POLS 6481, Spring 2021

Professor Scott Basinger

Reading Assignment Week 5

Distributed Friday, February 12

Due Thursday, February 18

Required reading: Wooldridge 2.3 + 2.4a + 3.2f + 4.5 + 6.1 + 6.3

Iacobucci, *Mediation Analysis*, 1–23

1. Suppose you have two explanatory variables, *x*1 and *x*2, and you wish to “partial out” the effects of *x*1 and *x*2, using in place of *x*1. More specifically, let *x*1 be the *number of clubs* someone participated in during high school, and let *x*2 be *adolescent height*. Write the precise steps you would follow to explore how the *number of clubs* – after having “netted out”the effect of *adolescent height* – affects *wages*.

2. Suppose you followed the steps listed in 1….

A. How would (the coefficient from a simple regression of *wages* on *number of clubs*) relate to (the coefficient on *number of clubs* from the multiple regression of *wages* on *adolescent height* and )? Explain your responses or provide a demonstration using *R*.

B. How would (the coefficient from a simple regression of *wages* on *adolescent height*) relate to (the coefficient on *adolescent height* from the multiple regression of *wages* on *adolescent height* and )? Explain your responses or provide a demonstration using *R*.

3. Take the example in 1. and 2.; write out a mediation model similar to Iacobucci’s p. 11, in which *number of clubs* mediates the effect of *adolescent height* on *wages*. What would be in the numerator of the *z*-statistic on Iacobucci’s p. 12?

Using the MLB1 dataset, Wooldridge shows regression models in which *log(salary)* is the dependent variable, and the independent variables are *years*, *gamesyr*, *bavg*, *hrunsyr*, and *rbisyr*. He performed a second analysis that drops the last three of these variables from the regression model. The results are:

|  |  |
| --- | --- |
|  |  |

4. Calculate the SST (sum of squared deviations in *log*(*salary*)based on either model’s *R*2 and SSR. (This may require a little algebra, but don’t be intimidated ☺ )

5. The results shown above display *R*2 but not the adjusted *R*2. Calculate the adjusted *R2* for each model.

6A. Suppose you wish to perform an F test of the joint significance of the variables *bavg*, *hrunsyr*, & *rbisyr*. State the null hypothesis that you would testing, and the appropriate alternative hypothesis.

H0 :

H1 :

6B. State the numerator degrees of freedom for the F test. Explain your answer.

6C. State the denominator degrees of freedom for the F test. Explain your answer.

7. As you can see, the R2 from the model that uses all five variables equals 0.6728, and the R2 from the model with just *years* and *gamesyr* equals 0.5971. Answer the following:

(i) What is the value of *q* (using Wooldridge’s notation)?

(ii) Calculate the F-statistic for exclusion of these variables.

(iii) What is the critical value of the F distribution at 5% significance?

(iv) Do we reject or retain the null hypothesis stated in 6A.?

8. Perform an F test of the regression model shown on the right which has just *years* and *gamesyr*.

Be sure to state the value of *k* = \_\_\_\_ , the value of *n – k –* 1 = \_\_\_\_\_ , the value of *F* = \_\_\_\_\_ , and the critical value based on ( *k* , *n – k –* 1 ) degrees of freedom = \_\_\_\_\_ .