Session #2: Job Training

A persistent challenge in US public policy is how to help individuals who are out of work find jobs. A common solution is enrollment in job training programs to build skills sought by employers. Every year, tens of billions of dollars are spent on such programs, which are often very expensive. However, it is not clear whether the programs actually work—even if individuals get jobs after having been enrolled in the programs, they might have gotten a job anyway. If the programs don't work, then that money could be reallocated to more productive programs, and policymakers should think about other approaches to help those out of work.

Prior to LaLonde (1986), there was significant debate among researchers about the effect of job training programs. This paper used a randomized experiment for low-income men during 1976-1977 to test the effect of job training programs. We will discuss randomized experiments in the social sciences in more detail later in the course. Today we will use this data to explore the relationship between job training and employment.

The data `(jtrain_small.dta)` can downloaded from the class web site. If you want to put all your Stata code into a .do file you can, but this week you don't have to—You can answer all the questions by working interactively. You will likely find the Stata Basics handout (found on the Learning Stata page of the course web site) helpful if this is your first experience with Stata.

At the end of the session, you should have a Word document that contains your answer and the relevant output (including graphs) that you pasted in.

1. How many variables are in the data set? How many observations are there? To do this, type `describe` into the command line. You can use the `browse` command to look directly at the data.

2. `idno` is a variable that contains a unique individual identifier for each observation. Use the command `tab` to check the values of this variable.

3. Use the command `codebook` to verify that indeed this variable takes 435 unique values, and that there are no duplicates and no missing values in this variable.

4. Use the command `egen max_idno=max(idno)` to calculate the maximum value of this variable. Verify that this is indeed 435.

5. We first want to look at the characteristics of the sample, starting with the initial levels of education. Use the `summarize`, `tabulate` and `histogram` commands with the variable educ (e.g. `summarize educ`)

    a. What are the mean and standard deviation of educational attainment? How about the median? To get the median, type `summarize educ, detail`

    b. Create a histogram of the data on education. If you want to use code to save the graph to your computer, you can use the command `graph export` after you create the histogram (e.g. `graph export "../histo.png"`, replace would save it to that folder as a .png file)

6. Visualizing our data is often helpful. The variable re78 is earnings from 1978, measured in thousands of 1982 dollars. Create a scatter plot of how earnings are related to education, using the command `scatter`.

7. Sometimes there are problems in the data that we are given that we need to fix due to data entry error or other issues. It is thus useful to go through the data and check that there is not anything that might not be quite right.

    a. Check through the variables married, age, and black using the commands that you learned earlier. What is the mean of each variable? Do you see anything that looks off?

    b. In cases where you think an observation has been recorded incorrectly, we need to figure out what to do. Oftentimes, we will "drop" that observation of the variable by replacing it with a "missing value", basically a blank space. To do that, you can use the replace command, such as
    `replace var = . if var == 200.`
    For any observations of var that are equal to 200, they get replaced as a missing value in the variable var.

    c. What is the new mean for each of the variables after you make this adjustment?

8. We will now look at the training.

    a. How do the characteristics of men who received training compare to those who did not? Compare the average years of education, likelihood of being married and age in 1977 (e.g. `summarize educ if train == 0`).

    b. Do the groups look similar or different in their characteristics? Do you think this is a good or bad thing if we want to compare between the groups to figure out the effect of the training?

9. We now want to look at the effects of the job training program.

    a. Find the averages of re78 for the sample of men receiving job training and the sample not receiving job training. Is the difference economically large?

    b. Now plot earnings (re78) by whether the individual did or did not get training. (Hint: use the command `hist re78, by(train)`.) Do you note differences between the two distributions? Does it appear that the job training program was effective?

References

LaLonde, Robert J. (1986). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *American Economic Review 76*(4), 604–620.