Homework 12: ECO220Y – SOLUTIONS

Required Problems:

(1) $\hat{P} \pm z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = 0.48 \pm 1.96\sqrt{\frac{0.48(1-0.48)}{81}} = 0.48 \pm 1.96 * 0.06 = 0.48 \pm 0.11$

We are 95% confident that among *all* felons, between 37 to 59 percent will commit another crime. The point estimate is 48% of felons will commit another crime and the margin of error is 11 percentage points. (Note: If you picked another confidence level that is alright so long as your tabular z value and interpretation match.)

(2) (a) The correct answer is **(D)**. In answering on p. 348 the book says "This shows the power of statistical analysis. From a sample of 1000 people we can make conclusions about a population of 100 million just as accurately as we can about a population of 1 million." (Note: (E) is not a correct answer (even though a sample size of 1,000 would be large enough for the Normal approximation to be appropriate) because it does not address the question asked.)

(b) There is no contradiction. The sample sizes did not have to be proportional. However, if you take a random sample of *Canadians* you will end up with more observations from more populous provinces. In expectation the number of observations from each province will be proportional to the number of people in that province meeting the survey's criteria.

(3) (a) Check your answers against the table: make sure you got all 12 of them in Panel C: the response rates $\left(\hat{P} = \frac{X}{n}\right)$, standard errors $\left(\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right)$, and number of observations (sample sizes, n) for all four columns.

(b) Control group: $\hat{P}_1 = \frac{X_1}{n_1} = \frac{97}{6{,}648} = 0.01459$ and All treatments: $\hat{P}_2 = \frac{X_2}{n_2} = \frac{318}{13{,}594} = 0.02339$. Hence the point estimate of the effect of offering any match is 0.00880 $(= 0.02339 - 0.01459)$.

(c) $(0.02339 - 0.01459) \pm 1.96\sqrt{\frac{0.02339(0.97661)}{13{,}594} + \frac{0.01459\ (0.98541)}{6{,}648}}$
$0.00880 \pm 1.96 * 0.00196$
0.00880 ± 0.00384

Hence the margin of error is 0.00384 and the LCL = 0.0050 and UCL = 0.0126.

(d) At the time of the study, amongst previous donors to the (undisclosed) charitable organization who lived in "Red States" (Republican dominated states like Texas and Alabama), we are 95 percent confident that the (causal) effect of offering any match (1:1, 2:1 or 3:1) to donors increased the percent donating money by between 0.5 to 1.26 percentage points. While a bit noisy (interval is wide), this is clearly a huge effect as only 1.5% donate when offered no match (so even a 0.5 percentage point increase is a sizeable one).

(e) 1:1 Match: $\hat{P}_1 = \frac{X_1}{n_1} = \frac{93}{4{,}490} = 0.02071$ and 3:1 Match: $\hat{P}_2 = \frac{X_2}{n_2} = \frac{116}{4{,}547} = 0.02551$

$$(0.02551 - 0.02071) \pm 1.96\sqrt{\frac{0.02551(0.97449)}{4{,}547} + \frac{0.02071\ (0.97929)}{4{,}490}}$$

$0.00480 \pm 1.96 * 0.00316$
0.00480 ± 0.00619

Hence the margin of error is 0.00619 and the LCL = -0.0014 and UCL = 0.0110. At the time of the study, amongst previous donors to the (undisclosed) charitable organization who lived in "Red States" (Republican dominated states like Texas and Alabama), we are 95 percent confident that the (causal) effect of offering a 3 to 1 versus 1 to 1 match to donors affects the percent donating money by between -0.1 to 1.1 percentage points. In other words, we cannot even be confident that the effect of increasing the match from \$1 for every \$1 to \$3 for every \$1 has a positive effect on donors' generosity. This is consistent with the abstract from the original research shown in lecture: "Larger match ratios (i.e., \$3:\$1 and \$2:\$1) relative to a smaller match ratio (\$1:\$1) had no additional impact, however."

(4) (a) The first shows a 90% CI, the second a 95% CI, and the third a 99% CI. This is clear from the values of $z_{\alpha/2}$.

(b) As the confidence level is increased, precision goes down, which means there is a direct trade-off. To achieve greater confidence the interval will include the parameter means the margin of error (the width of the interval) goes up. In the extreme case, we can be 100% confident that any proportion is between 0 and 1 (which is not saying anything).

(c) The answer is **(E)**. Probability statements about parameters (which are constants) are nonsensical. The given probability statements about $\hat{P}$ are also nonsensical: the interval is centered at the actually sample proportion we got, hence it is in that interval with certainty. Further, it also is incorrect even if you think about taking a fresh random sample with an unknown (ahead of time) $\hat{P}$.

(d) The answer is **(C)**. Remember the population proportion is a parameter (a constant): it cannot "happen to land" somewhere like a random variable. Also, remember the CI allows us to make an inference about the unknown population proportion. Hence the "all" in the correct answer **(C)** is very important. **(A)** and **(B)** are silly: we are not making an inference about a sample. We have the sample in hand and can see it. We do not need to make *inferences* about the sample. We need to make *inferences* about what we cannot see: the population parameter.

(5) (a) It shows the results from a computer simulation. A computer drew 20 random samples from a population. We do not know the sample size (i.e. the number of observations). In real life – not a computer simulation – a researcher will not have 20 random samples but only 1 random sample: for example a poll speaks to 1,049 randomly selected Torontonians about voting intentions (one sample). However, Figure 11.3 is using a hypothetical population (exists on a computer) and drawing 20 random samples from it. (This should sound familiar as Lecture 11 and HW 11 addressed computer simulation methods.) For each of the 20 random samples the computer calculated the 95% confidence interval estimate of the population proportion. Because the population is entirely known in this hypothetical example, Figure 11.3 actually plots the population parameter: the population proportion, which is a constant (as all parameters are) and depicted as a horizontal (green) line. The figure shows that most of the 95% confidence intervals include the true parameter. However, the CI estimate from the sixth sample fails to include that correct answer: it looks like that sample had a very high sample proportion ($\hat{P}$) (i.e. because of sampling error ended up with a lot of yes answers), and the entire interval missed the target. The point is: if you are 95% confident, you will be wrong about 5% of the time. Hence it is not surprising that in 20 samples one of the samples produced a CI estimate that missed the correct answer.

(b) If the figure used a 99% confidence level (i.e. a 1% significance level) than all of the intervals would be wider (and hence more likely to include the correct answer even if there is quite a bit of sampling error).

(c) If the sample sizes were larger, TWO things would change about the figure. The most obvious is that that the width of the intervals would be narrower. However, the dots would also be generally closer to the green line: with a larger sample size $\hat{P}$ is less affected by sampling error (and the dots represent $\hat{P}$, which is at the center of the CI estimate).