C4. Use the data in GPA2 for this exercise.

Natchaya Laokham
6104641425

i. Consider the equation

$$colgpa = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 hsperc + \beta_4 sat + \beta_5 female + \beta_6 athlete + u,$$

where *colgpa* is cumulative college grade point average; *hsize* is size of high school graduating class, in hundreds; *hsperc* is academic percentile in graduating class; *sat* is combined SAT score; *female* is a binary gender variable; and *athlete* is a binary variable, which is one for student-athletes. What are your expectations for the coefficients in this equation? Which ones are you unsure about?

The two signs that are pretty clear are $\beta_3 < 0$ (because hsperc is defined so that the smaller the number the better the student) and $\beta_4 > 0$. The effect of size of graduating class is not clear. It is also unclear whether males and females have systematically different GPAs. We may think that $\beta_6 < 0$, that is, athlets do worse than other students with comparable characteristics. But remember, we are controlling for ability to some degree with hsperc and sat.

ii. Estimate the equation in part (i) and report the results in the usual form. What is the estimated GPA differential between athletes and nonathletes? Is it statistically significant?

The estimated equation is

$$\widehat{colgpa} = \underset{(0.079)}{1.241} - \underset{(.0164)}{.0569}\,hsize + \underset{(.00225)}{.00468}\,hsize^2 - \underset{(.0006)}{.0132}\,hsperc + \underset{(.00007)}{.00165}\,sat + \underset{(.018)}{.155}\,female + \underset{(.042)}{.169}\,athlete$$

$n = 4,137 \quad R^2 = .293$

Holding other factors fixed, an athlete is predicted to have a GPA about .169 points higher than a nonathlete.

The t statistic $\frac{.169}{.042} \approx 4.02$, which is very significant

iii. 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 (ii).

With sat dropped from the model, the coefficient on athlete becomes about .0054 (se ≈ .0448), which is practically not different from zero. this happens because we do not control for SAT scores, and athletes score lower on average than nonathletes. Part (ii) shows that, once we account for SAT differences, athletes do better than nonathletes. Even if we do not control for SAT score, there is no difference.

iv. In the model from part (i), 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.

We should choose one of these as the base group. We choose female nonathletes. The estimated equation is ...

$$\widehat{colgpa} = \underset{(.076)}{1.39} - \underset{(.0164)}{.0568}\,hsize + \underset{(.00225)}{.00467}\,hsize^2 - \underset{(.0006)}{.0132}\,hsperc$$

$$+ \underset{(.00007)}{.00165}\,sat + \underset{(.084)}{.175}\,femath + \underset{(.049)}{.013}\,maleath - \underset{(.018)}{.155}\,malenonath$$

$n = 4{,}137 \quad R^2 = .293$

The coefficient on femath = female · athlete shows that colgpa is predicted to be about .175 points higher for a female athlete than a female nonathlete, other variables in the equation fixed. The hypothesis that there is no difference between female athletes and female nonathetes is testing by using the t statistic on femath. In this case, t = 2.08, which is staticticalLy significant at the 5% level against a two-sided alternative.

v. Does the effect of *sat* on *colgpa* differ by gender? Justify your answer.

Whether we add the interaction female · sat to the equation in part (ii) or part (iv), the outcome is practically the same. For example, when female·sat is added to the equation in part (ii), its coefficient is about .000051 and its t statistic is about .40. There is very little evidence that the effect of sat differs by gender.