

Female Labour Supply coursework August 2020

We will use the Current Population Survey (CPS). This is a monthly survey in the United States, which significantly informs public policy decisions. For example, monthly unemployment statistics are derived from this survey. We will use three years (1989-1991) of the March surveys, because those contain an extended set of variables. If you want to know more about the CPS (and you probably will need to look at the documentation) here are two useful links:

http://nber.org/data/cps_basic.html

<http://www.nber.org/data/current-population-survey-data.html>

Fortunately for you, I have already downloaded the data and dropped many (but not all) of the useless variables. Here is what you should do:

1. First of all, open your do-file, clear Stata's memory, and include the command line "set more off". Append the three Stata files (stack them on top of each other) that I have put on Moodle, but make sure that you mark the year, which the observations are from. To make sure you have done this correctly, show how many observations you have for each year (there should be 144,687 in 1989). **(5 marks)**
2. Drop individuals that are in the armed forces. How many observations were dropped? **(6 marks)**
3. Create a dummy called "self_emp" for those whose main income is from any form of self-employment or who work without pay. How many self-employed observations are there? **(10 marks)**
4. Create a variable indicating the age of the individual. Name this variable 'age1'. Drop anyone below the age of 16. What is the average age of men in the sample? What is the average age of women in the sample? **(10 marks)**
5. Create 'age2' by squaring 'age1'. **(7 marks)**
6. Drop anyone that is self-employed. How many observations did you drop? **(5 marks)**
7. Create a variable for annual hours worked (in the previous year). For some observations, this variable will be based on *allocated* values. That is, they were not directly answered by the respondent, but inferred by the statisticians from previous records. Drop observations whose annual hours are allocated. What is the average of hours worked in the remaining sample? **(10 marks)**

8. We will now find the hourly wage. The CPS records annual wage and salary income from the main job, and other wage and salary income. Both of these variables are top-coded at \$99,999. Multiply top-coded values by 1.45, then add the two incomes, and find the hourly wage. To make the wages comparable across the three years, convert the hourly wage to year 2012-dollar values by inflating them with the personal consumption expenditure price index (implicit deflator). You can find this index here <http://research.stlouisfed.org/fred2/categories/21>. What is the average hourly wage in 1989? **(10 marks)**
9. Create the log of hourly wage. What is the mean log wage for females? **(7 marks)**
10. We will now create a years of schooling variable. Assign the following years of schooling to the observations

| Schooling category | Assigned years of schooling |
|-----------------------------------|-----------------------------|
| None or did not finish elementary | 0 |
| Elementary 1-4 | 4 |
| Elementary 5-7 | 7 |
| Elementary 8 | 8 |
| High school 1-3 | 11 |
| High school 4 | 12 |
| College 1-3 | 14 |
| College 4 | 16 |
| College 5 | 17 |
| College 6+ | 20 |

What is the average education in the sample? **(10 marks)**

11. Estimate a Mincer regression for women, that is, regress log of wages on years of schooling, age and age squared. Give an interpretation of your results. Why did we include 'age2' in the regression? **(20 marks)**