

## Exercise 8

The file **JTRAIN2.GDT** contains data on a job training experiment for a group of men. Men could enter the program starting in January 1976 through about mid-1977. The program ended in December 1977. The idea is to test whether participation in the job training program had an effect on unemployment probabilities and earnings in 1978.

- (i) The variable *train* is the job training indicator. How many men in the sample participated in the job training program? What was the highest number of months a man actually participated in the program?
- (ii) Run a linear regression of *train* on several demographic and pretraining variables: *unem74*, *unem75*, *age*, *educ*, *black*, *hisp*, and *married*. Are these variables jointly significant at the 5% level?
- (iii) Estimate a probit version of the linear model in part (ii). Compute the likelihood ratio test for joint significance of all variables. What do you conclude?
- (iv) Based on your answers to parts (ii) and (iii), does it appear that participation in job training can be treated as exogenous for explaining 1978 unemployment status? Explain.
- (v) Run a simple regression of *unem78* on *train* and report the results in equation form. What is the estimated effect of participating in the job training program on the probability of being unemployed in 1978? Is it statistically significant?
- (vi) Run a probit of *unem78* on *train*. Does it make sense to compare the probit coefficient on *train* with the coefficient obtained from the linear model in part (v)?
- (vii) Find the fitted probabilities from parts (v) and (vi). Explain why they are identical. Which approach would you use to measure the effect and statistical significance of the job training program?
- (viii) Add all of the variables from part (ii) as additional controls to the models from parts (v) and (vi). Are the fitted probabilities now identical? What is the correlation between them?