Economics of Discrimination
Professor Patrick Button
Total points possible = 10

Group Current Population Survey Stata Data Assignment #1

Introduction

In this assignment you will work with Current Population Survey (CPS) data from the IPUMS-CPS. The Current Population Survey is commonly used by labor economists to investigate labor market trends. For example, it is used to calculate official unemployment rates. As a reminder, you can do this assignment solo or with a partner (and then hand in one copy with both of your names on it).

Downloading the Data

You can download the data for this assignment from: https://cps.ipums.org/cps/

During one or more lectures I will walk you through how to download and process the data. You will need to download the data and process it yourself. This is especially important since you will need to learn how to download and process the data yourself as you will be doing your own analysis using this data for your brief assignment later.

Go to https://cps.ipums.org/cps/ to download the data. Follow the instructions above and ensure you have the right variables and the right time periods. To ensure you have everything, you should read the entire assignment to confirm which variables and time periods you are asked for. In class when I show you how to download the data I will show you how to find what you need, so you could follow along with me.

You will be prompted to make an account to download any data you've added to your "cart". You'll get an email with a link to the data when it's ready (took less than a minute for me to get the data after I submitted the request, but it may take up to an hour if you're requesting a lot of data).

When you click the link in your email to download the data, you'll see several columns with options. You want to download it in Stata format, which provides a Stata-ready ".dta" file.

The One Question for this Assignment

1. (8 points) Calculate the employment rate for the nation for August 2021 (the most recently available month as of October 5, 2021). Please use survey weights and please ensure you are using the correct weight (in this case, it's the weight for person-level analysis of basic monthly data). Report the estimated employment rate, its standard error, and its 95% confidence interval.

Tips: I will demo this in class. First you need to create an indicator variable for being employed. I will show you how to do this during class. You will use the employment status variable (empstat) for this. Essentially you need to generate a variable called "employed" that equals 1 if the individual is employed, 0 if they are unemployed or not in the labor force (NILF). You also need to change armed forces and not in universe (NIU) observations to ".". The "." means blank, so those who are in the armed forces or who were not asked the employment status question (they are NIU) are not used in the calculation.

*Here is what the code looks like if you want to get started before I demo this and provide you with the code. The lines with * in front of them are comments which won't run in Stata (you include them in your code file to help explain your code).*

*Creating a blank variable to start

gen employed = .

*Replacing this variable with 0s for those who are employed

replace employed = 1 if empstat==10 | empstat==12

*Note: the "|" means "OR", so it's setting employed = 1 if empstat equals 10 or 12, which are both ways to code someone as employed. Also, not that after an if statement, you need to use two equal signs for "equals".

*Now, replacing the variable with 0s for those who are unemployed

replace employed = 0 if empstat>=20 & empstat<=22

*Now, replacing the variable with 0s for those who are NILF

replace employed = 0 if empstat>=30

*Based on how we set up the variable, those in the armed forces and "not in universe" (NIU), meaning they were not asked this question are coded as missing since we didn't change those values to 0 or 1. But if we wanted to confirm our coding, we could run:

*replace employed = . if empstat==0 | empstat==1

*If we were to run that line of code above, it should result in no changes

To calculate the employment rate using population weights (I will explain what those are when I demo the code), use the mean command with the [pw=] option. So, the command looks like this:

mean employed [pw=wtfinl]

where wtfinl is the name of the weight variable to be used for most calculations.

<u>Submitting the Assignment (2 points)</u>

Please submit the three following documents. These can be uploaded with your submission:
1) **Your answers to the questions in a document.** This has just the answers in an organized format using a word processing document. This saves me from having to dig through your log file to find the correct results. It will probably just be three sentences: The estimated employment rate is …, the standard error of the estimate is…, the 95% confidence interval of the estimate is…
2) **Your .do file for your code.** Your .do file must be nicely annotated so that I can follow along with what you are doing (I will show you what I mean in class, or see the annotations in *s above).
3) **Your log file for your code** (this log file saves all the output that appears on the screen, usually has the extension ".smcl"). Please don't submit a log file where you opened it, let Stata run a bunch of erroneous commands (then corrected commands), then closed the file. Please submit a log file that runs all of your commons in a clean way. This involves having "log open" at the start of your final .do file, and "log close" at the end, and then running that final .do file to produce a clean log file.

I am awarding 2 points for correct submission of these files. Points will be deducted for missing files, log or do files that are difficult to interpret (e.g., contain incorrect code or output).