Introduction to Computational Statistics INSH 5301

Homework 10

03/23/2020

For this homework you'll need to use some real world data to answer some research questions using multiple regression. The data, along with the data description, can be downloaded from the course material section in Blackboard. You can also download this dataset using the AER package.

AER package: Just

Run 'library(AER)' first, and then 'data("CollegeDistance")'

> College Distance (Re-Re-Revisited)
> For this problem we are going to explore the effect of distance from college on educational attainment. The dependent variable (Y) is years of completed education ed. Run all the regressions using the lm command and display regression tables using the stargazer command.

> 1. Run a regression of ed on dist, female, bytest, tuition, black, hispanic, incomehi, ownhome, dadcoll, momcoll, cue80, and, Stwmfg80. If dist increase from 2 to 3, how are years of education expected to change? If dist increases from 6 to 7, how are years of education expected to change?

> 2. Run a regression of the natural logarithm of ed on dist, female, bytest, tuition, black, hispanic, incomehi, ownhome, dadcoll, momcoll, cue80, and, Stwmfg80. If dist increase from 2 to 3, how are years of education expected to change? If dist increases from 6 to 7, how are years of education expected to change?

> 3. Run a regression of the natural log ed on dist, dist^2, female, bytest, tuition, black, hispanic, incomehi, ownhome, dadcoll, momcoll, cue80, and, Stwmfg80. If dist increase from 2 to 3, how are years of education expected to change? If dist increases from 6 to 7, how are years of education expected to change?

> 4. Compare the results of (2), and, (3) to (1). Do you think (1) is a correct specification? (Explain)

> 5. Add the interaction term dadcoll*momcoll to the regression (3). Interpret the coefficient of the interaction term.

> 6. Test if the effect of distance on education depends on the family's income.

> 7. Use the F-test on (5) as the complete model and (1) as the nested model. What specification do you prefer? (Explain)sK