



IPUMS Data Training Exercise:

An introduction to IPUMS NHIS

(Exercise 2 for SAS)



Learning goals

- Create and download an NHIS data extract
- Decompress data file and read data into SAS
- Analyze the data 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 associations among BMI, poverty, health status, sleep, and frequency of exercise. You will create a data extract that includes the variables AGE, SEX, POORYN, HEALTH, BMI, HRSLEEP, and VIG10FWK; then you will use the sample code to analyze these data.

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

Common Mistakes to Avoid

- Not fully decompressing the data
- Giving the wrong filepath to indicate the dataset
- Forget to close a procedure with "run;"
- Forget to terminate a command with a semicolon ";"

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 2004 through 2010 samples. 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 sleep under the "Health Behaviors" 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:
 - 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

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.



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://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 SAS 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

- Open the "nhis_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 "nhis_000##.dat";
- Choose Submit under the Run file menu.



Analyze the Sample

Part 1: Group Means

1. 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? _____

2. 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;
```

3. Is there a noticeable trend between health status and hours of sleep using this sample? _____
4. 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;
```



Part 2: Weighting the Data

To get a more accurate estimation of demographic patterns from the sample, you will have to use the person weight as well as the `proc surveyfreq` and `proc surveymeans` commands for appropriate variance estimation.

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

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

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

```
proc surveyfreq data=ipums.nhis_000##;  
    tables year*pooryn / row ;  
    strata strata;  
    cluster psu;  
    weight perweight;  
run;
```

- 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 surveyfreq data=nhis_000##;
  where pernum = 1;
  tables poorn*year;
  strata strata;
  cluster psu;
  weight hhweight;
run;
```

Part 3: 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.

8. Check the NHIS website 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.

9. 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 to exclude unknown and missing codes.



```

data ipums.nhis_000##;
    set ipums.nhis_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", i.e., replace the # symbols with the number of your data extract.

10. What percent of more frequent exercisers report excellent health? _____
11. What percent of less frequent exercisers report excellent health? _____

```

proc freq;
    where health<6;
    tables health*exer3 / col;
run;

```

Part 4: Graphing

Create a graph to show the mean BMI over age for males and females. Note: SAS graph procedures do not allow for WEIGHT options, so graph analyses are at the sample level.

12. How does the universe for BMI appear on this graph? _____



13. 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;
```

Introduce the variable POORYN

14. 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.nhis_000##;
  where bmi>0 and bmi<99 and pooryn^=9;
  panelby sex exer3;
  vbar pooryn/ stat = mean
  response=bmi;
run;
```



Answers

Part 1: Group Means

1. 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
2. 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
3. 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 health and people with excellent health are associated with getting more sleep.
4. 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.



Part 2: Weighting the Data

5. Without weights, what proportion of people in this sample was below the poverty threshold in 2010? 16.48% of the sample
6. Using weights, what proportion of the population was below the poverty threshold in 2010? 13.76% of the sample
7. 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

Part 3: Generating Variables

8. Check the output of the do file in the Log window to find the codes for VIG10FWK. Which code means "Never"? 95 Never
9. 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)
10. What percent of more frequent exercisers report excellent health? 41.37%
11. What percent of less frequent exercisers report excellent health? 34.19%

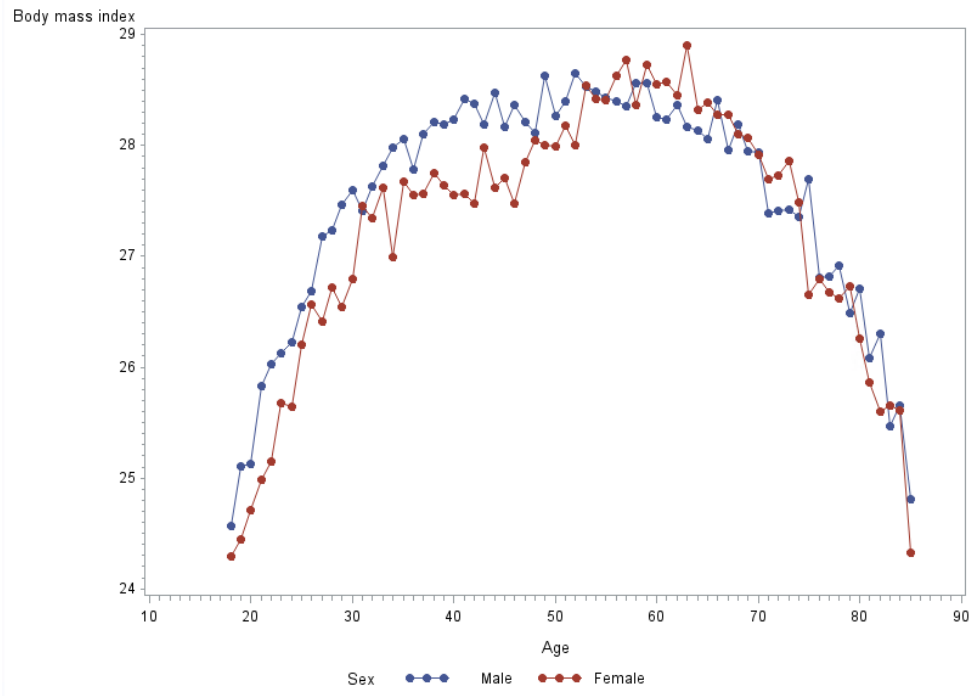
Part 4: Graphing

12. How does the universe for BMI appear on this graph? There appears to be no BMI for individuals below 18, because the universe for BMI is only for adults older than 18.
13. Approximately at what age does BMI peak



for women? ~ 61 years old

For men? ~50 years old



Introduce the variable POORYN

14. 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.



