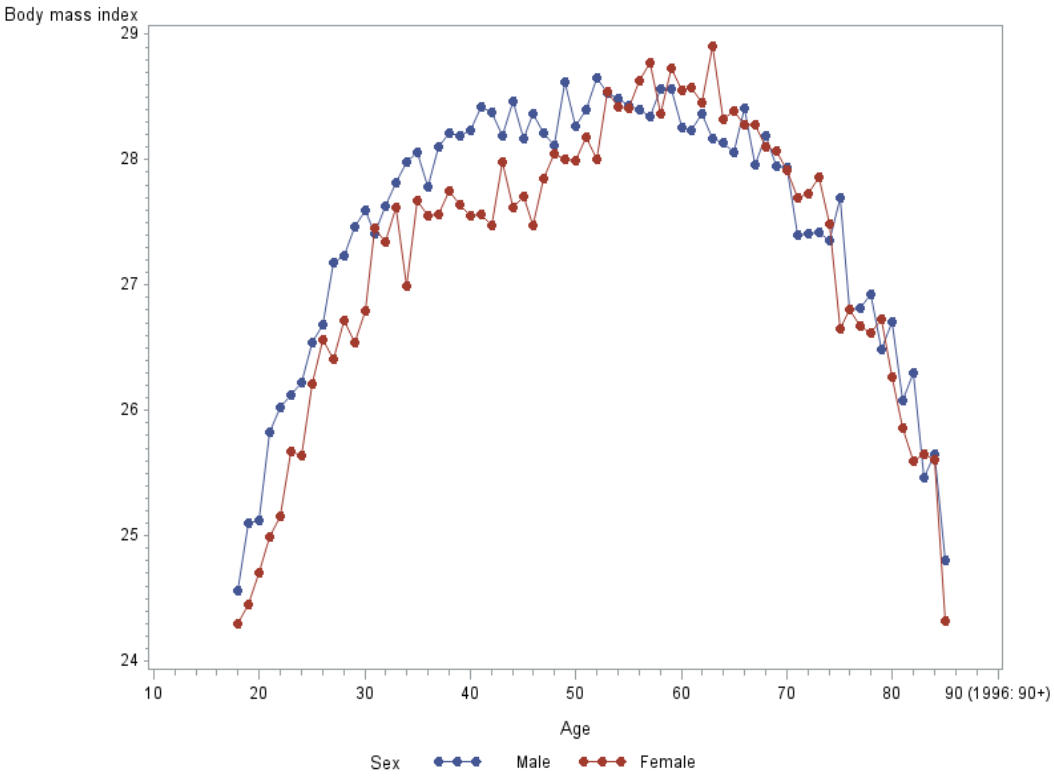
Analyze the Data

A) How does the universe for BMI appear on this graph? There appears to be very low or no BMI for individuals below 18, because the universe for BMI is only for adults older than 18.

B) Approximately at what age does BMI peak for women? For men? Women: ~ 61 years old; Men: ~50 years old

```
proc means;
where bmi > 0 and bmi < 99;
var bmi;
class sex age;
output out=sumstat mean=AverageBmi;
run;
symbol Interpol=join
            value=dot;
proc gplot data=sumstat;
where sex ne .;
plot AverageBmi*age=sex;
run;
```

Note: SAS graph procedures do not allow for WEIGHT options, so graph analyses are at the sample level.



C) Create a graph to show how an associated effect of poverty status on BMI differs with gender, controlling for frequent exercise. Comment on three apparent trends. Women under the poverty threshold are more likely to have a higher BMI on average whether or not they exercise. Frequent exercisers have lower BMI's on average in each category. Men under the poverty threshold seem to have a lower BMI on average controlling for exercise.

```
proc sgpanel data=ipums.ihis_000##;
        where bmi > 0 and bmi < 99 and pooryn^=9;
        panelby sex exer3;
        vbar pooryn/ stat = mean
        response=bmi;
run;
```

• • •

Section 2

Include a New Variable

Section 2

Graph the Data