

Solutions to Problem Set 1 – KAN-CCMVV2401U

Part 1

1 (a) The average of *educ* is about 12.6 years. There are two people reporting zero years of education, and 19 people reporting 18 years of education.

(b) The average of *wage* is about \$5.90, which seems low in 2005.

(c) Using Table B-60 in the 2004 *Economic Report of the President*, the CPI was 56.9 in 1976 and 184.0 in 2003. To convert 1976 dollars into 2003 dollars, we use the ratio of the CPIs, which is $184/56.9 \approx 3.23$. Therefore, the average hourly wage in 2003 dollars is roughly $3.23(\$5.90) \approx \19.06 , which is a reasonable figure.

(d) The sample contains 252 women (the number of observations with *female* = 1) and 274 men.

2 (a) The largest is 100, the smallest is 0.

(b) 38 out of 1,823, or about 2.1 percent of the sample.

(c) 17

(d) The average of *math4* is about 71.9 and the average of *read4* is about 60.1. So, at least in 2001, the reading test was harder to pass.

(e) The sample correlation between *math4* and *read4* is about .843, which is a very high degree of (linear) association. Not surprisingly, schools that have high pass rates on one test have a strong tendency to have high pass rates on the other test.

(f) The average of *exppp* is about \$5,194.87. The standard deviation is \$1,091.89, which shows rather wide variation in spending per pupil. [The minimum is \$1,206.88 and the maximum is \$11,957.64.]

Part 2

3 (a) The constant elasticity equation is

$$\log(\text{salary}) = 4.62 + .162 \log(\text{sales}) + .107 \log(\text{mktval})$$

$$n = 177, R^2 = .299.$$

(b) We cannot include profits in logarithmic form because profits are negative for nine of the companies in the sample. When we add it in levels form we get

$$\log(\text{salary}) = 4.69 + .161 \log(\text{sales}) + .098 \log(\text{mktval}) + .000036 \text{profits}$$

$$n = 177, R^2 = .299.$$

The coefficient on *profits* is very small. Here, *profits* are measured in millions, so if profits increase by \$1 billion, which means $\Delta profits = 1,000$ – a huge change – predicted salary increases by about only 3.6%. However, remember that we are holding sales and market value fixed.

Together, these variables (and we could drop *profits* without losing anything) explain almost 30% of the sample variation in $\log(salary)$. This is certainly not “most” of the variation.

(c) Adding *ceoten* to the equation gives

$$\log(salary) = 4.56 + .162 \log(sales) + .102 \log(mktval) + .000029 profits + .012 ceoten$$

$$n = 177, R^2 = .318.$$

This means that one more year as *CEO* increases predicted salary by about 1.2%.

(d) The sample correlation between $\log(mktval)$ and *profits* is about .78, which is fairly high. As we know, this causes no bias in the OLS estimators, although it can cause their variances to be large. Given the fairly substantial correlation between market value and firm profits, it is not too surprising that the latter adds nothing to explaining CEO salaries. Also, *profits* is a short term measure of how the firm is doing while *mktval* is based on past, current, and expected future profitability.

4 (a) The results of the regression are

$$\log(bwght) = 8.06 - .0032 cigs + .0056 npvis$$

$$n = 1,656, R^2 = .0159$$

The signs of the estimated slopes imply that smoking more cigarettes during pregnancy reduces birth weight, a common finding. The positive coefficient on *npvis* implies that more prenatal care leads to higher birth weight. (This latter effect should be interpreted with caution, however, as a very high number of prenatal visits could reflect a difficult pregnancy.)

(b) The sample standard deviation of *npvis* is about 3.68, so a one standard deviation increase in *npvis* increases predicted $\log(bwght)$ by $.0056(3.68) \approx .021$, so about a 2.1% increase in birth weight.

(c) The simple regression results are

$$\log(bwght) = 8.12 - .0034 cigs$$

$$n = 1,722, R^2 = .0053$$

The estimated effect of smoking is very similar using simple and multiple regression.

(d) The sample correlation between *cigs* and *npvis* is only about $-.0387$, which helps to explain why dropping *npvis* from the regression has relatively little effect on the *cigs* coefficient.

(e) When *mage*, *meduc*, *fage*, and *feduc* are added to the regression in part (i), the number of observations is 1,624 and the *R*-squared is .0169. In other words, even adding parents' age and education, we explain less than 2% of the variation in $\log(bwght)$. Birth weight is not an easy variable to explain.