Parte 1: Variáveis Binárias

EXAMPLE 7.2 EFFECTS OF COMPUTER OWNERSHIP ON COLLEGE GPA

In order to determine the effects of computer ownership on college grade point average, we estimate the model

$$colGPA = \beta_0 + \delta_0 PC + \beta_1 hsGPA + \beta_2 ACT + u,$$

where the dummy variable *PC* equals one if a student owns a personal computer and zero otherwise. There are various reasons PC ownership might have an effect on *colGPA*. A student's work might be of higher quality if it is done on a computer, and time can be saved by not having to wait at a computer lab. Of course, a student might be more inclined to play computer games or surf the Internet if he or she owns a PC, so it is not obvious that δ_0 is positive. The variables *hsGPA* (high school GPA) and *ACT* (achievement test score) are used as controls: it could be that stronger students, as measured by high school *GPA* and *ACT* scores, are more likely to own computers. We control for these factors because we would like to know the average effect on *colGPA* if a student is picked at random and given a personal computer.

Using the data in GPA1.RAW, we obtain

$$\begin{aligned} \widehat{colGPA} &= \underset{(.33)}{1.26} + \underset{(.057)}{.157}\, PC + \underset{(.094)}{.447}\, hsGPA + \underset{(.0105)}{.0087}\, ACT \\ n &= 141,\ R^2 = .219. \end{aligned} \tag{7.6}$$

This equation implies that a student who owns a PC has a predicted GPA about .16 points higher than a comparable student without a PC (remember, both *colGPA* and *hsGPA* are on a four-point scale). The effect is also very statistically significant, with $t_{PC} = .157/.057 \approx 2.75$.

What happens if we drop *hsGPA* and *ACT* from the equation? Clearly, dropping the latter variable should have very little effect, as its coefficient and *t* statistic are very small. But *hsGPA* is very significant, and so dropping it could affect the estimate of β_{PC}. Regressing *colGPA* on *PC* gives an estimate on *PC* equal to about .170, with a standard error of .063; in this case, $\hat{\beta}_{PC}$ and its *t* statistic do not change by much.

In the exercises at the end of the chapter, you will be asked to control for other factors in the equation to see if the computer ownership effect disappears, or if it at least gets notably smaller.

5 In Example 7.2, let *noPC* be a dummy variable equal to one if the student does not own a PC, and zero otherwise.

(i) If *noPC* is used in place of *PC* in equation (7.6), what happens to the intercept in the estimated equation? What will be the coefficient on *noPC*? (*Hint:* Write $PC = 1 - noPC$ and plug this into the equation $\widehat{colGPA} = \hat{\beta}_0 + \hat{\delta}_0 PC + \hat{\beta}_1 hsGPA + \hat{\beta}_2 ACT$.)

(ii) What will happen to the *R*-squared if *noPC* is used in place of *PC*?

(iii) Should *PC* and *noPC* both be included as independent variables in the model? Explain.

C1 Use the data in GPA1.RAW for this exercise.

(i) Add the variables *mothcoll* and *fathcoll* to the equation estimated in (7.6) and report the results in the usual form. What happens to the estimated effect of PC ownership? Is *PC* still statistically significant?

(ii) Test for joint significance of *mothcoll* and *fathcoll* in the equation from part (i) and be sure to report the *p*-value.

(iii) Add $hsGPA^2$ to the model from part (i) and decide whether this generalization is needed.

C2 Use the data in WAGE2.RAW for this exercise.

(i) Estimate the model

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 married + \beta_5 black + \beta_6 south + \beta_7 urban + u$$

and report the results in the usual form. Holding other factors fixed, what is the approximate difference in monthly salary between blacks and nonblacks? Is this difference statistically significant?

(ii) Add the variables $exper^2$ and $tenure^2$ to the equation and show that they are jointly insignificant at even the 20% level.

(iii) Extend the original model to allow the return to education to depend on race and test whether the return to education does depend on race.

(iv) Again, start with the original model, but now allow wages to differ across four groups of people: married and black, married and nonblack, single and black, and single and nonblack. What is the estimated wage differential between married blacks and married nonblacks?

Parte 2 : Heterocedasticidade

1 Which of the following are consequences of heteroskedasticity?
(i) The OLS estimators, $\hat{\beta}_j$, are inconsistent.
(ii) The usual F statistic no longer has an F distribution.
(iii) The OLS estimators are no longer BLUE.

2 Consider a linear model to explain monthly beer consumption:

$$beer = \beta_0 + \beta_1 inc + \beta_2 price + \beta_3 educ + \beta_4 female + u$$

$$\mathrm{E}(u|inc, price, educ, female) = 0$$

$$\mathrm{Var}(u|inc, price, educ, female) = \sigma^2 inc^2.$$

Write the transformed equation that has a homoskedastic error term.

3 True or False: WLS is preferred to OLS when an important variable has been omitted from the model.

4 Using the data in GPA3.RAW, the following equation was estimated for the fall and second semester students:

$$\begin{aligned}
\widehat{trmgpa} = {} & -2.12 + .900\, crsgpa + .193\, cumgpa + .0014\, tothrs \\
& \quad (.55) \quad (.175) \qquad\quad (.064) \qquad\quad (.0012) \\
& \quad [.55] \quad [.166] \qquad\quad [.074] \qquad\quad [.0012] \\
& + .0018\, sat - .0039\, hsperc + .351\, female - .157\, season \\
& \quad (.0002) \qquad (.0018) \qquad\quad (.085) \qquad\quad (.098) \\
& \quad [.0002] \qquad [.0019] \qquad\quad [.079] \qquad\quad [.080] \\
n = {} & 269,\ R^2 = .465.
\end{aligned}$$

Here, *trmgpa* is term GPA, *crsgpa* is a weighted average of overall GPA in courses taken, *cumgpa* is GPA prior to the current semester, *tothrs* is total credit hours prior to the semester, *sat* is SAT score, *hsperc* is graduating percentile in high school class, *female* is a gender dummy, and *season* is a dummy variable equal to unity if the student's sport is in season during the fall. The usual and heteroskedasticity-robust standard errors are reported in parentheses and brackets, respectively.

(i) Do the variables *crsgpa*, *cumgpa*, and *tothrs* have the expected estimated effects? Which of these variables are statistically significant at the 5% level? Does it matter which standard errors are used?
(ii) Why does the hypothesis H_0: $\beta_{crsgpa} = 1$ make sense? Test this hypothesis against the two-sided alternative at the 5% level, using both standard errors. Describe your conclusions.
(iii) Test whether there is an in-season effect on term GPA, using both standard errors. Does the significance level at which the null can be rejected depend on the standard error used?

C8 Use the data set GPA1.RAW for this exercise.
(i) Use OLS to estimate a model relating *colGPA* to *hsGPA*, *ACT*, *skipped*, and *PC*. Obtain the OLS residuals.
(ii) Compute the special case of the White test for heteroskedasticity. In the regression of $\hat{u}_i^2$ on $\widehat{colGPA}_i$, $\widehat{colGPA}_i^2$, obtain the fitted values, say $\hat{h}_i$.
(iii) Verify that the fitted values from part (ii) are all strictly positive. Then, obtain the weighted least squares estimates using weights $1/\hat{h}_i$. Compare the weighted least squares estimates for the effect of skipping lectures and the effect of PC ownership with the corresponding OLS estimates. What about their statistical significance?