BOSTON COLLEGE
Department of Economics
EC 228 Econometrics, Prof. Baum, Mr. Barbato, Spring 2003
Problem Set 2

2.3

(i) Let $y_i = GPA_i, x_i = ACT_i$, and $n = 8$. Then $\overline{x} = 25.875, \overline{y} = 3.2125, \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = 5.8125$, and $\sum_{i=1}^{n} (x_i - \overline{x})^2 = 56.875$. From equation (2.9), we obtain the slope as $\widehat{\beta}_1 = 5.8125/56.875 \approx .1022$, rounded to four places after the decimal. From (2.17), $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} \approx 3.2125 - (.1022)25.875 \approx .5681$. So we can write

$$\begin{aligned} \widehat{GPA} &= .5681 + .1022 ACT \\ n &= 8 \end{aligned}$$

The intercept does not have a useful interpretation here because ACT is not close to zero for the population of interest. If ACT is 5 points higher, $\widehat{GPA}$ 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	$\widehat{GPA}$	$\widehat{\mu}$
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, $\sum_{i=1}^{n} \widehat{\mu}_i^2$, is about 1.0288. So the R-squared from the regression is

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

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 119.7 ounces. When $cigs = 20, \widehat{bwght} = 109.49$. This is about a 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, sometimes such as caffeine consumption can affect birth weight, and might also be correlated with cigarette smoking.

(iii) If we want a predicted *bwight* 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%

2.11. (i) Average salary is about 865.864, which means \$865,864 because *salary* is in thousands of dollars. Average *ceoten* is about 7.95.

(ii) There are five CEOs with $ceoten = 0$. The longest tenure is 37 years.

(iii) The estimated equation is

$$\begin{aligned} \widehat{\log(salary)} &= 6.51 + .00971 totwrk \\ n &= 177, R^2 = .013 \end{aligned}$$

We obtain the approximate percentage change in *salary* given $\Delta ceoten = 1$ by multiplying the coefficient on *ceoten* by 100, $100(.0097) = .97\%$. Therefore, one more year as CEO is predicted to increase salary by almost 1%

.

2.14. (i) The constant elasticity model is a log-log model:

$\log(rd) = \beta_0 + \beta_1 \log(sales) + u$

where β_1 is the elasticity of rd with respect to *sales*

(ii) The estimated equation is

$$\widehat{\log(rd)} = -4.105 + 1.076 \log(sales)$$
$$n = 32, R^2 = .910.$$

The estimated elasticity of *rd* with respect to *sales* is 1.076, which is just above one. A one percent increase in *sales* is estimated to increase *rd* by about 1.08%.

3.1. (i) *hsperc* is defined so that the smaller it is, the lower the student's standing in high school. Everything else equal, the worse the student's standing in high school, the lower is his/her expected college GPA.

(ii) Just plug these values into the equation:

$$\widehat{co\lg pa} = 1.392 - .0135(20) + .00148(1050) = 2.676$$

(iii) The difference between A and B is simply 140 times the coefficent on *sat,* because *hsperc* is the same for both students. So A is predicted to have a score $.00148(140) \approx .207$ higher.

(iv) With *hsperc* fixed, $\Delta\widehat{co\lg pa} = .00148\Delta sat$. Now, we want to find Δsat such that $\Delta\widehat{co\lg pa} = .5$, so $.5 = .00148(\Delta sat)$ or $\Delta sat = .5/(.00148) \approx 338$. Perhaps not surprisingly, a large ceteris paribus difference in SAT score - almost two and one-half standard deviation - is needed to obtain a predicted difference in college GPA or a half a point.

3.3. (i) If adults trade off sleep for work, more work implies less sleep (other things equal), so $\beta_1 < 0$.

(ii) The signs of β_2 and β_3 are not obvious to me. One could argue that more educated people like to get more out of life, and so, other things equal, they sleep less ($\beta_2 < 0$). The relationship between sleeping and age is more complicated than this model suggests, and economists are not in the best position to judge such things.

(iii) Since *totwrk* is in minutes, we must convert five hours into minutes: $\Delta totwrk = 5(60) = 300$. Then *sleep* is predicted to fall by $.148(300) = 44.4$ minutes. For a week, 45 minutes less sleep is not an overwhelming change.

(iv) More education implies less predicted time sleeping, but the effect is quite small. If we assume the difference between college and high school is four years, the college graduate sleeps about 45 minutes less per week, other things equal.

(v) Not surprisingly, the three explanatory variables explain only about 11.3% of the variation in *sleep*. One important factor in the error term is general health. Another is marital status, and whether the person has children. Health (however we measure that), marital status, and number and ages of children would generally be correlated with *totwrk*. (For example, less healthy peoplewould tend to work less.)