Solutions to Problem Set 2 (Due October 2)

EC 228 01, Fall 2013   Prof. Baum, Mr. Lim

Maximum number of points for Problem set 2 is: 67

Problem 2.4

(i) (2 pts.) When cigs $= 0$, predicted birth weight is 119.77 ounces. When cigs $= 20$, bwght $= 109.49$. This is about an 8.6 percent drop.

(ii) (2 pts.) 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) (2 pts.) If we want a predicted $bwght$ of 125, then $cigs = (119.77 - 125)/(524) \approx -10.18$, or about negative 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.

(iv) (2 pts.) Yes. Since about 80 percent of women did not smoke, but we only have one birthweight estimate when $cigs = 0$ (since we are only using cigs to explain birth weight) then the predicted birthweight for $cigs = 0$ is in the middle of the entire distribution of birth weights when $cigs = 0$. If we believe that non-smokers have heavier babies than smokers, then we would under-predict high birth weights.

Problem 2.5

(i) (2 pts.) The intercept implies that when income is 0, consumption is predicted to be negative 124.84 dollars. This, of course, cannot be true, and is reflective of this consumption function being a poor predictor of consumption at very low income levels. On the other hand, relative to annual income, 124.84 dollars is not so far from zero.

(ii) (2 pts.) Just plug 30,000 into the equation: $= 124.84 + .853(30,000) = 25,465.16$ dollars.

(iii) (2 pts.) The MPC is a straight line at $\beta_1 = .853$. While the APC has a negative intercept, for incomes of very low range, ie $inc = 1000$, the APC is greater than zero. As inc goes to infinity, APC approaches MPC from below.

Problem 2.6

(i) (2 pts.) The coefficient is .312 meaning there is a positive relationship between housing prices and distance from a garbage incinerator. This is what we would expect. Increasing a house's distance from a garbage incinerator, holding everything else constant, should result in an increased home value.

(ii) (2 pts.) If the city chose to locate the incinerator in an area away from more expensive neighborhoods, then $log(dist)$ is positively correlated with housing quality. This would make OLS estimation biased.

(iii) (2 pts.) Size of the house, number of bathrooms, size of the lot, age of the home, proximity to parks, and quality of the neighborhood (including school quality) are just a handful of factors that could influence price. As mentioned in part (ii), these could certainly be correlated with dist [and $log(dist)$]. For example, it might be likely that a city planner would want to place the garbage incinerator away from the park and closer to a manufacturing area.

Problem 2.11

(i) (4 pts.) We would want to randomly assign the number of preparation hours to different students so that *hours* is independent of other factors that affect total SAT score. Then we collect the information on SAT score for each student in the experiment, yielding a dataset $\{(sat_i, hours_i), i = 1, 2, ...n\}$, where n is the number of total students we can afford to have in the experiment. From equation (2.7), we should try to get as much variation in the $hours_i$ as is feasible.

(ii)(2 pts.) Here are three factors: innate ability, family income, and general health on the day of 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 exam should be roughly uncorrelated with *hours* spent in the preparation course.

(iii) (2 pts.) If the preparation course is effective, β_1 should be positive: other factors equal, an increase in preparation *hours* should increase the SAT score.

(iv) (2 pts.) Since $E(\mu) = 0$, β_0 is the average SAT score for students in the population with *hours* = 0, those who do not take the preparation courses.

Problem C 2.5

(i) (3 pts.) 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) (3 pts.) The estimated equation is:

$$\widehat{\log(rd)} = \text{-4.105} + 1.076 \log(sales)$$
$$n=32, \ R^2=0.910$$

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 1.076%.

Problem C 2.7

(i) (2 pts.) The average gift is about 7.44 Dutch guilders. Out of 4268 respondents, 2561 did not give a gift, or about 60 percent.

(ii) (1 pt.) The average mailings per year is about 2.05. The minimum value is 0.25 (which presumably means that someone has been on the mailing list for at least four years) and maximum is 3.5.

(iii) (2 pts.) The estimated equation is:

$$\widehat{gift} = 2.01 + 2.65 \; mailsyear$$
$$\text{n=4268, } R^2\text{=0.0138}$$

(iv) (2 pts.) The slope coefficient means that each mailing per year is associated with on average 2.65 additional guilders. Therefore, if each mailing costs one guilder, the expected profit from each mailing is estimated to be 1.65 guilders. This is only the average, however. Some mailings generate no contributions, or a contribution less than the mailing cost; other mailings generated much more than the mailing cost.

(v) (2 pts.) The smallest *mailsyear* in the sample is 0.25, the smallest predicted value of *gifts* is 2.01+2.65(0.25)≈2.67. Even if we look at the overall population, where some people have received no mailings, the smallest predicted value is about two. So, with this estimated equation, we never predict zero charitable gifts.

Problem 3.1

(i) (1 pts.) *hsperc* is defined in a way so that the smaller it is, the lower the student's standing in high school. Everything else equal, the worse the student's standing, i.e. the larger the *hsperc*, the lower is his/her expected college GPA.

(ii) (1 pt.) 2.676

(iii) (1 pt.) The difference between A and B is simply 140 times the coefficient on *sat*, *hsperc* is the same for both, then A will have a higher college GPA by 0.00148(140)≈0.207. Seems not too large.

(iv) (1 pt.) With *hsperc* fixed, $\Delta \widehat{colgpa} = 0.00148\Delta sat$. To get a o.5 difference in *colgpa*, SAT difference of 338 is needed. A large SAT difference is needed to generate only a half point difference in college GPA.

Problem 3.4

(i) (1 pt.) A larger rank number means the school has less prestige, which lowers the starting salaries.

(ii) (2 pts.) $\beta_1 > 0, \beta_2 > 0$. Both LSAT and GPA are measures of the quality of the entering class. No matter where better students attend law school, we expect them to earn more, on average. $\beta_3 > 0, \beta_4 > 0$. The number of volumes in the law library and the tuition cost are both measures of the school quality.

(iii) (1 pt.) 24.8%

(iv) (1 pt.) This is an elasticity: one percent increase in library volumes implies a 0.095% increase in predicted median starting salary, other thins equal.

(v) (1 pt.) It is definitely better to attend a higher ranked law school. A difference in ranking of 20 is worth 0.0033(20)=6.6% higher in starting salary.

Problem 3.10

(i) (2 pts.) Because x_1 is highly correlated with x_2, x_3, and these latter variables have large partial effects on y, the simple and multiple regression coefficients on x_1 can differ by large amounts. We have not done this case explicitly, but given equation (3.46) and the discussion with a single omitted variable, the intuition is pretty straightforward.

(ii) (2 pts.) Here we would expect $\tilde{\beta}_1$ and $\hat{\beta}_1$ to be similar. The amount of correlation between x_2 and x_3 does not directly affect the regression estimate on x_1 if x_1 is essentially uncorrelated with x_2 and x_3.

(iii) (2 pts.) In this case we have (unnecessarily) introducing multicollinearity into the regression. Adding x_2 and x_3 increases the standard error of the coefficient on x_1 substantially, so $se(\hat{\beta}_1)$ is likely to be much larger than $se(\tilde{\beta}_1)$.

(iv) (2 pts.) In this case, adding x_2 and x_3 will decrease the residual variance without causing much collinearity (since they are uncorrelated with x_1), so we should see $se(\hat{\beta}_1)$ smaller. The correlation between x_2 and x_3 does not directly affect $se(\hat{\beta}_1)$.

Problem C 3.6

(i) (1 pt.) The slope from regression IQ on $educ$ is $\tilde{\delta}_1 = 3.53383$.

(ii) (1 pt.) The slope coefficient from $\log(wage)$ on $educ$ is $\tilde{\beta}_1 = 0.05984$.

(iii) (1 pt.) The coefficients from $\log(wage)$ on $educ$ and IQ are $\hat{\beta}_1 = 0.03912$ and $\hat{\beta}_2 = 0.00586$, respectively.

(iv) (1 pt.) We have $\hat{\beta}_1 + \tilde{\delta}_1\hat{\beta}_2 = 0.03912 + 3.53383(0.00586) \approx = 0.05983$, which is very close to 0.05984.