

## GPA Exercise - Solutions

Please refer to the do file *gpa\_exercise.do* on *classesv2* for how to use Stata to solve this exercise.

1. We estimate the regression:

$$colgpa = \alpha + \beta_1 hspc + \beta_2 sat + u$$

- (a) It make sense for the coefficient on *hspc* to be negative since lower values of *hspc* mean higher percentiles in the high school class. Higher percentiles students are the ones that we expect to have higher GPA in college.
- (b)  $col\hat{gpa} = \hat{\alpha} + \hat{\beta}_1 \cdot 20 + \hat{\beta}_2 \cdot 1050 = 1.391757 - 0.0135192 \cdot 20 + 0.0014762 \cdot 1050$
- (c) The predicted difference in college GPA for these two students is  $\hat{\beta}_2 \cdot 140 = 0.0014762 \cdot 140 = 0.206668$ . The difference is quite small.
- (d) The difference in SAT scores that leads to a predicted *colgpa* difference of .50 is the  $x$  such that  $\hat{\beta}_2 \cdot x = 0.5$ . We find that  $x = 338.7$ . This is a very large difference in SAT, equivalent to approximately 2.5 standard deviations of SAT.

2. We obtain estimates for the new model.

- (a) The equation estimated in Stata is:

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

- (b) The coefficient on  $hsize^2$  is statistically significant at any standard confidence level. This means that a linear relation between *hsize* and *sat* does not describe the data well, and that instead a model which also includes the variable  $hsize^2$ , meant to capture nonlinearities, is more appropriate. The estimates tell us that the size of the high school graduating class positively affects SAT score for small high school size, but does so at a decreasing rate and eventually becomes negative.
- (c) The estimated difference in SAT score between nonblack females and nonblack males is  $\hat{\beta}_3 = -45.09145$ . The difference is statistically significant at any standard confidence level.
- (d) The estimated difference in SAT score between nonblack males and black males is  $\hat{\beta}_4 = -169.8126$ . The p-value that Stata reports for  $\hat{\beta}_4$  tests exactly the null hypothesis that there is no difference between nonblack males and black males, against the alternative that there is a difference. Since the p-value is very small, the difference between their scores is statistically significant at any standard confidence level.
- (e) The estimated difference in SAT score between black females and nonblack females is given by  $\hat{\beta}_4 + \hat{\beta}_5$ . If we wanted to test whether the difference is statistically significant, the we would estimate the specification

$$sat = a + b_1 hsize + b_2 hsize^2 + b_3 male + b_4 black + b_5 male \cdot black + u$$

and test the null hypothesis  $H_0 : b_4 = 0$ .

3. Consider now the equation:

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

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

- (a) Based on the simple regression model estimated in part 1, we expect the coefficients on the variables `hspc`, `sat` and `female` to have the same signs as before. We can also build some expectations about the coefficients on `hsize`,  $hsize^2$ , `female` from the SAT model estimated in part 2. We are unsure about what to expect for athletes.
- (b) The estimated GPA differential between athletes and non-athletes is  $\hat{\beta}_6 = 0.1693064$ . It is statistically significant at any standard confidence level.
- (c) When we drop `sat` from the model and reestimate the equation, the estimated effect of being an athlete is not statistically significant. This suggests that there may be a negative correlation between SAT score and being an athlete.
- (d) If we wanted to allow the effect of being an athlete to differ by gender, we would introduce in the specification, together with the dummies for `female` and `athlete`, the interaction between `female` and `athlete`. When we estimate such a model, the interaction term turns out not to be statistically significant, suggesting that the effect of being an athlete does not differ by gender.
- (e) We introduce in the specification, together with the variables `female` and SAT score, the interaction between `female` and SAT score. The interaction term is not statistically significant, suggesting that the effect of `sat` on `colgpa` does not differ by gender.