Introduction to Computational Statistics INSH 5301

Homework 09

03/16/2020

PLEASE copy and paste the whole question number and text into submission so I can grade easily.
When I grade easier, you might get better grade!
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.

Problem 1: College Distance (Revisited)
1. 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 any regression using the lm command and display regression tables using the stargazer command.
1.1 Test the hypothesis that college distance is related with educational attainment using a bivariate regression model.
1.2 A regression will suffer from omitted variable bias when two conditions hold. What are these two conditions? Do these conditions seem to hold in this case?
1.3 Consider the various control variables in the dataset. Which do you think should be included in the regression? (Explain)
1.4 Estimate a set of multivariate regression models. Add one additional variable at a time – that is, each new regression should have at most one additional variable than the previous one –. Display all regressions (including (1.1)) in a single table using stargazer. At this stage only consider OV B when deciding to add more variables. Compare the results of the multivariate regressions with the bivariate model.
1.5 Check if the estimated model sufferes from multicollinearity.

2. Blood pressure
For this problem we are going to explore the effect of weight on blood pressure. The dependent variable (Y) is blood pressure bp. Run any regression using the lm command and display regression tables using the stargazer command.
2.1 Test the hypothesis that weight is related with blood presure using a bivariate model.
2.2 Use the other variables in the model to correct for OV B.
2.3 Test for multicollinearity
2.4 Choose a final functional form and interpret the results.