Instructor's Manual

Answers to Odd-numbered Exercises

Chapter 1

1-3. (a) The coefficient of Li represents the change in the percentage chance of making a putt when the length of the putt increases by one foot. In this case, the percentage chance of making the putt decreases by 4.1 for each foot longer the putt is.

(b) The equations are identical. To convert one to the other, note that $\hat{P}_i = P_i - e_i$, which is true because $e_i = P_i - \hat{P}_i$ (or more generally, $e_i = Y_i - \hat{Y}_i$).

(c) 42.6 percent, yes; 79.5 percent, no (too low); -18.9 percent, no (negative!).

(d) One problem is that the theoretical relationship between the length of the putt and the percentage of putts made is almost surely non-linear in the variables; we'll discuss models appropriate to this problem in Chapter 7. A second problem is that the actual dependent variable is limited by zero and one but the regression estimate is not; we'll discuss models appropriate to this problem in Chapter 13.

1-5. (a) β_Y is the change in S caused by a one-unit increase in Y, holding G constant and β_G is the change in S caused by a one-unit increase in G, holding Y constant.

(b) +, -

(c) Yes. Richer states spend at least some of their extra money on education, but states with rapidly growing student populations find it difficult to increase spending at the same rate as the student population, causing spending per student to fall, especially if you hold the wealth of the state constant.

(d) $\hat{S}_i = -183 + 0.1422Y_i - 59.26G_i$. Note that $59.26 \cdot 10 = 5926 \cdot 0.10$, so nothing in the equation has changed except the scale of the coefficient of G.

1-7. (a) β_2 represents the impact on the wage of the *i*th worker of a one-year increase in the education of the *i*th worker, holding constant that worker's experience and gender.

(b) β_3 represents the impact on the wage of the *i*th worker of being male instead of female, holding constant that worker's experience and education.

(c) There are two ways of defining such a dummy variable. You could define $COLOR_i = 1$ if the *i*th worker is a person of color and 0 otherwise, or you could define $COLOR_1 = 1$ if the *i*th worker is not a person of color and 0 otherwise. (The actual name you use for the variable doesn't have to be "COLOR." You could choose any variable name as long as it didn't conflict with the other variable names in the equation.)

(d) We'd favor adding a measure of the quality of the worker to this equation, and answer iv, the number of employee of the month awards won, is the best measure of quality in this group. As tempting as it might be to add the average wage in the field, it would be the same for each employee in the sample and thus wouldn't provide any useful information.

Chapter 2

2-3. (a) The squares are "least" in the sense that they are being minimized.

(b) If $R^2 = 0$, then RSS = TSS, and ESS = 0. If R^2 is calculated as ESS/TSS, then it cannot be negative. If R^2 is calculated as 1 – RSS/TSS, however, then it can be negative if RSS > TSS, which can happen if $\hat{Y}$ is a *worse* predictor of Y than $\bar{Y}$ (possible only with a non-OLS estimator or if the constant term is omitted).

(c) positive.

(d) We prefer Model T because it has estimated signs that meet expectations and also because it includes an important variable that Model A omits. A higher R^2 does not *automatically* mean that an equation is preferred.

2-5. (a) Even though the fit in Equation A is better, most researchers would prefer equation B because the signs of the estimated coefficients are as would be expected. In addition, X_4 is a theoretically sound variable for a campus track, while X_3 seems poorly specified because an especially hot *or* cold day would discourage fitness runners.

(b) The coefficient of an independent variable tells us the impact of a one-unit increase in that variable on the dependent variable holding constant the other explanatory variables in the equation. If we change the other variables in the equation, we're holding different variables constant, and so the $\hat{\beta}$ has a different meaning.

2-7. (a) Yes. We'd expect bigger colleges to get more applicants, and we'd expect colleges that used the common application to attract more applicants. It might seem at first that the rank of a college ought to have a positive coefficient, but the variable is defined as 1 = best, so we'd expect a negative coefficient for RANK.

(b) The meaning of the coefficient of SIZE is that for every increase of one in the size of the student body, we'd expect a college to generate 2.15 more applications, holding RANK and COMMONAP constant. The meaning of the coefficient of RANK is that every one-rank improvement in a college's USNews ranking should generate 32.1 more applications, holding SIZE and COMMONAP constant. These results do not allow us to conclude that a college's ranking is 15 times more important than the size of that college because the units of the variables SIZE and RANK are quite different in magnitude. On a more philosophical level, it's risky to draw any general conclusions at all from one regression estimated on a sample of 49 colleges.

(c) The meaning of the coefficient of COMMONAP is that a college that switches to using the common application can expect to generate 1222 more applications, holding constant RANK and SIZE. However, this result does not prove that a given college would increase applications by 1222 by switching to the common application. Why not? First, we don't trust this result because there may well be an omitted relevant variable (or two) and because all but three of the colleges in the sample use the common application. Second, in general, econometric results are evidence that can be used to support an argument, but in and of themselves they don't come close to "proving" anything.

(e) If you drop COMMONAP from the equation, $\bar{R}^2$ falls. This is evidence (but not proof) that COMMONAP belongs in the equation.