Introductory Econometrics

Tutorial 6

<u>PART A:</u> To be done before you attend the tutorial. The solutions will be made available at the end of the week.

1. *Using reported information to work out other important statistics:* It is quite possible that we are given a report with some regression result, with some statistics reported. But something like R^2 may not be reported. However, we may want to be able to compute the missing statistics from the reported ones. This is one example:

$$\begin{aligned} \hat{y}_i &= 150 - 0.2x_{i1} + 2.1x_{i2} + 1.2x_{i3},\ i = 1, \ldots 44 \\ \hat{\sigma} &= 21.5 \text{ (standard error of the regression)} \\ \hat{\sigma}_y &= 50 \text{ (sample standard deviation of the dependent variable)} \end{aligned}$$

   (a) Compute the R^2 of this regression. Remember $R^2 = 1 - \frac{SSR}{SST}$. Compute SSR and SST using the information provided above, and then compute the R^2.

   (b) Test the overall significance of this reported regression at the 5% level of significance (use the R^2 to compute the F statistic for overall significance).

2. File named TeachingRatings.WF1 contains data on unit evaluation (course_eval), unit characteristics and professor characteristics for 463 units at the University of Texas at Austin. Professor characteristics include an index of the professor's beauty as rated by a panel of six judges. This index is constructed to have sample average of 0, so positive values of the index mean above average beauty and negative values mean below average beauty. Is the professor's beauty a significant predictor of unit evaluations?

   (a) Use Eviews to construct a scatterplot of unit evaluation on the professor's beauty. Does there appear to be a relationship between the two variables?

   (b) Run a regression of average unit evaluation against professor's beauty. What is the estimated intercept? What is the estimated slope? Explain why the estimated intercept is equal to the sample mean of the unit evaluation variable.

   (c) Test the null hypothesis that professor's beauty has no predictive power to predict unit evaluations versus the alternative that it does at the 5% level of significance.

   (d) Does beauty explain a large fraction of the variance in evaluations across modules?

Do not forget to bring your answers to PART A and a copy of the tutorial questions to your tutorial.

<u>Part B:</u> This part will be covered in the tutorial. It is still a good idea to attempt these questions before the tutorial.
The purpose of this tutorial is to practice hypothesis testing.

1. *Hypothesis test on a single parameter, the meanings of the size of a test and a confidence interval*: Consider the classical linear model

$$y_i = \beta_0 + \beta_1 x_i + u_i,\ i = 1, 2, \ldots, n$$

   A random sample of size $n = 22$ is drawn and the estimated model based on this sample is:

$$\begin{aligned} \hat{y}_i &= \underset{(3.1)}{5.4} + \underset{(1.5)}{3.2}\, x_i,\ i = 1, 2, \ldots, 22 \\ R^2 &= 0.26. \end{aligned}$$

(a) Test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ at the 5% level.

(b) Construct a 95% confidence interval for β_1.

(c) Suppose you learn that y and x were independent. Would you be surprised? Explain.

(d) Suppose that y and x are independent and many samples of size $n = 22$ are drawn, regressions estimated, and (a) and (b) answered. In what fraction of the samples would H_0 from (a) be rejected? In what fraction of samples would the confidence intervals from (b) include the value $\beta_1 = 0$?

2. *Practice with t-test and F-test:* (This is based on problem 3 at the end of Chapter 3 of the textbook): The following multiple regression model is used to study the trade-off between time spent sleeping and working and to look at other factors affecting sleep:

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + u$$

where *sleep* and *totwrk* are measured in minutes per week and *educ* and *age* are measured in years.

(a) If adults trade-off sleep for work, what is the sign of β_1?

(b) What signs do you think β_2 and β_3 will have?

(c) Using data from a random sample of 706 adults, we have estimated the following equation:

$$\begin{aligned} \widehat{sleep} &= \underset{(112.27)}{3638.25} - \underset{(0.017)}{0.148} totwrk - \underset{(5.88)}{11.13} educ + \underset{(1.45)}{2.20} age \\ R^2 &= 0.113, \; SSR = 123455057 \end{aligned} \tag{1}$$

Test the hypothesis that adults do not trade-off sleep for work against the alternative that they do at the 1% level of significance.

(d) Construct a 95% confidence interval for β_3. Interpret this confidence interval.

(e) We have also estimated the following regression:

$$\begin{aligned} \widehat{sleep} &= \underset{(38.91)}{3586.38} - \underset{(0.017)}{0.151} totwrk \\ SSR &= 124858119 \end{aligned} \tag{2}$$

Test the joint hypothesis given work time, education and age have no effect on sleep time versus the alternative that at least one of them does. Perform this test at the 5% level of significance.

(f) Compute the R^2 of the regression (2).

(g) Suppose that someone suggests that one year of education keeping all else constant has the same effect but with opposite sign of the effect of one more year of age keeping all else constant. That is, $\beta_2 = -\beta_3$. Explain how you would test this hypothesis with an F-test. You need to state the alternative hypothesis that can be tested with an F-test, specify any extra regression that you need to estimate, and explain how you would use the results of that regression to test this hypothesis.

(h) Suppose the alternative hypothesis of interest was $\beta_2 < -\beta_3$. Explain how you would test $H_0 : \beta_2 = -\beta_3$ against this one-sided alternative.

(i) If time permits, perform (g) and (h) at the 5% level of significance using data in sleep75.wf1. In the light of the results of these tests, comment on how focusing on the magnitude of OLS estimates without any notice of their standard errors can be misleading.