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.

$$
\begin{aligned}
\hat{\sigma}_y^2 &= \frac{\sum_{i=1}^{44} (y_i - \bar{y})^2}{44 - 1} = \frac{SST}{43} \Rightarrow SST = 43 \times 50^2 = 107500 \\
\hat{\sigma}^2 &= \frac{\sum_{i=1}^{44} \hat{u}_i^2}{44 - 3 - 1} = \frac{SSR}{40} \Rightarrow SSR = 40 \times 21.5^2 = 18490 \\
R^2 &= 1 - \frac{SSR}{SST} = 1 - \frac{18490}{107500} = 0.828
\end{aligned}
$$

(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).

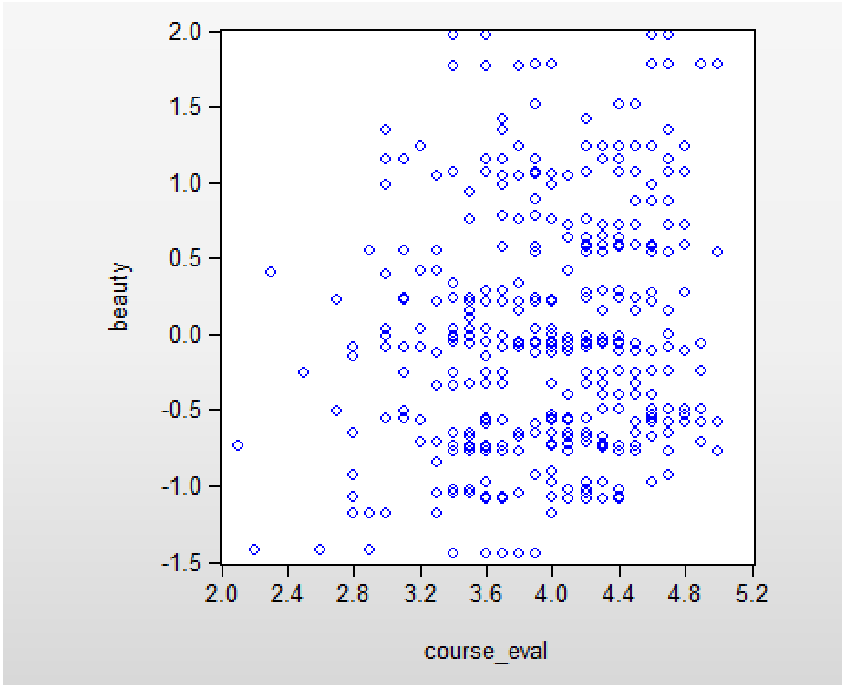
$$
\begin{aligned}
H_0 &: \beta_1 = \beta_2 = \beta_3 = 0 \\
H_1 &: \text{at least one of the above is not zero} \\
F &= \frac{R^2/3}{(1 - R^2)/(44 - 3 - 1)} \sim F_{3,40} \ \text{ under } H_0 \\
F_{calc} &= \frac{0.828/3}{(1 - 0.828)/40} = 64.2 \\
F_{crit} &= 2.84 \\
F_{calc} &> F_{crit} \text{ therefore we reject the null}
\end{aligned}
$$

We conclude that at least one of the explanatory variables is a significant predictor of y.

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 average module evaluations on the professor's beauty. Does there appear to be a relationship between the two variables?

Answer The scatterplot between module evaluation and beauty is shown below:



There appears to be a weak positive relationship between course evaluation and the beauty index.

b. Run a regression of average module 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 module evaluation variable.

Answer The regression output produced by Eviews is shown in the table below:

Dependent Variable: COURSE_EVAL
Method: Least Squares
Date: 03/29/18 Time: 20:33
Sample: 1 463
Included observations: 463

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.998272	0.025349	157.7272	0.0000
BEAUTY	0.133001	0.032178	4.133368	0.0000

R-squared	0.035736	Mean dependent var	3.998272
Adjusted R-squared	0.033644	S.D. dependent var	0.554866
S.E. of regression	0.545452	Akaike info criterion	1.629905
Sum squared resid	137.1556	Schwarz criterion	1.647779
Log likelihood	-375.3231	Hannan-Quinn criter.	1.636942
F-statistic	17.08473	Durbin-Watson stat	1.410317
Prob(F-statistic)	0.000042		

The estimated regression line is given by:

$$\widehat{Course\_\text{eval}} = \underset{(0.03)}{4.00} + \underset{(0.03)}{0.133} Beauty$$

We are told that the Beuty variable is constructed such that its mean in this sample is 0. We know, that using the OLS estimation method, the estimated intercept is equal to the mean of the dependent variable (Course_eval) minus the estimated slope (0.133) times the mean of the regressor (Beauty), or

$$\hat{\beta}_0 = \overline{Course\_\text{eval}} - \hat{\beta}_1 \overline{Beauty}.$$

Thus, the estimated intercept is equal to the mean of Course_eval.

c.

$$\begin{aligned}
H_0 &: \beta_1 = 0 \\
H_1 &: \beta_1 \neq 0 \text{ (OK if you think a one-sided alternative makes better sense,} \\
&\quad \text{then need to use a critical value appropriate for your one sided alternative)} \\
t &= \frac{\hat{\beta}_1}{se\left(\hat{\beta}_1\right)} \sim t_{433-2} \text{ under } H_0 \\
t_{calc} &= 4.333 \\
t_{crit} &= 1.980 \text{ (2-tailed 5\%)} \\
t_{calc} &> t_{crit} \text{ therefore we reject the null hypothesis} \\
&\quad \text{Our conclusion is that beauty is a significant predictor of course evaluations.}
\end{aligned}$$

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

Answer The regression R^2 is 0.036, so that Beauty explains only 3.6% of the variance in the course evaluations, which is very small.

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.

$$\begin{aligned}
H_0 &: \beta_1 = 0 \\
H_1 &: \beta_1 \neq 0 \\
t &= \frac{\hat{\beta}_1}{se\left(\hat{\beta}_1\right)} \sim t_{20} \text{ under } H_0 \\
t_{calc} &= \frac{3.2}{1.5} = 2.133 \\
t_{crit} &= 2.086 \\
t_{calc} &> t_{crit} \Rightarrow \text{ we reject the null}
\end{aligned}$$

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

$$3.2 \pm 2.086 \times 1.5 = [0.071, 6.329]$$

(c) Suppose you learn that y and x were independent. Would you be surprised? Explain.
I will be somewhat surprised, but will understand that there is a 5% change of wrongly rejecting the null when it is true and this has been one such instance.

(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 $\beta_1 = 0$?
The null hypothesis that $\beta_1 = 0$ would be rejected at the 5% level in 5% of the samples; 95% of the confidence intervals would contain 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?
Negative.

(b) What signs do you think β_2 and β_3 will have?
Other things equal, I guess older people need more sleep, but I also know that university students sleep a lot more than 25-35 year olds. So, don't know really. I have no idea about if two people have the same age and work the same number of hours, why the more educated one should sleep more or sleep less. Perhaps if we consider that less educated people may be working more physical work, then we can say that less educated people need more sleep to recover, so the sign would be negative if that is true.

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

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

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

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_1 &: \beta_1 < 0 \\ t &= \frac{\hat{\beta}_1}{se\left(\hat{\beta}_1\right)} \sim t_{706-3-1} \text{ under } H_0 \\ t_{calc} &= \frac{-0.148}{0.017} = -8.706 \\ t_{crit} &= -2.358 \\ t_{calc} &< t_{crit} \Rightarrow \text{ we reject the null} \\ & \quad \text{There is a trade-off between work time and sleep time} \end{aligned}$$

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

$$\hat{\beta}_3 \pm 1.98 * se\left(\hat{\beta}_3\right) = 2.20 + 1.98 * 1.45 = [-0.671, 5.071]$$

In repeated samples, 95% of such intervals will contain β_3. We can give the range given by the confidence interval as our interval estimate for β_3 and state that there is a 5% chance that we are wrong. Note that the interval includes 0, which means that age is not a significant predictor of sleep time after controlling for work time and education.

(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.

$$\begin{aligned} H_0 &: \beta_2 = \beta_3 = 0 \\ H_1 &: \text{at least one of } \beta_2 \text{ or } \beta_3 \text{ is not equal to zero} \\ F &= \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/702} \sim F_{2,702} \text{ under } H_0 \\ F_{calc} &= \frac{(124858119 - 123455057)/2}{123455057/702} = 3.989 \\ F_{crit} &= 3.07 \\ F_{calc} &> F_{crit} \Rightarrow \text{ we reject the null} \\ & \text{Given work time, at least one of education or age has a significant effect on sleep time} \end{aligned}$$

(f) Compute the R^2 of the regression (2).
Both equations have the same left hand side variable, so they both have the same SST. From equation 1:

$$R_1^2 = 1 - \frac{SSR_1}{SST} \Rightarrow SST = \frac{SSR_1}{1 - R_1^2} = \frac{123455057}{1 - 0.113} = 139182702.4$$

Using this, we can calculate the R^2 of equation 2:

$$R^2 = 1 - \frac{SSR_2}{SST} = 1 - \frac{124858119}{139182702.4} = 0.103$$

(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.

$$\begin{aligned} H_0 &: \beta_2 = -\beta_3 \\ H_1 &: \beta_2 \neq -\beta_3 \\ F &= \frac{(SSR_r - SSR_{ur})/1}{SSR_{ur}/702} \sim F_{1,702} \text{ under } H_0 \\ F_{calc} &= \text{using the regressions explained below} \\ F_{crit} &= \text{from the } F \text{ table given the size of the test} \\ \text{if } F_{calc} &> F_{crit} \Rightarrow \text{ we reject the null, and we don't reject otherwise} \end{aligned}$$

SSR_{ur} is from equation (1) stated above. The restricted model is:

$$\begin{aligned} sleep &= \beta_0 + \beta_1 totwrk - \beta_3 educ + \beta_3 age + u \\ &= \beta_0 + \beta_1 totwrk + \beta_3 (age - educ) + u \end{aligned}$$

Estimating this restricted model gives us SSR_r, and then we can compute F_{calc}.

(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.
Under the null $\beta_2 + \beta_3 = 0$. We denote $\beta_2 + \beta_3 = \delta$, which implies $\beta_2 = \delta - \beta_3$. We use this to reparameterise the model:

$$\begin{aligned} sleep &= \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + u \\ &= \beta_0 + \beta_1 totwrk + (\delta - \beta_3) educ + \beta_3 age + u \\ &= \beta_0 + \beta_1 totwrk + \delta educ + \beta_3 (age - educ) + u \\ \widehat{sleep} &= \hat{\beta}_0 + \hat{\beta}_1 totwrk + \hat{\delta} educ + \hat{\beta}_3 (age - educ) \\ H_0 &: \delta = 0 \Rightarrow \beta_2 = -\beta_3 \\ H_1 &: \delta < 0 \Rightarrow \beta_2 < -\beta_3 \\ t_{\hat{\delta}} &= \frac{\hat{\delta}}{se\left(\hat{\delta}\right)} \sim t_{706-3-1} \text{ under } H_0 \\ t_{calc} &= \text{using the estimated reparameterised model} \\ t_{crit} &= \text{from the } t \text{ table for a 1 tailed test at the given size} \\ \text{if } t_{calc} &< t_{crit} \Rightarrow \text{ we reject the null, and we don't reject otherwise} \end{aligned}$$

(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.
For part (g), the estimated restricted model is:

Dependent Variable: SLEEP
Method: Least Squares
Sample: 1 706
Included observations: 706

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3500.007	52.34247	66.86744	0.0000
TOTWRK	-0.148581	0.016704	-8.894940	0.0000
AGE-EDUC	3.140886	1.278651	2.456406	0.0143

R-squared	0.110918	Mean dependent var	3266.356
Adjusted R-squared	0.108389	S.D. dependent var	444.4134
S.E. of regression	419.6381	Akaike info criterion	14.92090
Sum squared resid	123795567	Schwarz criterion	14.94028
Log likelihood	-5264.079	Hannan-Quinn criter.	14.92839
F-statistic	43.85182	Durbin-Watson stat	1.941704

$$F_{calc} = \frac{(123795567 - 123455057)/1}{123455057/702} = 1.936$$

$$F_{crit} = 3.92$$

$$F_{calc} < F_{crit} \Rightarrow \text{ We cannot reject the null}$$

There is not enough evidence to reject the hypothesis that $\beta_2 = -\beta_3$. For part (h) the estimated reparameterised model is:

Dependent Variable: SLEEP
Method: Least Squares
Sample: 1 706
Included observations: 706

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3638.245	112.2751	32.40474	0.0000
TOTWRK	-0.148373	0.016694	-8.888075	0.0000
EDUC	-8.933928	6.420423	$\boxed{-1.391486}$	0.1645
AGE-EDUC	2.199885	1.445717	1.521657	0.1285

R-squared	0.113364	Mean dependent var	3266.356
Adjusted R-squared	0.109575	S.D. dependent var	444.4134
S.E. of regression	419.3589	Akaike info criterion	14.92098
Sum squared resid	1.23E+08	Schwarz criterion	14.94681

$t_{calc} = -1.3915$

$t_{crit} = -1.658$

$t_{calc} \nless t_{crit}$ therefore we cannot reject the null

there is no evidence to reject the hypothesis that $\beta_2 = -\beta_3$ in favour of the one-sided alternative that $\beta_2 < -\beta_3$.
Please emphasise that the results of the reparameterised model are exactly the same as the results of the unrestricted model. The only reason we do the reparameterisation is to compute the standard error of $\hat{\beta}_2 + \hat{\beta}_3 = \hat{\delta}$ so that we can do a t-test on it.
Also emphasise that it was very hard to say that data cannot reject $\beta_2 = -\beta_3$ by just considering the parameter estimates -11.13 and 2.20 without noticing that they were not very precisely estimated.