

IPUMS – HigherEd Extraction and Analysis

Exercise 2 - SPSS

OBJECTIVE: Gain an understanding of how an IPUMS dataset is structured and how it can be leveraged to explore your research interests. This exercise will use the IPUMS dataset to explore the factors that affect student loan debt and the relatedness between highest degree and work activities.

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?

Objectives

- Explore a sample of variables from IPUMS-HIGHER ED
- Analyze the data using example code

IPUMS Variables

- GENDER: Respondent's gender
- WAPRI: Primary work activity
- AGEP: Age
- MINRTY: Minority indicator
- GOVSUP: Indicator of government support for work
- DGRDG: Degree type of respondent's highest degree
- BAFLN: Financial support during bachelor's degree included loans
- UGBALPB: Balance left on undergraduate student loans
- UGLOANPB: Total amount of undergraduate student loans

SPSS Code to Review

Code	Purpose
compute	Creates a new variable
freq	Displays a simple tabulation and frequency of one variable
crosstabs	Displays a cross-tabulation for up to 2 variables and a control
~=	Not equal to

Review Answer Key (page 6)

Common Mistakes to Avoid

1 Excluding cases you don't mean to. Avoid this by turning off weights and select cases after use, otherwise they will apply to all subsequent analyses

2 Terminating commands prematurely or forgetting to end commands with a period (.) Avoid this by carefully noting the use of periods in this exercise

Registering with IPUMS

Go to <http://highered.ipums.org>, click on "Register to Use Data System" and apply for access. On login screen, enter email address and password and submit it!

Step 1

Make an Extract

- Go back to homepage and go to Select Data
- 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
- Using the drop down menu or search feature, select the following variables:

GENDER: Gender

MINRTY: Minority indicator

WAPRI: Primary work activity

AGEP: 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

UGBALPB: Balance of undergraduate student loans

WEIGHT: SESTAT sample weight variable

UGLOANPB: Total amount of undergraduate student loans

...

Step 2

Request the Data

- Click the green VIEW CART button under your data cart
- Review variable selection. Click the green Create Data Extract button
- Review the 'Extract Request Summary' screen, describe your extract and click Submit Extract
- You will get an email when the data is available to download
- To get to the page to download the data, follow the link in the email, or follow the Download and Revise Extracts link on the homepage

Getting the data into your statistics software

The following instructions are for SPSS. If you would like to use a different stats package, see: http://highered.ipums.org/highered/extract_instructions.shtml

Step 1

Download the Data

- Go to <http://highered.ipums.org> and click on Download or Revise Extracts
- Right-click on the data link next to extract you created
- Choose "Save Target As..." (or "Save Link As...")
- Save into "Documents" (that should pop up as the default location)
- Do the same thing for the SPSS 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 here
- Double-check that the Documents folder contains three files starting "highered_000..."
- Free decompression software is available at <http://www.ironis.net/soft/wingzip/>

•••

Step 3

Read in the Data

- Double click on the ".sps" file, which should automatically have been named "highered_000....."
- The first two lines should read:

```
cd "."  
data list file = 'highered_000.../'
```
- Change the first line to read: cd (location where you've been saving your files). For example:

```
cd "C:\Documents".
```
- Change the second line to read:

```
data list file = "C:\Documents\highered_000...dat"/
```
- Under the "Run" menu, select "All" and an output viewer window will open
- Use the Syntax Editor for the SPSS code below, highlight the code, and choose "Selection" under the Run menu

Analyze the Sample – Part I Frequencies

Step 1

Analyze the Data

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

B) Describe the kind of information WAPRI provides.

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

```
compute filter1 = (WAPRI<90).  
filter by filter1.  
crosstabs /tables = year by wapri  
/cells= count row.
```

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

```
filter off.  
compute filter2 = (DGRDG=3 & WAPRI<90).  
filter by filter2.  
crosstabs /tables = year by wapri  
/cells= count row.
```

Using weights (WEIGHT)

Step 2

Analyze the Data

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.

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

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

...

Step 3

Weighting the Data

```
filter by filter2.  
weight by weight.  
crosstabs /tables = year by wapri  
/cells= count row.
```

Analyze the Sample – Part II Relationships in the Data

Section 1

Crosstabs

A) What is the universe for GOVSUP?

B) What percentage of the science and engineering workforce in the United States received funding from the federal government for their work in 2010? Note that we are excluding respondents who were out of universe for GOVSUP. _____

```
filter off.  
compute filter3 = (GOVSUP<90 & YEAR==2010).  
filter by filter3.  
crosstabs /tables year by govsup /cells = count row.
```

C) Do doctorate or bachelor's degree holders tend to receive federal funding for their work at a higher rate? _____

```
filter by filter3.  
crosstabs /tables dgrdg by govsup /cells = count row.
```

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

```
filter off.  
compute filter4 = (BAFLN<90).  
filter by filter4.  
crosstabs /tables year by bafln /cells = count row.
```

Analyze the Sample – Part II Relationships in the Data

Section 2

Means

A) What is the average amount of remaining undergraduate loans for those who used loans to finance their undergraduate degree in 1999? _____

```
filter off.  
compute filter5 = (UGBALPB<999990 & YEAR==1999).  
filter by filter5.  
means tables ugbalpb.
```

B) What is the average amount of total undergraduate loans borrowed for those who used loans to finance their undergraduate degree in 1999? _____

```
filter off.  
compute filter6 = (UGLOANPB<999990 & UGLOANPB>0 &  
YEAR=1999).  
filter by filter6.  
means tables ugloanpb.
```

Note: We exclude zeros from original loans, but not from the remaining balance. We want to include the people who paid off the loans, not the ones who never took out loans.

C) Now let's use some common variables to predict the value of the remaining undergraduate student loans of the science and engineering workforce in 1999. _____

...

Complete!
Check
your
Answers!

```
filter off.  
filter by filter5.  
weight by wtsurvey.  
regression  
/statistics coeff r  
/dependent ugbalpb  
/method=enter agep gender minrty.
```

ANSWERS - Analyze the Sample – Part I Frequencies

Step 1

Analyze the Data

A) 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

B) Describe the kind of information WAPRI provides. The work activity the respondent performs for the majority of his/her job.

C) What were the top three most common work activities for SESTAT respondents in 2010? Management and Administration, Professional Services, Teaching

```
compute filter1 = (WAPRI<90).  
filter by filter1.  
crosstabs /tables = year by wapri  
/cells= count row.
```

D) 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

```
compute filter2 = (DGRDG=3 & WAPRI<90).  
filter by filter2.  
crosstabs /tables = year by wapri  
/cells= count row.
```

ANSWERS - Using weights (WEIGHT)

Step 2

Analyze the Data

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.

A) How many doctorate recipients nationally had an occupation in which the primary work activity was teaching? 234,630

B) What proportion of doctorate recipients nationally had an occupation in which the primary work activity was basic research? 17.7%

...

Step 3

Weighting the Data

```
filter by filter2.  
weight by weight.  
crosstabs /tables = year by wapri  
/cells= count row.
```

ANSWERS - Analyze the Sample – Part II Relationships in the Data

Section 1

Crosstabs

A) What is the universe for GOVSUP? Worked during calendar year 2009

B) What percentage of the science and engineering workforce in the United States received funding from the federal government for their work in 2010? Note that we are excluding respondents who were out of universe for GOVSUP. 16%

filter off.

compute filter3 = (GOVSUP<90 & YEAR=2010).

filter by filter3.

crosstabs /tables year by govsup /cells = count row.

C) Do doctorate or bachelor's degree holders tend to receive federal funding for their work at a higher rate? Doctorates 30% (vs Bachelors, of which only 14% receive funding)

filter by filter3.

crosstabs /tables dgrdg by govsup /cells = count row.

D) How many individuals in the science and engineering workforce in 1999 used loans from lending institutions to finance their bachelor's degree? 49%

filter off.

compute filter4 = (BAFLN<90).

filter by filter4.

crosstabs /tables year by bafln /cells = count row.

ANSWERS - Analyze the Sample – Part II Relationships in the Data

Section 2

Means

A) What is the average amount of remaining undergraduate loans for those who used loans to finance their undergraduate degree in 1999? \$7089.37

```
filter off.  
compute filter5 = (UGBALPB<999990 & YEAR=1999).  
filter by filter5.  
means tables ugbalpb.
```

B) What is the average amount of total undergraduate loans borrowed for those who used loans to finance their undergraduate degree in 1999? \$8948.41

```
filter off.  
compute filter6 = (UGLOANPB<999990 & UGLOANPB>0 &  
YEAR=1999).  
filter by filter6.  
means tables ugloanpb.
```

Note: We exclude zeros from original loans, but not from the remaining balance. We want to include the people who paid off the loans, not the ones who never took out loans.

C) Now let's use some common variables to predict the value of the remaining undergraduate student loans of the science and engineering workforce in 1999. (Coefficients below)

...

Complete!
Check
your
Answers!

Variable	Coef	t-stat
Age	-134	-127
Gender	-568	-52
Minority	564	39
Constant	12008	319