EMLE
Exam
Klick
2019

This exam is open book, open notes (i.e., you may use your own materials), but not open internet (i.e., you may not search for answers on the internet). Each question is worth an equal number of potential points. Good luck.

1. What conditions must hold for omitted variable bias to be a concern?

The regression omits an x or many x variables that 1) should be in the model (i.e., affect the outcome y) and 2) are correlated with an x or many x's that are included in the model.

2. What does the term "statistically significant" mean?

When testing a hypothesis (e.g., that the effect of some control variable is 0), statistically significant implies that the observed coefficient/correlation/difference/etc would be observed less than p% of the time (where p is the chosen type 1 error) if the hypothesis were correct in the population.

3. What does it mean to say that an empirical result has "external validity"?

External validity is related to generalizability. That is, assuming that the research conclusion is correct, how well does it generalize to a particular policy/legal setting.

4. In an event study, if multiple changes are occurring on the given event date, what problem does this create with respect to inferring the causal effect of the event studied?

The regression cannot, in general, determine whether the event effect is due to change 1, change 2, …, change n, or some combination of the changes.

5. When doing a cost benefit analysis (or impact assessment) of a proposed regulation, how does one account for any effects of the regulation on lives saved or lives lost as a result of the regulation?

The value of a statistical life is calculated statistically by comparing, for example, wages paid in jobs that differ only (or at least the other differences can be accounted for) in terms of their underlying mortality risk. These estimates can then be used in the cost benefit analysis.

Questions 6-10 refer to the following regression and resulting output:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_1x_2$$

6. If your computer program indicates that $b_3 = -100$ and the associated standard error is 75, interpret the b_3 coefficient

This coefficient suggests that, all other things equal, when x3 increases by one unit, y declines by 100 units on average.

7. Is the coefficient described in question 6 statistically significant at the 5% type 1 error level?

The 5% type 1 error corresponds with a decision rule of rejecting the null hypothesis (i.e., declaring something statistically significant) if the estimate is more than 2 standard errors away from the null. Since 100 is not more than 2 X 75 (i.e., the standard error), we would not normally refer to this as statistically significant.

8. Assume x_1 is an indicator variable taking the value of 1 if the observation in the data is female and x_2 is a continuous age variable, what does the coefficient b_4 represent?

b4 would indicate, on average, how much does a woman's y value change when age goes up by 1 unit over an above any age effect observed for men.

9. If the above regression were re-run just on female observations (noting what the x_1 variable means according to question 8 above), what is your best guess as to what the new *a* parameter would be?

Since the a parameter is our overall intercept, the b1 parameter tells us how different are women than the average man. Thus, if we only were to look at women, the new overall intercept would likely become a + b1, since the overall average would now just be the average for women.

10. If you thought the effect of age in the regression above was non-linear, how would you incorporate that in the model?

There are a number of ways this could be incorporated. One could include a set of individual age dummy variables (x20 = 1 if the person is 20; 0 otherwise; and so on) or one could include a polynomial in age (age, age*age, age*age*age, etc).