ECO220Y, Term Test #3: SOLUTIONS

February 10, 2017, 9:10 – 11:00 am

(1) (a) Given that the question asks for the size of an effect, we should us a confidence interval estimation approach for inference, not a hypothesis testing approach. Further, we must recognize that these are paired data.

$\bar{X}_d \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$ with degrees of freedom $\nu = n - 1$

$$s_d = SD[X_2 - X_1] = \sqrt{V[X_2] + V[X_1] - 2 * r * SD[X_1] * SD[X_2]}$$
$$= \sqrt{17.72117^2 + 15.29706^2 - 2 * 0.6387 * 17.72117 * 15.29706} = 14.2$$

$(60.8 - 53.04) \pm 2.064 \frac{14.204}{\sqrt{25}}$ with degrees of freedom $\nu = 25 - 1 = 24$

$7.76 \pm 2.064 * 2.84$

7.76 ± 5.86

$LCL = 1.9; UCL = 13.6$ We are 95% confident that marks on the test, which are out of 100 points, are between 1.9 points and 13.6 points higher after the one-day course. [Note this is quite wide: we're talking about getting anywhere between less than one additional multiple-choice question right to getting almost 7 additional questions right.]

(b) You should expect this suggestion to result in an unacceptably wide confidence interval. Even with paired data in part (a), which controls for differences in pre-existing financial literacy across people, the CI is wide. With independent samples the standard error would be much bigger, which can be explained conceptually or mathematically. Conceptually: with independent samples we do not control for variation in pre-existing financial literacy, which adds noise to the analysis. Mathematically: with independent samples the numerator would be much larger: $SD[X_2 - X_1] = \sqrt{V[X_2] + V[X_1]}$ if the correlation is zero (we do not subtract the last term).

(2) (a) The OLS intercept of -$74,486 has no interpretation because year equal to 0 is far outside the range of the data *and* negative real GDP per capita is not possible. The OLS slope means that in the 1990's in Bangladesh, real GDP on average rose by about $38.10 per capita, in 2011 USDs, per year. [**NOTE:** An interpretation requires clearly specifying the units: failing to specify that the number 38.10 is measured in dollars per capita is a serious omission.]

(b) The R-squared measures how well the OLS line fits the data: the closer it is to 1 the better the fit and the closer it is to 0 the worse. The very high R-squared value for Bangladesh of 0.977, means that <u>*GDP growth measured in levels is very STEADY*</u>: about 98 percent of the variation in the annual level GDP per capita during that decade is simply explained by the year. In contrast, in Kenya, GDP growth is much <u>*less STEADY*</u>: year is not as strong of a predictor of GDP and the R-squared value is much lower, 0.556. [Looking at all eight graphs: The very high R-squared values for Bangladesh, which are all greater than 0.97, means that *GDP growth, whether measured in levels or as an annual rate, is very steady*. In contrast, in Kenya, GDP growth is much more variable: year is not as strong of a predictor of GDP and the R-squared values are much lower, ranging from 0.56 to 0.88.]

(c) Neither shows substantial violations of the linearity assumption. While one applies the natural log to the y-variable (GDP per capita) and the other does not, and we usually associate using logs with straightening scatter plots, they are both fairly straight. At least for Kenya in the 2000's, the logs simply offer the convenience of the interpretation of the OLS coefficient as an estimate of the average annual growth rate, which allows easy comparison across countries with very different GDP per capita levels (i.e. poor versus rich countries). [In fact, none of the eight show substantial violations of the linearity assumption. Hence for these decades and countries, logs are just a convenience.]

(d) Given that two points determine a line, the OLS line is just the same as an algebraic line. We need to use the OLS results in Graphs #3, #4, #7 and #8. For Bangladesh the estimated average annual growth rate from 1990 to 2000 is 2.5% and from 2000 to 2010 is 4.3%. For Kenya the estimated average annual growth rate from 1990 to 2000 is -0.7% and from 2000 to 2010 is 1.7%. Hence, the slope $m = \frac{\Delta y}{\Delta x} = \frac{(4.3-1.7)}{(2.5--0.7)} = \frac{2.6}{3.2} \approx 0.8$ and the line can be obtained from the point-slope formula $(y - y_1) = m(x - x_1)$ to be $(y - 4.3) = 0.8(x - 2.5)$, which we can now write as $\hat{y} = 2.3 + 0.8x$, where y is the growth rate (%) for 2000 to 2010 and x is the growth rate (%) for 1990 to 2000. (Note: If one measures growth in terms of rates, instead of as percentages, then the line would be $\hat{y} = 0.023 + 0.8x$.)

(3) $P - value = P(Z < -3.59) + P(Z > 3.59) \approx 0$ There is a large difference in the percent called back, with 9.1% of females being called back compared to 7.2% of males: a huge 1.9 percentage point difference. Alternatively, we can say that the callback rate for females is 26 percent higher (100*(9.1-7.2)/7.2). These results are clearly economically significant: a big difference that policy makers (and applicants!) would care about. These results are also statistically significant: the P-value is essentially 0, which means we can decisively reject the null hypothesis that there is no difference at even the most stringent significance levels: there is overwhelming evidence in favor of the conclusion that there is a difference in the callback rates for male and female applicants. Finally, we *can* infer that the candidate's sex alone, and not other factors, is causing a *higher* callback rate for female candidates. As clearly stated in the excerpt of the paper, sex (along with all other applicants' characteristics) was randomly assigned and hence is exogenous. Hence there can be no lurking/confounding/omitted/unobserved variables. [**NOTE:** Assessing economic significance requires looking at the *point estimate*. Some thought that the solution I provided was inappropriate for the questions asked and also computed a CI. That is not only unnecessary – the answer guide clearly indicates that you were only supposed to compute the P-value and then interpret the results with words – but also, including a CI weakens the answer by distracting your interpretation with tangential analysis that does not directly address any of the questions asked. CI estimates get at the precision of the estimates whereas economic significance is about the *size* of the estimate. How precisely it is estimated is a separate question that goes towards statistical significance.]

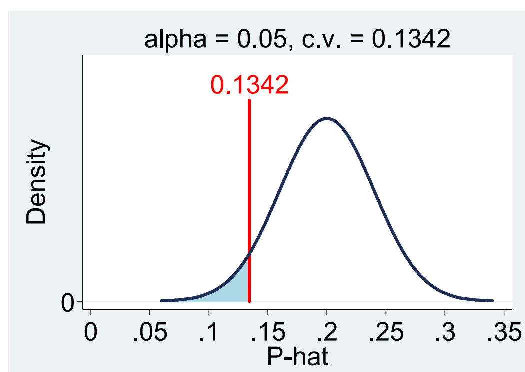
(4)

$H_0: p = 0.20$

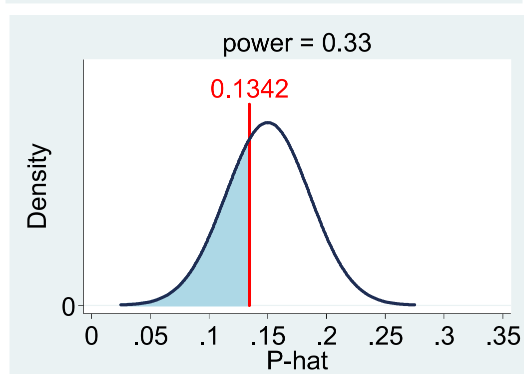
$H_1: p < 0.20$

Find the rejection region: $P(Z < -z_\alpha) = \alpha;\ P(Z < -1.645) = 0.05$

$$Z = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \Rightarrow -1.645 = \frac{\hat{P} - 0.20}{\sqrt{\frac{0.20(1-0.20)}{100}}} = \frac{\hat{P} - 0.20}{0.04} \Rightarrow c.v. = 0.1342$$



The unstandardized rejection region is $(-\infty,\ 0.1342)$ or $(0, 0.1342)$. In a random sample of 100 borrowers, we need less than 13.42% in default to prove that fewer than 1/5 of the population would default at a 5% sig. level. Power is the probability of being in the rejection region (i.e. correctly rejecting the false null and proving the true research hypothesis). $Power = P(\hat{P} < 0.1342 \mid p = 0.15, n = 100) = ?$



$$P(\hat{P} < 0.1342) = P\left(Z < \frac{0.1342 - 0.15}{\sqrt{\frac{0.15(1-0.15)}{100}}}\right) = P\left(Z < \frac{-0.0158}{0.0357}\right) = P(Z < -0.44) = 0.5 - 0.17 = 0.33$$

[**NOTE:** The question did *not* ask for an interpretation. Given that it could have, a full interpretation is provided next.] Hence, the chance that the researcher proves the true research hypothesis that default rates are below 20% is only about one-third with the planned sample size of 100 borrowers. This is low power (i.e. a pretty high chance of a Type II error). [Note: Some students may find and graph the probability of making a Type II error (failing to reject the false null hypothesis): $\beta = 0.67$. Power is the complement of making a Type II error: $power = 1 - \beta$.] Most likely the researcher would end up with an inconclusive result: insufficient proof for the research hypothesis. (The obvious remedy for this problem is to use a larger sample size: e.g. track 300 such loans instead of only 100.)

(5) (a) We should not expect anything in particular to happen to "(2.632)" because that is the standard deviation (s_{X_C}) of the variable recording the number of visits for each person in the control group (X_C) after the lottery: there is no reason to expect that a larger sample would have more or less variability (i.e. if we imagine the histogram, there would be thinner bins, but the spread and shape should be about the same). In contrast, we should expect that "(0.116)" becomes much smaller because that is a standard error ($s_{\bar{X}_C - \bar{X}_T}$) of the difference in the sample means ($\bar{X}_C - \bar{X}_T$) and as the sample sizes go up, sampling error goes down.

(b) $\hat{P}_{Treatment(no\ visits)} = 0.292\ (= 0.225 + 0.067)$

(c) $\bar{X}_{Treatment(two+)} = 3.864\ (= 3.484 + 0.380)$

(d)

$H_0: (\mu_T - \mu_C) = 0$

$H_1: (\mu_T - \mu_C) \neq 0$

$t = \frac{(\bar{X}_T - \bar{X}_C) - \Delta_0}{SE(\bar{X}_T - \bar{X}_C)} = \frac{0.380}{0.648} = 0.59$

$P - value = P(t < -0.59) + P(t > 0.59) = 2 * (0.5 - 0.2224) \approx 0.56$ (Note: We can use the Normal table because the degrees of freedom will be very large.)

This result is not statistically significant at any of the conventional significance levels: it is not even close to meeting the low burden of proof of a 10% significance level.