Solution to suggested practice questions (Chapter 2)

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 $E(u|educ) \neq 0$, 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 $ACT = 20$, $GPA = .5681 + .1022(20) \approx 2.61$.

(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.

2.4 (i) When *cigs* = 0, predicted birth weight is 119.77 ounces. When *cigs* = 20, *bwght* = 109.49. This is about an 8.6% drop.

(ii) Not necessarily. There are many other factors that can affect birth weight, particularly overall health of the mother and quality of prenatal care. These could be correlated with cigarette smoking during birth. Also, something such as caffeine consumption can affect birth weight, and might also be correlated with cigarette smoking.

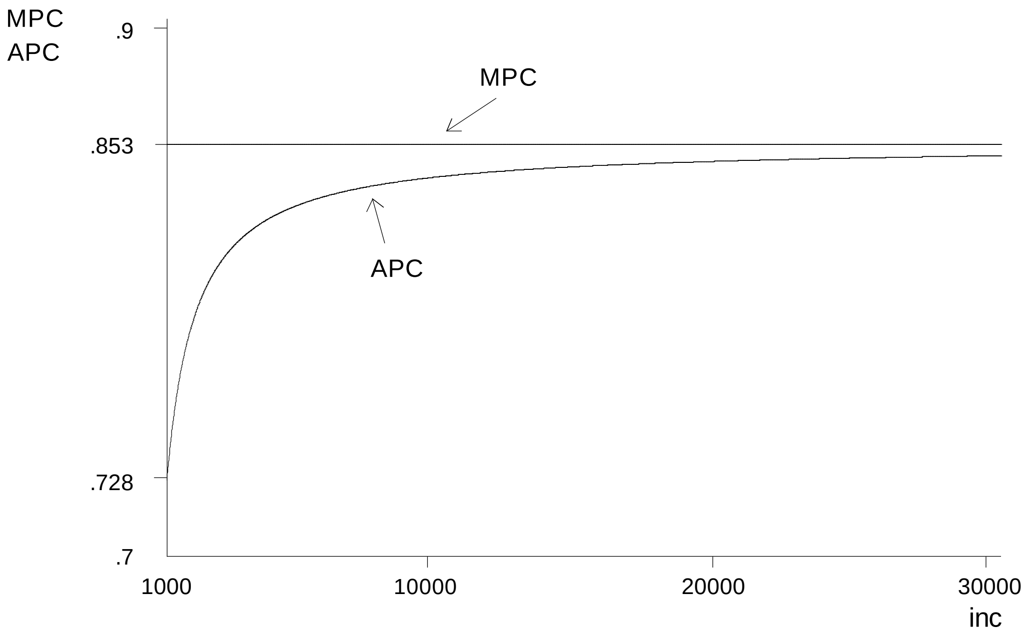
(iii) If we want a predicted *bwght* of 125, then *cigs* = (125 – 119.77)/(–.524) $\approx$ –10.18, or about –10 cigarettes. This is nonsense, of course, and it shows what happens when we are trying to predict something as complicated as birth weight with only a single explanatory variable. The largest predicted birth weight is necessarily 119.77. Yet, almost 700 of the births in the sample had a birth weight higher than 119.77.

(iv) 1,176 out of 1,388 women did not smoke while pregnant, or about 84.7%. Because we are using only *cigs* to explain birth weight, we have only one predicted birth weight at *cigs* = 0. The predicted birth weight is necessarily roughly in the middle of the observed birth weights at *cigs* = 0, and so we will under predict high birth rates.

2.5 (i) The intercept implies that when *inc* = 0, *cons* is predicted to be negative $124.84. This, of course, cannot be true, and reflects the fact that this consumption function might be a poor predictor of consumption at very low-income levels. On the other hand, on an annual basis, $124.84 is not so far from zero.

(ii) Just plug 30,000 into the equation: *cons* = –124.84 + .853(30,000) = 25,465.16 dollars.

(iii) The MPC and the APC are shown in the following graph. Even though the intercept is negative, the smallest APC in the sample is positive. The graph starts at an annual income level of $1,000 (in 1970 dollars).



2.6 (i) Yes. If living closer to an incinerator depresses housing prices, then being farther away increases housing prices.

(ii) If the city chooses to locate the incinerator in an area away from more expensive neighborhoods, then log(*dist*) is positively correlated with housing quality. This would violate SLR.4, and OLS estimation is biased.

(iii) Size of the house, number of bathrooms, size of the lot, age of the home, and quality of the neighborhood (including school quality), are just a handful of factors. As mentioned in part (ii), these could certainly be correlated with *dist* [and log(*dist*)].