Minnesota Population Center
Training and Development

IPUMS – HigherEd Extraction and Analysis

Exercise 1 - 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 doctorate recipient's salaries and the relatedness between doctorate recipients' field of degree, employer sector, and gender.

6/2/2015

IPUMS – Higher Ed Training and Development

Research Questions

How many doctorate recipients are working in an occupation related to his/her highest degree? What factors are most important in determining a doctorate recipient's salary?

Objectives

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

IPUMS Variables

- GENDER: Respondent's gender
- SALARP: Annual salary
- AGEP: Age
- EMSECPB: Employer sector
- NDGMEPD: Field of degree category
- CTZUSIN: US citizenship
- OCEDRLP: Degree to which respondent's job related to highest degree

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 7)

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 IPUMS-HIGHER ED" and apply for access. On the login screen, enter email address and password and submit it!

Step 1

Make an Extract

- Go back to the homepage and go to Select Data
- Click the Select Samples box and go to the Full SDR tab
- Check the very first check box labeled SDR – this will select all years of full SDR samples. Click on Submit sample selections
- Using the drop down menu or search feature, select the following variables:

  GENDER: Gender

  AGEP: Age

  MINRTY: Minority background indicator

  SALARP: Annual salary

  LFSTAT: Employment status

  EMSECPB: Employer sector

  HRSWKP: Hours typically worked per week

  CTZUSIN: US citizenship indicator

  OCEDRLP: Degree to which respondent's job related to highest degree

  NDGMEDP: Field of major for highest degree

  NDGMEMG: Field of major for highest degree (6 groups)

  WTSURVY: Full SDR weight variable

  SUPWK: Work includes supervisory role

- 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

•••

Step 2

Request the Data

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.irnis.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 OCEDRLP variable and write down each code value, and what category each code represents. _____

B) What is the universe for OCEDRLP?

_____

C) How many doctorate recipients were employed in an occupation closely related to their field in 2013 in the SDR survey? _____

D) What proportion of doctorate recipients in the SDR survey were employed in an occupation closely related to their field in 2013?

```
compute filter1 = (OCEDRLP~=98).
filter by filter1.
crosstabs /tables = year by ocedrlp
/cells= count row.
```

• • •

Step 2

Weighting the Data

Using weights (WTSURVY)

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.

A) How many doctorate recipients nationally had an occupation closely related to their field of degree in 2013?

_____

B) What proportion of doctorate recipients nationally had an occupation closely related to their field of degree in 2013?

_____

```
filter by filter1.
weight by wtsurvy.
crosstabs /tables = year by ocedrlp
/cells= count row.
```

C) How many doctorate recipients were working in the United States in 2013? _____

Analyze the Sample – Part II Relationships in the Data

Section 1

Crosstabs and Means

A) Which doctorate fields were dominated by women in 2013 (women comprised more than 50 percent)?

_____

```
filter off.
compute filter2=(SALARP <=150000 & YEAR=2013).
filter by filter2.
crosstabs  /tables = ndgmedp by gender /cells= count row.
```

B) What is the difference in the mean salary between employed female and male doctorate recipients in 2013? _____

```
filter by filter2.
weight by wtsurvy.
means tables salarp by gender.
```

C) What are the average salaries for doctorate recipients by employer sector in 2013? _____

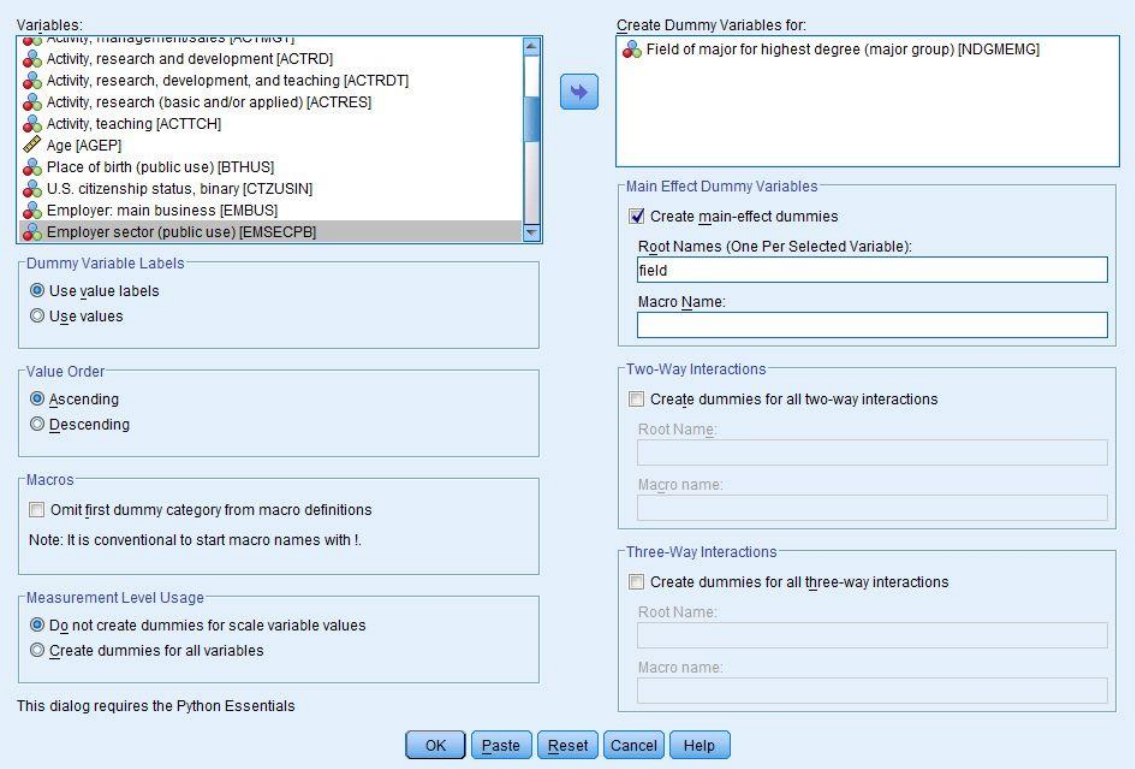
```
filter by filter2.
weight by wtsurvy.
means tables salarp by emsecpb.
```

Section 1

Regression

D) Use OLS regression to predict reported salaries of doctorate recipients for 2013. _____

Go to Transform-> Create Dummy variables and generate dummy variables for NDGMEMG (field) and EMSECPB (sector), as shown above for sector.



Note: Because SDR respondents are interviewed every 2 to 3 years, regressing over multiple years will lead to over-counting individuals and standard errors that are too small.

```
filter by filter2.
weight by wtsurvy.
regression
/statistics coeff r
/dependent salarp
/method=enter gender agep ctzusin minrty supwk sector_2 sector_3
sector_4 field_1 field_2 field_3 field_4 field_5 field_6
```

• • •

Complete! Check your Answers!

ANSWERS - Analyze the Sample – Part I Frequencies

Step 1

Analyze the Data

A) Find the codes page for the OCEDRLP variable and write down each code value, and what category each code represents. 1 Closely related; 2 Somewhat related; 3 Not related; 98 Logical Skip

B) What is the universe for OCEDRLP? Working during the week of *sample reference period.*

C) How many doctorate recipients were employed in an occupation closely related to their field in 2013 in the SDR survey? 17,696

D) What proportion of doctorate recipients in the SDR survey were employed in an occupation closely related to their field in 2013?

66.33%

```
compute filter1 = (OCEDRLP~=98).
filter by filter1.
crosstabs /tables = year by ocedrlp
/cells= count row.
```

• • •

Step 2

Weighting the Data

Using weights (WTSURVY)

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.

A) How many doctorate recipients nationally had an occupation closely related to their field of degree in 2013?

474,868

B) What proportion of doctorate recipients nationally had an occupation closely related to their field of degree in 2013?

65.9%

C) How many doctorate recipients were working in the United States in 2013? 720,802

```
weight by wtsurvy.
crosstabs
/tables = year by ocedrlp
/cells= count row.
```

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

Section 1

Crosstabs and Means

A) Which doctorate fields were dominated by women in 2013?
_Psychology, Sociology/Anthropology, Health, Non-science

```
filter off.
compute filter2=(SALARP <=150000 & YEAR=2013).
filter by filter2.
crosstabs  /tables = ndgmedp by gender /cells= count row.
```

B) What is the difference in the mean salary between employed female and male doctorate recipients in 2013?

__Women:_$84,409.26;_Men:$102,628.63;_Difference:_$18,228.37_

```
filter by filter2.
weight by wtsurvy.
means tables salarp by gender.
```

C) What are the average salaries for doctorate recipients by employer sector in 2013? _____

```
filter by filter2.
weight by wtsurvy.
means tables salarp by emsecpb.
```

Sector	Mean Salary ($)
2 Year College	63,194.48
4 Year College	86,864.94
Government	104,107.12
Business/industry	106,827.68

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

Section 1

D) Use OLS regression to predict reported salaries of doctorate recipients for 2013. _____

Note: Because SDR respondents are interviewed every 2 to 3 years, regressing over multiple years will lead to over-counting individuals and standard errors that are too small.

Regression

```
filter by filter2.
weight by wtsurvy.
regression
/statistics coeff r
/dependent salarp
/method=enter gender agep ctzusin minrty supwk sector_2
sector_3 sector_4 field_1 field_2 field_3 field_4 field_5 field_6
```

Variable	Coefficient	t-statistic
Gender	10024	103.4
Age	332	83.86
US citizen	6991	52.5
Minority	-4823	-31.48
Supervisory Work	23168	269.88
4 year college	17116	73.62
Government	34327	130.53
Business and Industry	34622	148.71
Computer and Math Sciences	24736	11.27
Biological Sciences	11862	5.42
Physical Sciences	15484	7.07
Social Sciences	8002	3.65
Engineering	25344	11.56
S&E related fields	17188	7.82
Constant	5172	2.34