MiniQuiz - Week 6

1. **01 - Week 6**

   MULTI | 1.0 point | 0.10 penalty | Single | Shuffle

   Albury and Wodonga are two adjacent towns on the border between New South Wales and Victoria. Albury is in NSW and Wodonga is in VIC. We have the selling price and the location of the last 32 properties sold in Albury-Wodonga. We have generated the binary variable W, which is equal to 1 if the property is in Wodonga, and is equal to 0 otherwise. It is believed that people prefer to locate in NSW than in VIC, and as a result mean of house prices in Albury is higher than the mean of house prices in Wodonga. The null and the alternative that would allow us to test this belief in the model:

   $$price_i = \beta_0 + \beta_1 W_i + u_i$$

   are:

   (a) $H_0 : \beta_0 = 0$ against $H_1 : \beta_0 < 0$
   (b) $H_0 : \beta_0 = 0$ against $H_1 : \beta_0 > 0$
   (c) $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 < 0$ (100%)
   (d) $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 > 0$
   (e) $H_0 : W_i = 0$ against $H_1 : W_i < 0$

   β_1 is the difference between house prices in Albury and Wodonga. The null of no difference is $H_0 : \beta_1 = 0$ and the alternative that confirms that the mean of house prices in Wodonga is less than the mean in Albury is $H_1 : \beta_1 < 0$

2. **02 - Week 6**

   MULTI | 1.0 point | 0.10 penalty | Single | Shuffle

   Suppose we use OLS to estimate the parameters of the model explained in the previous question and we denote the estimated model by:

   $$\widehat{price}_i = \widehat{\beta}_0 + \widehat{\beta}_1 W_i$$

   and we denote the standard error of $\widehat{\beta}_1$ by $se(\widehat{\beta}_1)$. The statistic $\frac{\widehat{\beta}_1}{se(\widehat{\beta}_1)}$ has:

   (a) a t_{32} distribution if $\beta_1 = 0$
   (b) a t_{30} distribution if $\beta_1 = 0$ (100%)
   (c) a t_{30} distribution if $\widehat{\beta}_1 = 0$
   (d) a t_{30} distribution if $\widehat{\beta}_1 < 0$
   (e) a t_{30} distribution if $\beta_1 < 0$

The t-statistic has a t with $n - k - 1 = 32 - 1 - 1 = 30$ degrees of freedom only if the null hypothesis $\beta_1 = 0$ is true.

3. **03 - Week 6**

MULTI | 1.0 point | 0.10 penalty | Single | Shuffle

Using the data explained in the previous question we have estimated the following equation using OLS:

$$\widehat{price_i} = \underset{(105.1)}{590.5} - \underset{(30.1)}{90.3} W_i$$

where the number in parentheses below each parameter estimate is its standard error. Which one of the following statements is correct?

(a) We do not reject the null hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_1 : \beta_1 < 0$ at the 5 percent level of significance because $t_{calc} = \frac{-90.3}{30.1} = -3$ is less than 1.697.
(b) We do not reject the null hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_1 : \beta_1 < 0$ at the 5 percent level of significance because $t_{calc} = \frac{-90.3}{30.1} = -3$ is less than 2.042.
(c) We do not reject the null hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_1 : \beta_1 < 0$ at the 5 percent level of significance because $t_{calc} = \frac{-90.3}{30.1} = -3$ is less than -1.697.
(d) We do not reject the null hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_1 : \beta_1 < 0$ at the 5 percent level of significance because $t_{calc} = \frac{-90.3}{30.1} = -3$ is less than -2.042.
(e) We reject the null hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_1 : \beta_1 < 0$ at the 5 percent level of significance because $t_{calc} = \frac{-90.3}{30.1} = -3$ is less than -1.697. (100%)

Given that the alternative is $H_1 : \beta_1 < 0$, the rejection region is in the left tail. The critical value for the left tail of t_{30} distribution for a 5 percent level is -1.697, and we reject if $t_{calc} < -1.697$.

4. **04 - Week 6**

MULTI | 1.0 point | 0.10 penalty | Single | Shuffle

The following equation reports the results of a multiple regression model estimated using OLS:

$$\hat{y}_i = 150 - 0.2x_{i1} + 2.1x_{i2} + 1.2x_{i3}, \; i = 1, \ldots 44$$

The standard error of the regression is reported to be $\hat{\sigma} = 21.5$ and its total sum of squares (SST) is also reported to be 107500. Based on this information,

(a) $R^2 = 1 - \frac{40 \times 21.5^2}{107500} = 0.828$ (100%)

(b) $R^2 = \frac{40\times21.5^2}{107500} = 0.172$
(c) $R^2 = 1 - \frac{44\times21.5^2}{107500} = 0.811$
(d) $R^2 = \frac{44\times21.5^2}{107500} = 0.189$
(e) the R^2 of the regression cannot be computed because the residual sum of squares is not reported.

$R^2 = 1 - \frac{SSR}{SST}$, *and* $\hat{\sigma} = \sqrt{\frac{SSR}{n-k-1}} \Rightarrow SSR = \hat{\sigma}^2 \times (44-4) = 21.5^2 \times 40 = 18490$. *Therefore,* $R^2 = 1 - \frac{18490}{107500} = 0.828$

5. **05 - Week 6**

MULTI | 1.0 point | 0.10 penalty | Single | Shuffle

In the multiple regression of the previous question, the F-statistic for the overall significance of the model has an

(a) $F_{4,44}$ distribution under the null
(b) $F_{3,44}$ distribution under the null
(c) $F_{3,41}$ distribution under the null
(d) $F_{4,40}$ distribution under the null
(e) $F_{3,40}$ distribution under the null (100%)

The test for the overall significance of a model has an $F_{k,n-k-1}$ *distribution under the null. Here* $n = 44$ *and* $k = 3$.

Total of marks: 5

Introductory Econometrics

Part B: 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. *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) We have also estimated the following regression:

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

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.

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

(e) 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:

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

(f) 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}.

(g) 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:

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

(h) We have performed the tests in (f) and (g) using our sample and in we could not reject the null hypothesis in either of these cases (you can verify these using sleep75.wf1 after the tutorial - solutions will be provided at the end of the week). 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 (f), 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 (g) 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 note 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 note that it was very hard to judge that there was not enough evidence to reject $\beta_2 = -\beta_3$ by considering the parameter estimates -11.13 and 2.20 only, without noticing that they were not very precisely estimated.