ECON 6306 Applied Econometrics — Fall 2016 Problem Set 1

1. A longitudinal experiment was conducted in Tennessee beginning in 1985 and ending in 1989. A single cohort of students was followed from kindergarten through third grade. In the experiment children were randomly assigned within schools into three types of classes: small classes with 13–17 students, regular-sized classes with 22–25 students, and regular-sized classes with a full-time teacher aide to assist the teacher. Student scores on achievement tests were recorded as well as some information about the students, teachers, and schools. Data for the kindergarten classes are contained in the data file *star.csv*.(Use either R or STATA for this question and attach the code and the output with the answers)

   (a) Using children who are in either a regular-sized class or a small class, estimate the regression model explaining students' combined aptitude scores as a function of class size, $TOTALSCORE_i = \beta_1 + \beta_2 SMALL_{+} e_i$. Interpret the estimates. Based on this regression result, what do you conclude about the effect of class size on learning?(*Hint*:Remember to exclude classes with teacher aide from the sample)

   (b) Repeat part (a) using dependent variables $READSCORE$ and $MATHSCORE$. Do you observe any differences?

   (c) Using children who are in either a regular-sized class or a regular-sized class with a teacher aide, estimate the regression model explaining student's combined aptitude scores as a function of the presence of a teacher aide,$TOTALSCORE_i = \gamma_1 + \gamma_2 AIDE + e_i$. Interpret the estimates. Based on this regression result, what do you conclude about the effect on learning of adding a teacher aide to the classroom? (*Hint*:Remember to exclude small classes from the sample)

   (d) Repeat part (c) using dependent variables $READSCORE$ and $MATHSCORE$. Do you observe any differences?

2. You have the results of a simple linear regression based on state-level data and the District of Columbia, a total of N = 51 observations.

   (a) The estimated error variance $\hat{\sigma}^2$=2.04672. What is the sum of the squared least squares residuals?

   (b) The estimated variance of b_2 is 0.00098. What is the standard error of b_2? What is the value of $\sum(x_i - \bar{x})^2$?

   (c) Suppose the dependent variable y_i=the state's mean income (in thousands of dollars) of males who are 18 years of age or older and x_i the percentage of males 18 years or older who are high school graduates. If $b_2 = 0.18$, interpret this result.

   (d) Suppose $\bar{x}$=69.139 and $\bar{y}$= 15.187, what is the estimate of the intercept parameter?

   (e) Given the results in (b) and (d), what is $\sum x_i^2$?

   (f) For the state of Arkansas the value of y_i= 12.274 and the value of x_i=58.3. Compute the least squares residual for Arkansas.(*Hint*:Use the information in parts (c) and (d))

3. The general manager of an engineering firm wants to know whether a technical artist's experience influences the quality of his or her work. A random sample of 24 artists is selected and their years of work experience and quality rating (as assessed by their supervisors) recorded. Work experience ($EXPER$) is measured in years and quality rating ($RATING$) takes a value of 1 through 7, with 7 = excellent and 1 = poor: The simple regression model $RATING = \beta_1 + \beta_2 EXPER + e$ is proposed. The least squares estimates of the model, and the standard errors of the estimates, are

$$\underset{(se)}{\widehat{RATING}} = \underset{(0.709)}{3.204} + \underset{(0.044)}{0.076EXPER}$$

   (a) Sketch the estimated regression function. Interpret the coefficient of $EXPER$.

   (b) Construct a 95 % confidence interval for β_2, the slope of the relationship between quality rating and experience. In what are you 95 % confident?

   (c) Test the null hypothesis that β_2 is zero against the alternative that it is not using a two-tail test and the $\alpha = 0.05$ level of significance. What do you conclude?

   (d) Test the null hypothesis that β_2 is zero against the one-tail alternative that it is positive at the $\alpha = 0.05$ level of significance. What do you conclude?

   (e) For the test in part (c), the p-value is 0.0982. If we choose the probability of a Type I error to be $\alpha = 0.05$; do we reject the null hypothesis, or not, just based on an inspection of the p-value?