IDS 702
Modeling and Representation of Data

Effects of Job Training on Wages

In the 1970s, researchers in the United States ran several randomized experiments intended to evaluate public policy programs. One of the most famous experiments is the National Supported Work Demonstration (NSWD), in which researchers wanted to assess whether or not job training for disadvantaged workers had an effect on their wages. Eligible workers were randomly assigned either to receive job training or not to receive job training. Since this is a randomized experiment, we can make causal claims about the effect of job training on wages for this population of workers.

We analyze a subset of the data from the NSWD. These and other data were originally analyzed in a highly influential paper by the economist Robert Lalonde. The reference for the study is

Lalonde, R. J. (1986), Evaluating the econometric evaluations of training programs with experimental data, *The American Economic Review*, 76, 604 - 620.

Here is a link to the paper. The data are in the Course Datasets folder on the IDS 702 Sakai course page; they are named "Lalonde data".

We will use linear and logistic regression modeling to answer the following questions of interest.

- Is there evidence that workers who receive job training tend to earn higher wages than workers who do not receive job training? What is a likely range for the effect of training? Is there any evidence that the effects differ by demographic groups? Are there other interesting associations with wages that are worth mentioning?
- Is there evidence that workers who receive job training tend to be more likely to have positive (non-zero) wages than workers who do not receive job training? What is a likely range for the effect of training? Is there any evidence that the effects differ by demographic groups? Are there other interesting associations with positive wages that are worth mentioning?

Write a report describing your findings. The report should have two separate sections, one answering the first set of questions about wages and the other answering the second set of questions about positive wages. Be sure to include the following in each section of your report: the model you ultimately decided to use, a justification for that model (e.g., why you chose certain transformations and why you decided the final model is reasonable to use based on residual diagnostics), the relevant regression output (includes: a table with coefficients and SEs, and p-values or confidence intervals; and somewhere in the text or table the estimated regression standard deviation and R-squared for linear regression, or area under the curve for logistic regression), your interpretation of the results in the context of the questions of interest, and any potential limitations of the analysis.

In this assignment, you work in pairs. You are free to divide the work any way you choose. Please turn in one report with both team members' names at the top. Please limit each section of your report to 5 typed pages, for a total of 10 typed pages. Upload all your analysis code to the Dropbox for the FIRST author listed on the report. You also can upload any supplemental material that is important for your analysis, such as diagnostic checks or exploratory plots that you feel justify the conclusions in your report. Please briefly describe the materials that you uploaded in a few sentences in s separate section at the end of your report.

Code Book

Variable	Description
treat	1 if received job training, 0 if did not receive job training.
age	age in years.
educ	years of education.
black	1 if race is black, 0 otherwise.
hisp	1 if Hispanic ethnicity, 0 otherwise.
married	1 if married, 0 otherwise.
nodegree	1 if dropped out of high school, 0 otherwise.
re74	real annual earnings in 1974.
re75	real annual earnings in 1975.
re78	real annual earnings in 1978.