ECO220Y1Y, Term Test #5, Prof. Murdock: SOLUTIONS

April 5, 2018, 9:10 – 11:00 am

(1) (a) Answering requires using Specification (2) in Table 1 because we are simply asked for the mean salary of full professors, regardless of sex. As the reference (omitted) category is full professor, the constant term reveals that the mean salary of full professors is $182,036.

(b) Regression (3) controls for job title whereas Regression (1) does not. Job title is clearly related to salary and must also be related to sex – female faculty tend to be in lower ranked positions – such that once we control for it, more than half of the salary difference by sex disappears: males are still paid more on average, but once we hold job title constant, the sex-based discrepancy lessens. (Note: For the purposes of interpretation, remember that job title is a single categorical variable. While we use a suite of dummy variables to include a categorical variable with multiple categories, remember that these categories are *mutually exclusive* (aka *disjoint*).)

(c) From the formula sheets, we plug into: $b_j \pm t_{\alpha/2} s_{b_j}$ with degrees of freedom $\nu = n - k - 1$. The degrees of freedom are $\nu = 1{,}067 - 5 - 1 = 1{,}061$ and referencing the Student t table we use 3.300 (to be conservative rather than using 3.291, which is for infinite degrees of freedom). Hence, we obtain $0.0403 \pm 3.300 * 0.0108$, which gives a lower confidence limit of 0.005 and an upper confidence limit of 0.076.

We are 99.9% confident that after controlling for job title, male faculty members on average have salaries that are between 0.5 percent to 7.6 percent higher than female faculty members. (*Note:* It is not correct to talk about percentage point differences because salaries are measured in dollars. We only talk about percentage points when talking about changes in something that is itself measured as a percent.)

(d) It would be 12,335.2 in Regression (1) and 0.0822 in Regression (4).

(e) The much higher R^2 in Regression (2) of 0.4513 means that about 45% of the variation in salaries of these faculty members can be explained by variation in their job title. In contrast, the much lower R^2 in Regression (1) of 0.0260 means that only 2.6% of the salary variation can be explained by variation in the sex of the faculty member. In other words, job title is much better predictor of salary than sex is (which is not surprising).

(f) There are zero female Clinical Lecturers. We can tell because the predicted salary of a female Clinical Lecturer is $135,800 (=(177.3490 – 41.5487)*1000) and there are no dots there in the provided diagnostic plot.

(g) The value 1198.11625 is the variance of salary when we include the president and the square root of it is the standard deviation of salaries, which is $34,614. It would be smaller for Regression (3) because that regression excludes the president – a clear outlier with the highest salary – and hence the variance (and s.d.) of salaries would be smaller without this extreme value.

(h) Answering requires using Regression (8). Since we included a dummy for the president, the regression will exactly predict his salary, which is $400,000 = (177.349 + 5.831735 + 216.8193)*1000.

(i) $H_0\!: \beta_{CL} = 0$ versus $H_0\!: \beta_{CL} \neq 0$, which tests if there is statistically significant difference in the mean salaries of clinical lecturers compared to full professors after controlling for sex (aka holding sex fixed). (You do not have to say anything about the president dummy because that effectively eliminates the president: notice that the coefficient estimates and standard errors are exactly the same in Regressions (3) and (8).)

(2) (a) For males aged 0 years (infants), the mortality rate has dropped substantially from 1990 to 2010 for everyone (the rich and the poor): the 2010 line is far below the 1990 line everywhere. Further, the amount of mortality inequality has declined between 1990 and 2010, which the figure shows as the 2010 line being less steep than the 1990 line. There is still inequality in 2010 – the mortality rates for male infants are higher in poor counties – but the difference in the mortality rates between rich and poor counties is less stark than it was in 1990.

(b) $mortality_{it} = \beta_0 + \beta_1 povertyrank_{it} + \beta_2 yr1990_{it} + \beta_3 yr1990_{it} * povertyrank_{it} + \varepsilon_{it}$ where yr1990 is a dummy variable and the last variable is an interaction term.

(c) We can approximate the equation of each line: for 1990 the intercept is about 10.2 and the slope is about 0.083 and for 2010 the intercept is about 5.6 and the slope is about 0.036. This yields either of the following two (depending on what you make the reference (aka omitted) category):

One correct answer: mortality-hat = 10.2 + 0.083*poverty_rank – 4.68*yr2010 – 0.047*poverty_rank*yr2010

Another correct answer: mortality-hat = 5.6 + 0.036*poverty_rank + 4.68*yr1990 + 0.047*poverty_rank*yr1990

(3) (a) $H_0: \beta_1 = 0$ versus $H_1: \beta_1 \neq 0$ using $F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{0.2833/1}{(1-0.2833)/(20-1-1)} = 7.12$. Using the F table we see that 7.12 is above the critical value for $\alpha = 0.05$ (which is somewhere between 4.54 and 4.35) but is below the critical value for $\alpha = 0.01$ (which is somewhere between 8.68 and 8.10), which means that this regression is statistically significant at the 5% significance level, but not a 1% significance level.

(b) The value of the s_e, which measures the amount of scatter around the OLS line, is larger for males aged 65+ years. Inspecting the figures, it is roughly 0.2 for males aged 20-24 and roughly 5 for males aged 65+. (Note the scale of the y-axis. Mortality is much higher for old men: the s_e is bigger, even though the R^2 is also bigger.)

(c) Recall this simple regression is the same as making an inference about the difference in two means. The homoscedasticity assumption (one of the six assumptions underlying regression) means we use the equal variances case for inference about the difference in means and plug into $\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$ recalling that $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$. Plugging in we obtain $s_p^2 = \frac{(15-1)6.42^2 + (5-1)4.64^2}{15+5-2} = 36.84$ and the standard error $\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{\frac{36.84}{15} + \frac{36.84}{5}} = 3.13$.

(4) (a) $\widehat{salary} = 46748.06 + 26097.98 * 1.44 + 2226.66 * (2015 - 2013) - 15.84 * (2015 - 2013)^2 + 1616.14 - 7349.55 - 12225.70 - 2210.70 = \$68,549.3$

(b) Graph (A) is the correct one: it shows that, after controlling for the other faculty characteristics, there are diminishing returns to seniority. There are two ways to see that Graph (B) is wrong: (1) it shows that the slope is negative at 40 years but the derivative with respect to years is: 2226.66 + 2*(-15.84)*years and if we plug in 40 we see the slope should still be positive (2226.66 + 2*(-15.84)*40 = 959.46) and (2) if we plug in years since hire equal to 1 we get something close to $115K, not $90K. There are two ways to see that Graph (C) is wrong: (1) it incorrectly shows increasing returns, which would correspond to a positive (not negative) coefficient on the years-squared term, and (2) if we plug in years since hire equal to 1 we get something close to $115K, not $100K. (To see the correct predicted salary for years since hire equal to 1: $116,611 = 46748 + 26098*1.6 + 2227*1 – 16*1^2 + 10084 + 15811.)