ECON 466 MIDTERM I ANSWER KEY

LIANG LIU

Question 1.

1. (a) Compute the t-values in the 4th and 8th columns of the above table.

Regressor	t	Regressor	t
comp	5	*ln(comp)*	5.253
staff	1.204	*staff*	1.131
enroll	-1	*ln(enroll)*	-1.793
intercept	0.372	*intercept*	-4.718

1. (b) In Model 1 interpret the coefficients of comp, staff, and enroll in simple words. (Note: if the passing rate is 40%, then $math10 = 40$, instead of 0.04.)

When *comp* increases by $1, predicted value of *math10* increases by 0.0005 percentage points.

When number of *staff* increases by one, predicted value of *math10* increases by 0.0479 percentage points.

When student enrollment (*enroll*) increases by one predicted value of *math10* decreases by 0.0002 percentage points.

1. (c) In Model 2 interpret the coefficients of ln(comp) and ln(enroll) in simple words.

When *comp* increases by 1%, predicted value of *math10* increases by 0.21173 percentage points.

When *enroll* increases by 1%, predicted value of *math10* decreases by 0.012443 percentage points.

1. (d) How would you test the hypothesis that there is no regression say at the 5% level of significance in Models 1 and 2? Please explain.

Model 1:

$$math10 = \beta_0 + \beta_1 comp + \beta_2 staff + \beta_3 enroll + u$$

H_0: $\beta_1 = \beta_2 = \beta_3 = 0$

H_1: H_0 is not true.

$$\begin{aligned} F &= \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} \\ &= \frac{(42394 + 2422 - 42394)/3}{(42394)/(100-3-1)} \\ &= 1.828 \end{aligned}$$

Or we could use the R^2 to find the F statistics,

$$R^2_{ur} = \frac{SSE}{SST} = \frac{2422}{42394 + 2422} = 0.054043$$

$$\begin{aligned} F &= \frac{R^2_{ur}/q}{(1-R^2_{ur})/(n-k-1)} \\ &= \frac{0.054043/3}{(1-0.054043)/96} \\ &= 1.828 \end{aligned}$$

Model 2:

$$math10 = \gamma_0 + \gamma_1 ln(comp) + \gamma_2 staff + \gamma_3 ln(enroll) + u$$

H_0: $\gamma_1 = \gamma_2 = \gamma_3 = 0$

H_1: H_0 is not true.

$$\begin{aligned} F &= \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} \\ &= \frac{(41853 + 2964 - 41853)/3}{(41853)/(100-3-1)} \\ &= 2.266 \end{aligned}$$

Or we could use the R^2 to find the F statistics,

$$R_{ur}^2 = \frac{SSE}{SST} = \frac{2964}{41853 + 2964} = 0.066136$$

$$\begin{aligned} F &= \frac{R_{ur}^2 / q}{(1 - R_{ur}^2)/(n-k-1)} \\ &= \frac{0.066136/3}{(1 - 0.066136)/96} \\ &= 2.266 \end{aligned}$$

Then we need to compare the F statics of the models with the critical value (c=$F_{3,95}$ at 5% significance level). If F is bigger than c, we reject the null; if F is lower than c, we cannot reject the null.

1. (e) If comp in Model 1 is measured in thousand dollars, what would happen to the estimated coefficients in column 2? How about the coefficients in Model 2 (column 6)? Please explain in details.

Model 1:

$$\widehat{math10} = \hat{\beta}_0 + \hat{\beta}_1 comp + \hat{\beta}_2 staff + \hat{\beta}_3 enroll$$

If *comp* in Model 1 is measured in thousand dollars, defining $comp^* = \frac{comp}{1000}$, then $comp = 1000 * comp^*$.

$$\widehat{math10} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 1000 * comp^* + \hat{\beta}_2 staff + \hat{\beta}_3 enroll$$

The estimated coefficients in column 2 are unchanged, except the coefficient of $comp^*$, $\hat{\beta}_1^* = 1000\hat{\beta}_1 = 1000 \cdot 0.0005 = 0.5$.

Model 2:

$$\widehat{math10} = \hat{\gamma}_0 + \hat{\gamma}_1 ln(comp) + \hat{\gamma}_2 staff + \hat{\gamma}_3 ln(enroll)$$

If *comp* in Model 2 is measured in thousand dollars, defining $comp^* = \frac{comp}{1000}$, then $comp = 1000 * comp^*$.

$$\begin{aligned} \widehat{math10} &= \hat{\gamma}_0 + \hat{\gamma}_1 ln(1000 * comp^*) + \hat{\gamma}_2 staff + \hat{\gamma}_3 ln(enroll) \\ &= \hat{\gamma}_0 + \hat{\gamma}_1 ln1000 + \hat{\gamma}_1 ln(comp^*) + \hat{\gamma}_2 staff + \hat{\gamma}_3 ln(enroll) \end{aligned}$$

The estimated coefficients in column 6 are unchanged, except the intercept, $\hat{\gamma_0^*} = \hat{\gamma_0} + \hat{\gamma_1} ln1000 = -194.15 + 21.173 \cdot 6.91 = -47.84457$.

1. (f) If comp in Model 1 is measured in natural logarithm (and everything else remains unchanged), how would you interpret the coefficient on ln(comp)? Please explain.

When *comp* increases by 1%, predicted value of *math10* increases by 0.000005 percentage points.

1. (g) How would you predict mean math10 and compute its confidence interval for enroll = 3,000, staff = 100, and comp = 45,000 in Model 1? Carefully explain all the steps.

The predicted mean $math10 = 2.2740+0.0005\cdot 45,000+0.0479\cdot 100-0.0002\cdot 3,000 = 28.964$.

To construct the confidence interval, we also need the standard error of the predicted value. The procedure is as following:

The parameter we would like to estimate is

$$\begin{aligned} \theta_0 &= \beta_0 + \beta_1 \cdot 45000 + \beta_2 \cdot 100 + \beta_3 \cdot 3000 \\ &= E(math10|comp = 45000, staff = 100, enroll = 3000) \end{aligned}$$

Write $\beta_0 = \theta_0 - \beta_1 \cdot 45000 - \beta_2 \cdot 100 - \beta_3 \cdot 3000$ and plug this into equation

$$math10 = \beta_0 + \beta_1 comp + \beta_2 staff + \beta_3 enroll + u$$

then

$$math10 = \theta_0 + \beta_1(comp - 45000) + \beta_2(staff - 100) + \beta_3(enroll - 3000) + u$$

We run the regression of $math10$ on $(comp - 45000)$,$(staff - 100)$ and $(enroll - 3000)$. The predicted mean value (say $\hat{\theta}_0$) and its standard error (say $\hat{\delta}_{\hat{\theta}}$) are obtained from the *intercept* in the above regression. Given the critical value (say c), we could construct the confidence interval as $\left[28.964 - c\hat{\delta}_{\hat{\theta}}, 28.964 + c\hat{\delta}_{\hat{\theta}}\right]$.

Question 2.

2. (a) Compute R^2 and df for Models 1-4 in the last row.

	Model 1	Model 2	Model 3	Model 4
R^2	0.164804	0.22514	0.269134	0.349162
df	524	523	522	520

2. (b) Compare the simple and multiple regressions in Models 1 and 2 in terms of the coefficient on *educ*. That is, explain why the coefficient on *educ* in Model 2 is different from that of Model 1. Is the coefficient on *educ* in Model 1 unbiased? Explain why or why not.

The relation between the simple regression and multiple regression in terms of the coefficient on *educ* is $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2\tilde{\delta}_1$, where $\tilde{\beta}_1$ denotes the coefficient on *educ* estimated by Model 1 (simple regression), $\hat{\beta}_2$ denotes the coefficient on *educ* estimated by Model 2 (multiple regression), $\tilde{\delta}_1$ represents the slope from the simple regression *exper* on *educ*.

Before we talking about the unbiasedness of the coefficient, we have to know which model is the true model. If Model 1 is the true model, then the coefficient on *educ* in model 1 is unbiased; if Model 2 is the true model, then the coefficient on *educ* is biased unless *educ* has no effect on *exper* ($\tilde{\delta}_1 = 0$) or *exper* has no effect on *wage* ($\hat{\beta}_2 = 0$).

2. (c) How would you test the hypothesis that experience does not matter in Model 2, Model 3, and Model 4? State the null hypothesis in each case and compute the value of test statistic in each case.

Model 2:

$$\begin{aligned} wage &= \beta_0 + \beta_1 educ + \beta_2 exper + u \\ H_0: & \quad \beta_2 = 0 \\ H_1: & \quad \beta_2 \neq 0 \end{aligned}$$

$$t_{\hat{\beta}_2} = \frac{0.070}{0.011} = 6.36$$

We need to compare the t statistics with some critical value. If the absolute value of t statistics is bigger than the critical value, we reject the null, which is to say *exper* does matter in Model 2; if lower, we cannot reject the null, which implies that *exper* probably does not matter in Model 2.

Model 3:

$$\begin{aligned} wage &= \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 expersq + u \\ H_0: & \quad \beta_2 = \beta_3 = 0 \\ H_1: & \quad H_0 \text{ is not true.} \end{aligned}$$

$$\begin{aligned} F &= \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(n-k-1)} \\ &= \frac{(0.269134 - 0.164804)/2}{(1 - 0.269134)/522} \\ &= 37.257 \end{aligned}$$

We next compare the F statistics with some critical value. If bigger, then reject the null, *exper* does matter in Model 3; if lower, we cannot reject the null.

Model 4:

$$\begin{aligned} wage &= \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 expersq + \beta_4 tenure + \beta_5 exper * tenure + u \\ H_0: & \quad \beta_2 = \beta_3 = \beta_5 = 0 \\ H_1: & \quad H_0 \text{ is not true.} \end{aligned}$$

$$\begin{aligned} F &= \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(n-k-1)} \\ &= \frac{(0.349162 - R^2_r)/3}{(1 - 0.349162)/520} \end{aligned}$$

We cannot calculate the F statistics with given information. We need the R_r^2 of the regression of *wage* on *educ*, *tenure*.

2. (d) How would you test the hypothesis, using Model 2, that $\beta_1 + 7\beta_2 = 1$? Write down the trick regression and show that the above hypothesis is equivalent to testing a single regression coefficient = 0. Describe step by step how you would conduct the test, if you have access to a computer.

$$H_0: \ \beta_1 + 7\beta_2 = 1$$
$$H_1: \ \beta_1 + 7\beta_2 \neq 1$$

We introduce a new variable θ_1 here,

$$\begin{aligned} H_0 \quad &\Rightarrow \theta_1 = \beta_1 + 7\beta_2 - 1 = 0 \\ &\Rightarrow \beta_1 = \theta_1 - 7\beta_2 + 1 \end{aligned}$$

substituting into Model 2

$$wage \quad = \beta_0 + \beta_1 educ + \beta_2 exper + u$$

then

$$\begin{aligned} wage &= \beta_0 + (\theta_1 - 7\beta_2 + 1)educ + \beta_2 exper + u \\ &= \beta_0 + \theta_1 educ - 7\beta_2 educ + educ + \beta_2 exper + u \end{aligned}$$

$$\Rightarrow (wage - educ) = \beta_0 + \theta_1 educ + \beta_2(exper - 7educ) + u$$

So the equivalent null hypothesis is $H_0: \ \theta = 0$. If we have access to a computer, we will run a regression of $(wage - educ)$ on $educ$ and $(exper - 7educ)$. Using the estimated value and the standard error of coefficient on $educ$, we could construct the t statistics for $\hat{\theta}_1$. By comparing the t statistics and some critical value, we will conclude whether we reject the null.

2. (e) If you add 10 to the *educ* variable what will happen to the estimated values of the intercept and slope coefficients in Model 1?

Model 1: $\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 educnew$

$$educnew = educ + 10$$

$$\Rightarrow educ = educnew - 10$$

$$\Rightarrow \hat{wage} = \hat{\beta}_0 + \hat{\beta}_1(educnew - 10)$$

$$\Rightarrow \hat{wage} = \hat{\beta}_0 - 10\hat{\beta}_1 + \hat{\beta}_1 educnew$$

The estimated value of the intercept will be $\hat{\beta}_0 - 10\hat{\beta}_1 = -0.905 - 10 \cdot 0.541 = -6.315$; the estimated value of the slope coefficient will be unchanged.

What will happen to the estimated values of the intercept and slope coefficients in Model 1 if the education variable is divided by 5?

$$educnew = \frac{educ}{5}$$

$$\Rightarrow educ = 5educnew$$

$$\Rightarrow \hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 5educnew$$

$$\Rightarrow \hat{wage} = \hat{\beta}_0 + 5\hat{\beta}_1 educnew$$

The estimated value of the intercept will be unchanged; the estimated value of the slope coefficient will be $5 \cdot \hat{\beta}_1 = 5 \cdot 0.541 = 2.705$.

2. (f) Compute the marginal effect of *exper* in Model 4 and explain it behavior in terms of experience and tenure.

Model 4: $\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 educ + \hat{\beta}_2 exper + \hat{\beta}_3 expersq + \hat{\beta}_4 tenure + \hat{\beta}_5 exper * tenure$

The marginal effect of *exper* is

$$\begin{aligned} \frac{\partial \hat{wage}}{\partial exper} &= \hat{\beta}_2 + 2\hat{\beta}_3 exper + \hat{\beta}_5 tenure \\ &= 0.175 - 2 \cdot 0.003 exper - 0.004 tenure \\ &= 0.175 - 0.006 exper - 0.004 tenure \end{aligned}$$

The marginal effect of *exper* will decrease as *exper* increases holding *tenure* constant. The marginal effect of *exper* will decrease as *tenure* increases holding *exper* constant.

When $exper = 20$, $tenure = 5$ and $educ = 15$, the marginal effect of tenure is $0.175 - 0.006 \cdot 20 - 0.004 \cdot 5 = 0.035$.

2. (g) Compute the marginal effect of *tenure* in Model 4 and explain it behavior in terms of experience and tenure.

Model 4: $\hat{wage} = \hat{\beta}_0 + \hat{\beta}_1 educ + \hat{\beta}_2 exper + \hat{\beta}_3 expersq + \hat{\beta}_4 tenure + \hat{\beta}_5 exper * tenure$
The marginal effect of *tenure* is

$$\begin{aligned} \frac{\partial \hat{wage}}{\partial tenure} &= \hat{\beta}_4 + \hat{\beta}_5 exper \\ &= 0.291 - 0.004 exper \end{aligned}$$

The marginal effect of *tenure* will decrease as *exper* increases; while the marginal effect of *tenure* will not be affected by the value of *tenure*. The return to experience will be maximized where $tenure = 72.75$.

When $exper = 20$, $tenure = 5$ and $educ = 15$, the marginal effect of tenure is $0.291 - 0.004 \cdot 20 = 0.211$.

2. (h) Find the level of *exper* in Model 3 at which the predicted wage is maximized. How can you tell that it is maximized and not minimized at that point?

Model 3: $\hat{wage} = -3.96 + 0.595 educ + 0.268 exper - 0.005 expersq$
To get the level at which the predicted *wage* is maximized, we need to find the first order condition:

$$\frac{\partial \hat{wage}}{\partial exper} = 0.268 - 0.01 exper = 0$$

$$\Rightarrow exper = 26.8$$

To check whether it is a maximum or minimum, we need to check the second order condition:

$$\frac{\partial^2 \hat{wage}}{\partial exper^2} = -0.01 < 0$$

So it is a maximum.

Department of Economics