GPA Exercise (adapted from exercises in Wooldridge, "Introductory Econometrics")

This exercise will be covered during sections on Monday 11th and Wednesday 13th. Sections will be held at the main StatLab classroom at the CSSSI - same location as past Stata tutorial. Please attend your assigned section time, as we won't have enough computer terminals for everyone otherwise. These sections are required.

1. The data in gpa2.dta contain information on 4,137 college students. The variable colgpa is the cumulative college grade point average, measured on a four-point scale, hsperc is the percentile in the high school graduating class (defined so that, for example, hsperc = 5 means the top five percent of the class), and sat is the combined math and verbal scores on the student achievement test. Using Stata, estimate an equation that relates hsperc and sat to colgpa.

   (a) Does it make sense for the coefficient on hsperc to be negative?

   (b) What is the predicted college GPA when hsperc=20 and sat=1050?

   (c) Suppose that two high school graduates, A and B, graduated in the same percentile from high school, but Student A's SAT score was 140 points higher (about one standard deviation in the sample). What is the predicted difference in college GPA for these two students? Is the difference large?

   (d) Holding hsperc fixed, what difference in SAT scores leads to a predicted colgpa difference of .50, or one-half of a grade point? Comment on your answer.

2. In the same data, the variable hsize is size of the student's high school graduating class, in hundreds, female is a gender dummy variable, and black is a race dummy variable equal to one for blacks, and zero otherwise.

   (a) Using Stata, estimate the following equation:

   $$sat = \alpha + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 female + \beta_4 black + \beta_5 female \cdot black + u$$

   (b) Is there strong evidence that hsize2 should be included in the model?

   (c) Holding hsize fixed, what is the estimated difference in SAT score between nonblack females and nonblack males? How statistically significant is this estimated difference?

   (d) What is the estimated difference in SAT score between nonblack males and black males? Test the null hypothesis that there is no difference between their scores, against the alternative that there is a difference.

   (e) What is the estimated difference in SAT score between black females and nonblack females? What would you need to do to test whether the difference is statistically significant?

3. Consider now the equation:

   $$colgpa = \alpha + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 hsperc + \beta_4 sat + \beta_5 female + \beta_6 athlete + u$$

   where athlete is a binary variable, which is one for student-athletes.

   (a) What are your expectations for the coefficients in this equation? Which ones are you unsure about?

   (b) Estimate the equation above and report the results. What is the estimated GPA differential between athletes and nonathletes? Is it statistically significant?

(c) Drop sat from the model and reestimate the equation. Now what is the estimated effect of being an athlete? Discuss why the estimate is different than that obtained in part (b).

(d) In the model from part (a), allow the effect of being an athlete to differ by gender and test the null hypothesis that there is no ceteris paribus difference between women athletes and women nonathletes.

(e) Does the effect of sat on colgpa differ by gender? Justify your answer.