CPP 523: Foundations of Eval I
Regression Specification

Prof. Jesse Lecy

PRACTICE EXAM

NAME____

Instructions: You have four hours to complete the exam once it is started. You can use notes, calculators, and statistical software. You are NOT allowed to work with anyone else, or share questions and solutions with others. Good luck!

Please give non-mathematical definitions to the following statistical concepts:

(1) The Standard Error:

(2) The 95% confidence interval of a slope:

(3) R-Squared:

(4) Name three things that will reduce the standard error of a regression slope.

(5) Name two sins of the Seven Sins that will <u>always</u> increase the standard error of a regression slope.

(6) Control variables that are uncorrelated with our policy variable will not cause omitted variable bias if we do not include them in a regression. Why do we include them in the model? For example, Teacher Quality in the Classroom Size example from class.

(7) Calculate the slope and the intercept for a simple bivariate regression model ($Y = b_0 + b_1 X + e$) from the following information:

$\bar{x}$:	4
$\bar{y}$:	11
var(x):	3
var(y):	7
cov(x,y):	6

b_1 =

b_0 =

(8) Now using the slope and intercept that you calculated above, calculate the residual (prediction error – column ***e***) for the following three cases.

X	**Y**	$\hat{Y}$	e
5	15		
6	11		
8	21		

(9) Consider the following regression:

$$Y = \beta_0 + \beta_1 X1 + \varepsilon$$

Consider the case where $B_1 = 6$, and $SE_{B1} = 2.61$.

Using t=2.58, calculate the 99% confidence interval for B_1. Is the slope statistically significant at this level? How do we know?

(10) Consider the following regression results.

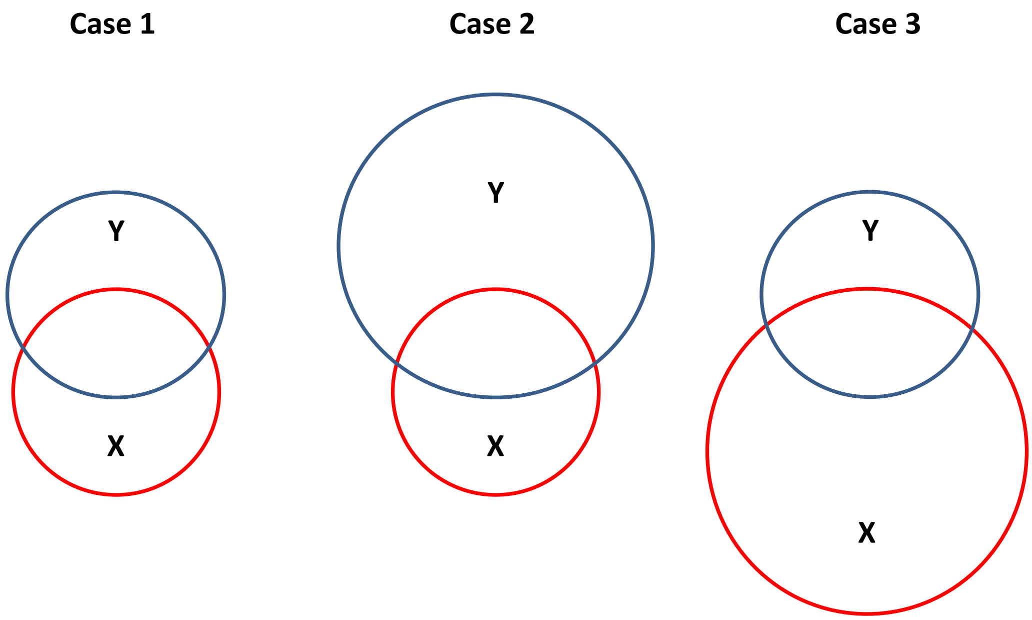
Table VI Multiple regression results on channel satisfaction

Independent variables	Beta	*T*-statistic	Significance
Constant	0	−0.76	0.4540
Relative performance	0.75	5.34	0.0001*
Experience	0.15	1.04	0.3063
Control	0.44	4.98	0.0560**
Changeability	0.14	0.80	0.4319
Uncertainty	0.04	0.31	0.7592
Monitoring	0.12	0.78	0.4406
Intermediate mode	0.21	2.58	0.5638
Hierarchy mode	−0.02	−0.06	0.9511

Notes: $R^2 = 0.52$; $F = 18.23$; $n = 45$; * = $P \leq 0.01$ (1 tailed *t*-test); ** = $P \leq 0.10$ (1 tailed *t*-test)

What is the largest level of confidence you can chose for the confidence interval around the slope estimate for the **Experience** coefficient before it crosses zero?

(11) Consider the following cases:



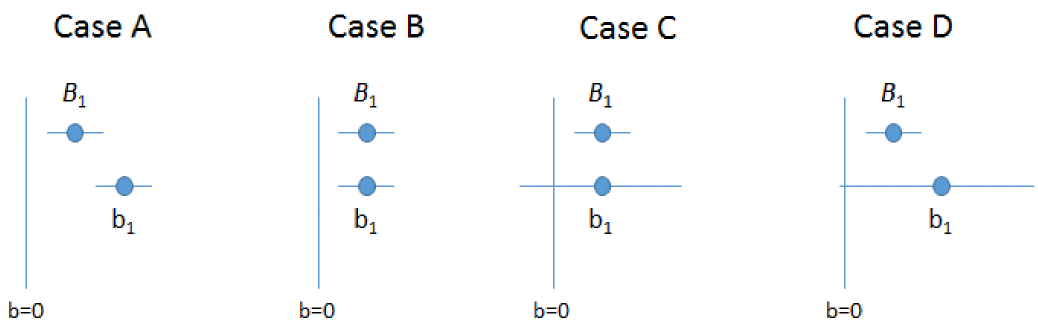
a. Holding *cov(x,y)* constant across all cases, which case(s) will have the <u>largest</u> standard error?

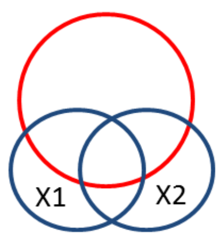
b. Holding *cov(x,y)* constant across all cases, which case(s) will have the <u>smallest</u> slope?

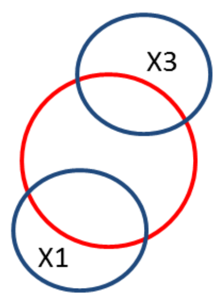
(12) Consider four cases below. The full regression in this case is with X1 being the policy variable:

$$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + e$$

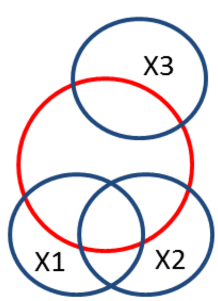
Write the correct case letter under each Venn diagram.

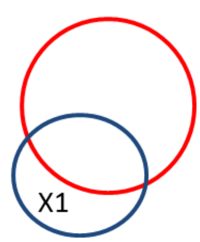






__________ __________





__________ __________

BONUS (3pts): Go back to the model that attempts to discern the effects of class size on test scores:

DV:Test-Scores	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	141.675***	-179.058***	141.479***	141.553***	-179.725***
	(4.571)	(1.609)	(4.601)	(4.596)	(1.575)
Class Size	-0.433***	-0.468***		-0.377	-0.267***
	(0.021)	(0.003)		(0.207)	(0.029)
Quality of Instruction		62.125***			62.169***
		(0.285)			(0.278)
Socio-Economic Status			43.420***	5.649	20.344***
			(2.076)	(20.853)	(2.918)
R-squared	0.307	0.986	0.305	0.307	0.986
N	1000	1000	1000	1000	1000

Now think about another model:

$$SES = \pi_0 + \pi_1 ClassSize + e$$

What is the exact slope for the regression of **SES** on **Class Size,** π_1? Show your math.

BONUS (4 pts): Think back to the model that we have studied looking at the relationship between classroom size and test scores:

$$TestScore = \beta_0 + \beta_1 ClassSize + \beta_2 SES + \beta_3 TeacherQuality + \varepsilon \qquad (1)$$

Now think about a different way to run the regression model. What if we constructed it in the following way:

$$ClassSize = \pi_0 + \pi_1 SES + e_1 \qquad (2)$$

$$TestScore = b_0 + b_1 e_1 + b_2 SES + b_3 TeacherQuality + \gamma \qquad (3)$$

In this case the e_1 in model (3) is the residual term from model (2). Using a Venn diagram to justify your response, answer the following questions:

Does $b_1 = \beta_1$?

Does $b_2 = \beta_2$?

Does $b_3 = \beta_3$?

Does $\varepsilon = \gamma$?