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:**

1. **R-Squared:**
2. Name three things that will reduce the standard error of a regression slope.

1. Name two sins of the Seven Sins that will always increase the standard error of a regression slope.
2. 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.
3. Calculate the slope and the intercept for a simple bivariate regression model (  ) from the following information:

: 4

: 11

*var(x):* 3

*var(y):* 7

*cov(x,y):* 6

b1 =

b0 =

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

|  |  |  |  |
| --- | --- | --- | --- |
| **X** | **Y** |  | e |
| 5 | 15 |  |  |
| 6 | 11 |  |  |
| 8 | 21 |  |  |

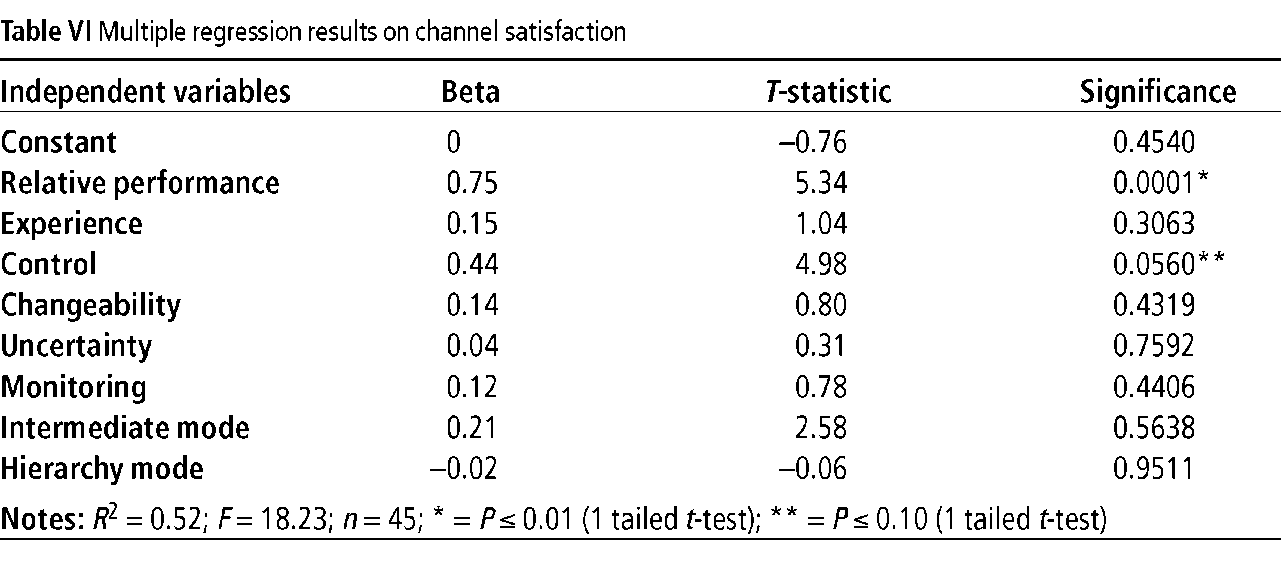
1. Consider the following regression:



Consider the case where *B1* = 6, and *SEB1* = 2.61.

Using t=2.58, calculate the 99% confidence interval for *B1.* Is the slope statistically significant at this level? How do we know?

1. Consider the following regression results.



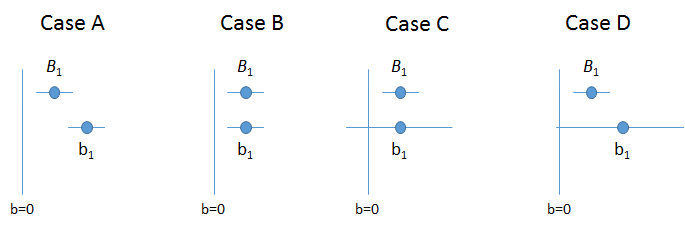
What is the largest level of confidence you can chose for the confidence interval around the slope estimate for the **Experience** coefficient before it crosses zero?

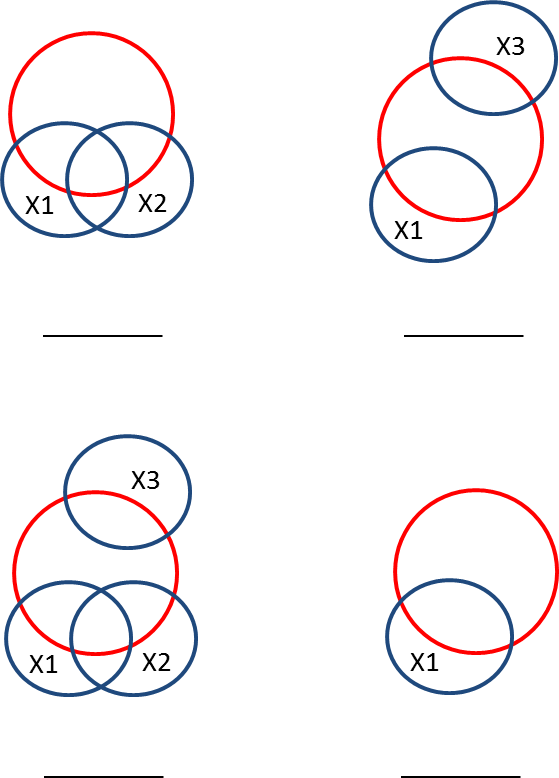
1. Consider the following cases:



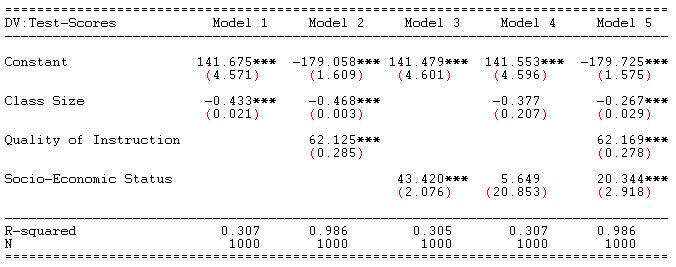
* 1. Holding *cov(x,y)* constant across all cases, which case(s) will have the largest standard error?
  2. Holding *cov(x,y)* constant across all cases, which case(s) will have the smallest slope?

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



Now think about another model:



What is the exact slope for the regression of **SES** on **Class Size,** π1? 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:

 (1)

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

 (2)

 (3)

In this case the e1 in model (3) is the residual term from model (2). Using a Venn diagram to justify your response, answer the following questions:

Does ?

Does ?

Does ?

Does ?