

**ECON 3035: Econometric Methods, Spring 2019**  
**Midterm One**

**Answer the following questions briefly.**

**Problem One**

1a) State assumption necessary to test the significance (inference) of Ordinary Least Square regression coefficients. Logically, state the why the assumption allows us to make inferences.

1b) Given a regression  $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u$ , sketch that  $\hat{\beta}_1$  is biased estimator of  $\beta_1$  if  $x_{1i}$  and  $x_{2i}$  are strongly correlated. (Hint: Directly, guess the formula for  $\hat{\beta}_1$  and show that it is biased.)

### Problem Two

Consider the multiple regression model with three independent variables, under the six classical linear model assumptions,  $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + u$ .

Specifically, let  $y$  be measure of how healthy a person is (weekly reported),  $x_1$  is number of packs of cigarettes a person smokes per week,  $x_2$  is the number of hours a person exercises per week, and  $x_3$  is individual's income per week.

2a) What are the expected signs for  $\beta_1, \beta_2, \beta_3$ ?

2b) You would like to test the null hypothesis  $H_0 : \beta_1 + \beta_2 = 0$ . Explain why does this test make sense. Let  $\hat{\beta}_1$  and  $\hat{\beta}_2$  denote the OLS estimators of  $\beta_1$  and  $\beta_2$ . Find  $Var(\hat{\beta}_1 + \hat{\beta}_2)$  and  $se(\hat{\beta}_1 + \hat{\beta}_2)$ .

2c) Write the t statistic for testing  $H_0 : \beta_1 + \beta_2 = 0$

2d) Write the regression equation that allows you to directly estimate and make inferences about your hypothesis in 2b).

### **Problem Three**

The following model can be used to study whether campaign expenditures affect election outcome:

$$voteA = \beta_0 + \beta_1 \log(expendA) + \beta_2 \log(expendB) + \beta_3 prtystA + u$$

where  $voteA$  is the percentage of the vote received by Candidate A,  $expendA$  and  $expendB$  are campaign expenditures by Candidate A and B (respectively), and  $prtystA$  is a measure of party strength for Candidate A (the percentage of the most recent presidential vote that went to A's party).

3a) What is the interpretation of  $\beta_1$ ?

3b) In terms of the parameters, state the null hypothesis that a 1% increase in A's expenditure is offset by a 1% increase in B's expenditure.

3c) The estimated model is given by the following equation:

$$\widehat{voteA} = 45.08 + 6.083 \log(expendA) - 6.615 \log(expendB) + .152 \text{prtystrA}$$

(3.93) (0.382)                      (0.379)                      (0.062)

$$n = 173, R^2 = .793.$$

Based on the result, do A's expenditures affect the outcome? What about B's expenditures? Can we directly use these results to test the null hypothesis in 3b)

3d) Show how you would test the null hypothesis stated in 3b) by explicitly writing out the regression you would estimate.

### Problem Four

Suppose you are interested in the determinants of cigarette smoking to provide some public health policy suggestion. You have data on *cigs* (cigarette smoked per day), *educ* (years of schooling), *lcigpric* (log(cigarette price)), *age* (in years), *lincome* (log(income)) for 807 individuals. You estimated an OLS regression and found the following result.

```
. regress cigs educ lcigpric age lincome
```

Source	SS	df	MS			
Model	<b>1937.3501</b>	<b>4</b>	<b>484.337526</b>	Number of obs =	<b>807</b>	
Residual	<b>149816.333</b>	<b>802</b>	<b>186.803407</b>	F( 4, 802) =	<b>2.59</b>	
Total	<b>151753.683</b>	<b>806</b>	<b>188.280003</b>	Prob > F =	<b>0.0354</b>	
				R-squared =	<b>0.0128</b>	
				Adj R-squared =	<b>0.0078</b>	
				Root MSE =	<b>13.668</b>	

  

cigs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	<b>-.3888199</b>	<b>.168512</b>	<b>-2.31</b>	<b>0.021</b>	<b>-.7195966</b>	<b>-.0580432</b>
lcigpric	<b>-2.912104</b>	<b>5.825768</b>	<b>-0.50</b>	<b>0.617</b>	<b>-14.34766</b>	<b>8.523449</b>
age	<b>-.042654</b>	<b>.0287616</b>	<b>-1.48</b>	<b>0.138</b>	<b>-.0991108</b>	<b>.0138028</b>
lincome	<b>1.777385</b>	<b>.713231</b>	<b>2.49</b>	<b>0.013</b>	<b>.3773655</b>	<b>3.177405</b>
_cons	<b>10.00436</b>	<b>24.26632</b>	<b>0.41</b>	<b>0.680</b>	<b>-37.62864</b>	<b>57.63736</b>

4a) Interpret the coefficients on the independent variables. Comment on the demand function of cigarettes (cigarette smoked per day) as a function of *lcigpric* (log(cigarette price)) taking in to account the significance of the coefficient.

4b) Suppose you regressed *cigs* (cigarette smoked per day) on *educ*(years of schooling) and *lincome* ( $\log(\text{income})$ ) and found the following result with the sum of squared residuals (SSR) equal to 150282.168.

<i>cigs</i>	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
<i>educ</i>	<b>-.3454854</b>	<b>.1658978</b>	<b>-2.08</b>	<b>0.038</b>	<b>-.6711293</b>	<b>-.0198416</b>
<i>lincome</i>	<b>1.735492</b>	<b>.7116314</b>	<b>2.44</b>	<b>0.015</b>	<b>.3386177</b>	<b>3.132367</b>
<i>_cons</i>	<b>-3.817261</b>	<b>6.564067</b>	<b>-0.58</b>	<b>0.561</b>	<b>-16.70199</b>	<b>9.067471</b>

Decide whether *age* (in years) and *lincome* ( $\log(\text{income})$ ) belong in regression 4a). What is the  $R^2$  of this regression (regression in 4b)?

4c) Suppose you regressed *cigs* (cigarette smoked per day) on *lcigpric* ( $\log(\text{cigarette price})$ ) and *lincome* ( $\log(\text{income})$ ) and found the following result with  $R^2$  equal to 0.0047.

cigs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lcigpric	-3.276784	5.839439	-0.56	0.575	-14.73913	8.185561
lincome	1.299833	.6793971	1.91	0.056	-.0337682	2.633435
_cons	9.516411	24.31977	0.39	0.696	-38.22132	57.25414

Comparing with your result in i), how do you decide whether educ and age belong in the regression. (the corresponding p-value for what you calculate is  $p = 0.04$ )

4d) Can you comment on a policy, if any? [ I want to see your argument– there is no right or wrong answer]