IPUMS Data Training Exercise:

An introduction to IPUMS NHIS

(Exercise 1 for SDA)

IPUMS HEALTH SURVEYS

Learning goals

- Understand how to navigate the IPUMS NHIS website and browse variables
- Select datasets and variables of interest from IPUMS NHIS in SDA
- Analyze smoking patterns and lack of assistive devices among the elderly population in the United States using sample code for SDA

Summary

In this exercise, you will gain basic familiarity exploring IPUMS NHIS data using our website, and will utilize the online analysis tool, SDA, to answer the following research questions: What are the patterns of smoking behaviors in the United States? How many elderly people lack an assistive device that need it? You will use sample code and the IPUMS NHIS variables SMOKESTATUS1, SMOKESTATUS2, CSQTRYYR, USEDEVICE, and DIFWODEVICE to answer these questions. After completing this exercise, you will have experience navigating the IPUMS NHIS website and should be able to leverage these data to explore your own research interests.

Register for an IPUMS account

Go to https://nhis.ipums.org, click on User Registration and Login and Apply for access. On the login screen, enter email address and password and submit your application.

Exploring the IPUMS NHIS website

- Navigate to https://nhis.ipums.org and click on "Browse and Select Data" on the left-hand menu underneath the Data header.
- The variable drop-down menus allow you to explore variables by topic. For example, you might expect to find variables about smoking under the Health Behaviors > Smoking 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 view more 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.
- Use either the drop down menu or the search feature to browse these variables.
    - SMOKESTATUS1: Smoking status current/former/never
    - SMOKESTATUS2: Smoking status detailed current/former/never
    - CSQTRYYR: Tried to quit smoking 1+ days, past 12 months
    - CIGSDAY: Number cigarettes per day (current smokers)
    - USEDEVICE: Uses assistive device
    - DIFWODEVICE: Frequency of difficulty, lacking assistive device

Getting started with SDA

Selecting datasets

- From the IPUMS NHIS home page, click on "Analyze Data Online" on the left-hand menu underneath the Data header.
- The next page summarizes information about SDA and allows you to select either a single year dataset, or a multi-year dataset.
- Select the multi-year dataset that covers the 1997-present samples.
- Note that you will need to log in to use SDA.

Browsing and adding variables

- You may browse variables under the Household and Person variable categories on the left-hand menu of SDA, or search for them on the IPUMS NHIS website and enter the variable names directly as inputs in SDA.
    - Note that some variables available through the IPUMS NHIS extract system are not available for analysis in SDA.
- When you browse variables within SDA, click on the plus symbol to the left of a variable topical group to see the available variables.
    - Click on a variable name, and it will appear in the "Selected" box at the top of the left-hand SDA variable browsing area.
    - To insert the selected variable as inputs for your SDA analysis, click on the appropriate input location (e.g., row, column).

Relevant Fields in online tabulator

- Row: Represents the primary variable of interest
- Column: Divides the analysis of the variable of interest into categories
- Control: Creates a separate chart for each category of the control

- Selection filter: Allows you to select cases; ex: year(2000-*) -> all years 2000-onward

Common mistakes to avoid

- Choosing a numerical rather than categorical variable for the Frequencies/Cross Tabulation Program. For continuous variables, use the Comparison of Means Program instead.
- Forgetting to exclude missing data categories.
- Forgetting to specify the years of interest.
- Failing to update the weight. The default value for weight is the person weight (PERWEIGHT), which is not appropriate for analyses of sample adults or sample children or variables requiring the use of a supplement weight.
- Not intentionally choosing column or row percentages (column is the default).

Analyze the Data

Part 1: Frequencies

1. What is the difference between SMOKESTATUS1 and SMOKESTATUS2?

   _____

   _____

   _____

   _____

   _____

2. What are the codes for SMOKESTATUS1? _____

   _____

   _____

   _____

3. What is the universe for SMOKESTATUS1 in 1997-forward? _____

   _____

4. What is the universe for CSQTRYYR in 1997-forward? _____

   _____

   _____

   _____

5. How could you address these universe changes for an analysis using SMOKESTATUS1 and CSQTRYYR? _____

   _____

   _____

6. For your revised sample, what percentage of current smokers tried to quit...
    a. in the last 12 months? _____
    b. in the year 2000? _____

```
ROW: csqtryyr
WEIGHT: sampweight
SELECTION FILTER(S) (2a): smokestatus1(20), year(1997-2003)
SELECTION FILTER(S) (2b): smokestatus1(20), year(2000)
```

7. What are the missing data codes for AGE and USEDEVICE?

    _____

    _____

8. What percentage of people over the age of 65 used an assistive device in 2002?

    _____

```
ROW: usedevice
WEIGHT: sampweight
SELECTION FILTER(S): year(2002), age(65-85), usedevice(1-2)
```

Part 2: Relationships in the data

In the upper-left corner of the SDA interface, hover over "Analysis" and then select "Comparison of Means".

9. What are the missing data codes for CIGSDAY? _____

    _____

10. In 2009, what is the average number of cigarettes smoked each day for…
    a. daily smokers? _____
    b. some-day smokers? _____

```
DEPENDENT: cigsday

ROW: smokestatus2

WEIGHT: sampweight

SELECTION FILTER(S): year(2009), cigsday(*-90)
```

Navigate back to the Frequencies and Cross Tabulations via the "Analysis" option.

11. Among persons who reported their frequency of difficulty without a device, what number and percentage of people over the age of 65 always had difficulty without a device? _____

```
ROW: difwodevice

WEIGHT: sampweight

SELECTION FILTER(S): year(2002), age(65-85), difwodevice(1-5)
```

To complete the next section, you can use the recode feature in SDA. Remember to consider if you want column or row percentages for your answer.

12. Among those who "always" or "often" had any difficulties, what percentage were women? _____

```
ROW: difwodevice(r:1-2 "Always or Often"; 3-5 "Sometimes,
Rarely, Never")

COLUMN: sex

WEIGHT: sampweight

SELECTION FILTER(S): year(2002), age(65-85), difwodevice(1-5)
```

Answers

Part 1: Frequencies

1. What is the difference between SMOKESTATUS1 and SMOKESTATUS2?

   Years of availability (SMOKESTATUS1 available through 2003 and in assorted earlier years; SMOKESTATUS2 available through present and assorted earlier years). These changes in availability reflect a comparability break between the two variables: SMOKESTATUS2 has less detail in the frequency of smoking among former smokers, but more detail about current smokers daily cigarette usage.

2. What are the codes for SMOKESTATUS1?

   00: NIU, 10: never smoked; 20: current smoker, 30: former smoker; 31: former regular smoker; 32: former occasional smoker; 90: unknown smoking status; 91: unknown if ever smoked 100+ cigarettes; 92: unknown if smokes currently but has smoked 100+ cigarettes

3. What is the universe for SMOKESTATUS1 in 1997-forward?

   1997-2003: Sample adults age 18+

4. What is the universe for CSQTRYYR in 1997-forward?

   1997: Sample adults age 18+ who have ever smoked 100 cigarettes and currently smoke every day or some days; 1998-2003: Sample adults age 18+ who have ever smoked 100 cigarettes and currently smoke every day or some days or whose current smoking status is unknown.

5. How could you address these universe changes for an analysis using SMOKESTATUS1 and CSQTRYYR?

   Restrict to only persons with a known smoking status and use 1997-2003, or restrict to 1998-2003 to avoid universe difference for persons with unknown current smoking status between 1997 and 1998.

6. For your revised sample, what percentage of current smokers tried to quit...
    a. in the last 12 months? 43.0%
    b. in the year 2000? 43.4%

7. What are the missing data codes for AGE and USEDEVICE?

   AGE: no missing data codes, but top-coded at 85; USEDEVICE: 0 is NIU, 7 is refused, 8 is not applicable, and 9 is don't know.

8. What percentage of people over the age of 65 used an assistive device in 2002?

   18.5%

Part 2: Relationships in the data

9. What are the missing data codes for CIGSDAY?

   Top-coded at 90, and 96+ are NIU and other missing data codes.

10. In 2009, what is the average number of cigarettes smoked each day...
    a. by daily smokers? 15.46
    b. by some-day smokers? 4.56

11. Among persons who reported their frequency of difficulty without a device, what number and percentage of people over the age of 65 always had difficulty without a device?

    29.8%; unweighted N = 81; weighted N = 392,904

12. Among those who "always" or "often" had any difficulties, what percentage were women? 61.1%