



IPUMS Data Training Exercise:

An introduction to IPUMS HigherEd

(Exercise 2 for Stata)



Learning goals

- Understand how IPUMS HigherEd dataset is structured
- Create and download an IPUMS data extract
- Decompress the data file and read the data into a statistical package
- Explore the factors that affect student loan debt and the relatedness between highest degree and work activities

Summary

In this exercise, you will gain basic familiarity with the IPUMS HigherEd data exploration and extract system to answer the following research questions: "What are the most common work activities for scientists and engineers in the US? What factors are most important when predicting remaining student loan debt for college graduates?" You will create a data extract that includes the variables GENDER, WAPRI, AGE, MINRTY, GOVSUP, DGRDG, BAFLIN; then you will use the sample code to analyze these data. After completing this exercise, you will have experience

navigating the IPUMS HigherEd website and should be able to leverage these data to explore your own research interests.

Stata Code to Review

Code	Purpose
<code>generate</code>	Creates a new variable, "replace" specifies a value according to cases
<code>mean</code>	Displays a simple tabulation and frequency of one variable
<code>tabulate</code>	Displays a cross-tabulation for up to 2 variables
<code>regress</code>	OLS regression

Common Mistakes to Avoid

- Not fully decompressing the data
- Giving the wrong filepath to indicate the dataset
- Mixing up `=` and `==` ; To assign a value in generating a variable, use `=`. Use `==` to specify a case when a variable is a desired value using an *if* statement.
- Forgetting to put `[weight=weightvar]` into square brackets

Register for an IPUMS Account

Go to <http://higher.ed.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. Registration is automatic.

Make a Data Extract

- Navigate to the IPUMS HigherEd homepage and click on "Browse Data."

Select Samples

- Click on the "Select Samples" button to choose the census samples to include in your extract
- Click the Select Samples box on the SESTAT tab
- Check the box labeled "Select All Samples" – this will select all years of SESTAT samples. Click on Submit sample selections

Select Variables

- The variable drop-down menus allow you to explore variables by topic. . For example, you might expect to find variables about a person's degree(s) under the "Education" 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 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, and add them to your data cart.
 - GENDER: Gender
 - MINRTY: Minority indicator
 - WAPRI: Primary work activity
 - AGE: Age
 - GOVSUP: Indicator of government support for work
 - DGRDG: Degree type of respondent's highest degree
 - BAFLN: Loans as financial support for bachelor's degree
 - WEIGHT: SESTAT sample weight variable (will appear as part of the SESTAT sample)

Review and submit your extract

- Click on the "View Cart" button underneath your data cart.
- Review your variable and sample selection to ensure your extract is complete.
 - You may notice a number of additional variables you did not select are in your cart; IPUMS preselects a number of key technical variables, which are automatically included in your data extract.
- Add additional variables or samples if they are missing from your extract, or click the "Create Data Extract" button.



- Review the Extract Request screen that summarizes your extract; add a description of your extract (e.g., "Loan Debt, Highest Degree, and Work Activities") and click "Submit Extract".
- You will receive an email when your data extract is available to download.

Getting the data into your statistics software

The following instructions are for Stata. If you would like instructions for a different stats package, see <https://ipums.org/support/exercises>.

Download the data

- Follow the link in the email notifying you that your extract is ready, or by clicking on the "Download and Revise Extracts" link on the left-hand side of the IPUMS HigherEd homepage.
- Right-click on the data link next to the extract you created.
- Choose "Save Target As..." or "Save Link As..."
- Save into your preferred working directory. This tutorial assumes you will save the file into "Documents" (which should pop up as the default location).
- Do the same thing to save the Stata command file.

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.
- Double click on the ".dat" file.
- In the window that pops up, press the "Extract" button.
- After the extract has completed, confirm that the Documents folder contains three files that begin with "higher_000###".

Read in the data

- Open the "higher_000###.do" file.
- In the syntax editing 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 higher_###.dat;
- Choose "Submit" under the Run file menu.



Analyze the Data

Part 1: Frequencies

For each variable below, search through the tabbed sections of the variable description to answer each question.

1. On the website find the codes page for the DGRDG variable and write down each code value, and what category each code represents.

2. Describe the kind of information WAPRI provides.

3. What were the top three most common work activities for SESTAT respondents in 2010? _____

```
tab wapri if year == 2010 & wapri < 90, sort
```

4. What were the top three most common work activities for SESTAT respondents in 2010 whose highest degree was a doctorate?

```
tab wapri if year == 2010 & wapri < 90 & dgrdg == 3, sort
```

Using Weights (WEIGHT)

In order to find a nationally representative estimate of doctorate recipients, we need to use a frequency weight. The above analysis finds frequencies specific to the survey samples, but a weight adjusts the analysis to be representative of the US target population for each year. The variable WEIGHT is specifically adjusted to analyze the entire SESTAT sample in a particular year.

5. How many doctorate recipients nationally had an occupation in which the primary work activity was teaching?

6. What proportion of doctorate recipients nationally had an occupation in which the primary work activity was basic research?

```
replace weight = round(weight)
```



```
tab wapri if year == 2010 & wapri < 90 & dgrdg == 3 [w=weight],  
sort
```

Part 2: Relationships in the Data

7. What is the universe for GOVSUP?

8. What percentage of the science and engineering workforce in the United States received funding from the federal government for their work? Note that we are excluding respondents who were out of universe for GOVSUP?

```
tab year govsup if govsup < 90 [w=weight], row
```

9. Do doctorate or bachelor's degree holders tend to receive federal funding for their work?

```
tab dgrdg govsup if govsup < 90 & year == 2010 [w=weight], row
```

10. How many individuals in the science and engineering workforce in 1999 used loans from lending institutions to finance their bachelor's degree?

```
tab year bafln if bafln < 90 [w=weight], row
```



Answers

Part 1: Frequencies

For each variable below, search through the tabbed sections of the variable description to answer each question.

1. On the website find the codes page for the DGRDG variable and write down each code value, and what category each code represents. 1 = Bachelor's; 2 = Master's; 3 = Doctorate; 4 = Professional; 5 = Other
2. Describe the kind of information WAPRI provides. The work activity the respondent performs for the majority of his/her job.
3. What were the top three most common work activities for SESTAT respondents in 2010? Management and Administration, Professional Services, Teaching
4. What were the top three most common work activities for SESTAT respondents in 2010 whose highest degree was a doctorate? Teaching, Basic research, Management and Administration

Using Weights (WEIGHT)

5. In 2010, how many doctorate recipients nationally had an occupation in which the primary work activity was teaching 234,594 doctorate recipients
6. What proportion of doctorate recipients nationally had an occupation in which the primary work activity was basic research in 2010? 17.7%

Part 2: Relationships in the Data

7. What is the universe for GOVSUP? Worked during calendar year prior to the survey



8. In 2010, what percentage of the science and engineering workforce in the United States received funding from the federal government for their work? 16.15%
9. Do doctorate or bachelor's degree holders tend to receive federal funding for their work? Doctorates; 30% of Doctorate degree holders vs Bachelors, of which only 14.1% receive funding
10. How many individuals in the science and engineering workforce in 1999 used loans from lending institutions to finance their bachelor's degree? 48.95%

