Exercise session 9

Problem 1. In 1985, neither Florida nor Georgia had laws banning open alcohol containers in vehicle passenger compartments. By 1990, Florida had passed such a law, but Georgia had not.

(i) Suppose you can collect random samples of the driving-age population in both states, for 1985 and 1990. Let arrest be a binary variable equal to unity if a person was arrested for drunk driving during the year. Without controlling for any other factors, write down a linear probability model that allows you to test whether the open container law reduced the probability of being arrested for drunk driving. Which coefficient in your model measures the effect of the law?

(ii) Why might you want to control for other factors in the model? What might some of these factors be?

Problem 2. Suppose you want to test whether girls who attend a girls' high school do better in math than girls who attend coed schools. You have a random sample of senior high school girls from a state in the United States, and score is the score on a standardized math test. Let girlhs be a dummy variable indicating whether a student attends a girls' high school.

(i) What other factors would you control for in the equation? (You should be able to reasonably collect data on these factors.)

(ii) Write an equation relating score to girlhs and the other factors you listed in part (i).

$$MathSC = \gamma_0 + \gamma_1 girlhs + \gamma_2 faminc + \gamma_3 meduc + \gamma_4 feduc + \varepsilon$$

(iii) Suppose that parental support and motivation are unmeasured factors in the error term in part (ii). Are these likely to be correlated with girlhs? Explain.

(iv) Discuss the assumptions needed for the number of girls' high schools within a 20-mile radius of a girl's home to be a valid IV for girlhs.

Problem 3. Let grad be a dummy variable for whether a student-athlete at a large university graduates in five years. Let *hsGPA* and *SAT* be high school grade point average and SAT score, respectively. Let *study* be the number of hours spent per week in an organized study hall. Suppose that, using data on 420 student-athletes, the following logit model is obtained:

$$\hat{P}(grad = 1 | hsGPA, SAT, study) = \Lambda(-1.77 + .24hsGPA + 0.00058SAT + 0.073study),$$

Where, $\Lambda(\text{z}) = \frac{\exp(\text{z})}{1+\exp(\text{z})}$ is the logit function. Holding hsGPA fixed at 3.0 and SAT fixed at 1,200, compute the estimated difference in the graduation probability for someone who spent 10 hours per week in study hall and someone who spent 5 hours per week.