EE325: Homework 2 Solution

2.1 (i) Income, age, and family background (such as number of siblings) are just a few possibilities. It seems that each of these could be correlated with years of education. (Income and education are probably positively correlated; age and education may be negatively correlated because women in more recent cohorts have, on average, more education; and number of siblings and education are probably negatively correlated.)

(ii) Not if the factors we listed in part (i) are correlated with *educ*. Because we would like to hold these factors fixed, they are part of the error term. But if u is correlated with *educ*, then $\mathrm{E}(u|educ) \neq 0$, and so SLR.4 fails.

2.3 (i) Let $y_i = GPA_i$, $x_i = ACT_i$, and $n = 8$. Then $\bar{x} = 25.875$, $\bar{y} = 3.2125$, $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 5.8125$, and $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 56.875$. From equation (2.19), we obtain the slope as $\hat{\beta}_1 = 5.8125/56.875 \approx .1022$, rounded to four places after the decimal. From (2.17), $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \approx 3.2125 - (.1022)25.875 \approx .5681$. So we can write

$$GPA = .5681 + .1022\,ACT$$

$$n = 8.$$

The intercept does not have a useful interpretation because *ACT* is not close to zero for the population of interest. If *ACT* is 5 points higher, GPA increases by .1022(5) = .511.

(ii) The fitted values and residuals — rounded to four decimal places — are given along with the observation number i and *GPA* in the following table:

i	GPA	GPA	$\hat{u}$
1	2.8	2.7143	.0857
2	3.4	3.0209	.3791
3	3.0	3.2253	–.2253
4	3.5	3.3275	.1725
5	3.6	3.5319	.0681
6	3.0	3.1231	–.1231
7	2.7	3.1231	–.4231
8	3.7	3.6341	.0659

You can verify that the residuals, as reported in the table, sum to –.0002, which is pretty close to zero given the inherent rounding error.

(iii) When $ACT = 20$, $\widehat{GPA} = .5681 + .1022(20) \approx 2.61$.

(iv) The sum of squared residuals, $\sum_{i=1}^{n} \hat{u}_i^2$, is about .4347 (rounded to four decimal places), and the total sum of squares, $\sum_{i=1}^{n} (y_i - \bar{y})^2$, is about 1.0288. So the R-squared from the regression is

$$R^2 = 1 - \text{SSR/SST} \approx 1 - (.4347/1.0288) \approx .577.$$

Therefore, about 57.7% of the variation in *GPA* is explained by *ACT* in this small sample of students.

2.11 (i) We would want to randomly assign the number of hours in the preparation course so that *hours* is independent of other factors that affect performance on the SAT. Then, we would collect information on SAT score for each student in the experiment, yielding a data set $\{(sat_i, hours_i): i = 1,\ldots,n\}$, where n is the number of students we can afford to have in the study. From equation (2.7), we should try to get as much variation in $hours_i$ as is feasible.

(ii) Here are three factors: innate ability, family income, and general health on the day of the exam. If we think students with higher native intelligence think they do not need to prepare for the SAT, then ability and *hours* will be negatively correlated. Family income would probably be positively correlated with *hours*, because higher income families can more easily afford preparation courses. Ruling out chronic health problems, health on the day of the exam should be roughly uncorrelated with hours spent in a preparation course.

(iii) If preparation courses are effective, β_1 should be positive; other factors equal, an increase in *hours* should increase *sat*.

(iv) The intercept, β_0 , has a useful interpretation in this example: because $E(u) = 0$, β_0 is the average SAT score for students in the population with *hours* = 0.

Computer exercise 2.3 (i) The estimated equation is

$$\widehat{sleep} = 3{,}586.4 - .151\ totwrk$$

$$n = 706,\ R^2 = .103.$$

The intercept implies that the estimated amount of sleep per week for someone who does not work is 3,586.4 minutes, or about 59.77 hours. This comes to about 8.5 hours per night.

(ii) If someone works two more hours per week, then $\Delta totwrk = 120$ (because *totwrk* is measured in minutes), and so $\Delta\widehat{sleep} = -.151(120) = -18.12$ minutes. This is only a few minutes a night. If someone were to work one more hour on each of five working days, $\Delta\widehat{sleep} =$

$-.151(300) = -45.3$ minutes, or about five minutes a night.