John Chao
Spring 2015
Due Date: 4/14/2015

Economics 422

Econometrics I

Problem Set 3

1. Use the data in WAGE2.DTA to estimate a simple regression explaining monthly salary ($wage$) in terms of IQ score (IQ).

   (a) Find the average salary and average IQ in the sample. What is the sample standard deviation of IQ? (IQ scores are standardized so that the average in the population is 100 with a standard deviation equal to 15.)

   (b) Estimate a simple regression model where a one-point increase in IQ changes $wage$ by a constant dollar amount. Use this model to find the predicted increase in wage for an increase in IQ of 15 points. Does IQ explain most of the variation in $wage$?

2. The data in 401K.DTA are a subset of the data that were analyzed in a scholarly paper by Papke (1995) which studies the relationship between participation in a 401(k) pension plan and the generosity of the plan. The variable $prate$ is the percentage of eligible workers with an active account; this is the variable we would like to explain. The measure of the generosity is the plan match rate, $mrate$. This variable gives the average amount the firm contributes to each worker's plan for each \$1 contribution by the worker. For example, if mrate = 0.50, then a \$1 contribution by the worker is matched by a 50¢ contirbution by the firm.

   (a) Find the average participation rate and the average match rate in the sample of plans.

   (b) Now, estimate the simple regression equation

   $$\widehat{prate} = \widehat{\beta}_0 + \widehat{\beta}_1 mrate,$$

   and report the results along with the sample size and R-squared.

   (c) Interpret the intercept in your equation. Interpret the coefficient on $mrate$.

   (d) Find the predicted $prate$ when $mrate = 3.5$. Is this a reasonable prediction? Explain what is happening here.

(e) How much of the variation in *prate* is explained by *mrate*? Is this a lot in your opinion?

3. Use the data in SLEEP75.DTA from Biddle and Hamermesh (1990) to study whether there is a tradeoff between the time spent sleeping per week and the time spent in paid work. Estimate the model

$$sleep = \beta_0 + \beta_1 totwrk + u,$$

where *sleep* is minutes spent sleeping at night per week and *totwrk* is total minutes worked during the week.

a. Report your results in equation form along with the number of observations and R^2. What does the intercept in this equation mean?

b. If *totwrk* increases by 2 hours, by how much is *sleep* estimated to fall? Do you find this to be a large effect?

4. Suppose that (Y_i, X_i) satisfy the assumptions for the linear regression model given in class; and, in addition, assume that u_i is $N(0, \sigma_u^2)$ and is independent of X_i. A sample of size $n = 30$ yields

$$\widehat{Y} = \underset{(10.2)}{43.2} + \underset{(7.4)}{61.5}X, \; R^2 = 0.54, \; SER = 1.52,$$

where the numbers in parentheses are the homoskedastic-only standard errors for the regression coefficients.

(a) Construct a 95% confidence interval for β_0.

(b) Test $H_0 : \beta_1 = 55$ versus $H_0 : \beta_1 \neq 55$ at the 5% level.

(c) Test $H_0 : \beta_1 = 55$ versus $H_0 : \beta_1 > 55$ at the 5% level.