ECO220Y: Homework 12

Required Exercises: Chapter 11: 3, 5, 8, 9, 11, 12, 15, 21, 23, 25, 41, 55, 61

Required Problems:

(1) A researcher tracks a random sample of 81 felons and finds that 39 of them commit another crime. Find and interpret the relevant confidence interval.

(2) When Angus Reid Strategies conducts a survey in Canada, it typically uses a sample size of around 1,000.

(a) Should Angus Reid use a larger sample in the U.S. (population about 10 times that of Canada)? Explain
- **(A)** Yes and the sample should be about 10 times as large
- **(B)** Yes but the sample does not need to be 10 times as large
- **(C)** Yes because larger populations have a higher variance the sample needs to be larger
- **(D)** No because sampling error depends only on n and not N when the 10% condition holds
- **(E)** No because the sampling distribution of the sample proportion and sample mean will be Normal given that n = 1,000 is sufficiently large

(b) But on p. 349 "Statistics Canada surveyed 10,811 households across Canada, and we can assume that the number in Ontario and Manitoba was in proportion to the populations of those provinces (0.384*10,811 = 4,151 in Ontario and 0.036*10,811 = 389 in Manitoba)." Is the book contradicting itself?

(3) Recall the Karlan and List (2007) paper in Lecture 12. Review your lecture notes. The key table of results is below.

	Control	**1:1 Ratio**	**2:1 Ratio**	**3:1 Ratio**
PANEL A: All States				
Response Rate	0.018 (0.001)	0.021 (0.001)	0.023 (0.001)	0.023 (0.001)
Observations	16,687	11,133	11,134	11,129
PANEL B: Blue States				
Response Rate	0.020 (0.001)	0.021 (0.002)	0.022 (0.002)	0.021 (0.002)
Observations	10,029	6,634	6,569	6,574
PANEL C: Red States				
Response Rate	0.015 (0.001)	0.021 (0.002)	0.024 (0.002)	0.026 (0.002)
Observations	6,648	4,490	4,557	4,547

(a) Using this cross-tabulation of the raw data for the Red States, reproduce *all 12 numbers* in "PANEL C: Red States" above.

```
. tabulate ratio gave if red0==1;
           |          gave
     ratio |         0          1 |     Total
-----------+----------------------+----------
         0 |     6,551         97 |     6,648
         1 |     4,397         93 |     4,490
         2 |     4,448        109 |     4,557
         3 |     4,431        116 |     4,547
-----------+----------------------+----------
     Total |    19,827        415 |    20,242
```

(b) Continuing with Part (a) with a focus on Red States, what is the point estimate of the effect of offering any match versus no offering no match at all?

(c) Continuing with Part (b), what is the margin of error of that point estimate? What is the lower confidence limit? Upper confidence limit?

(d) Continuing with Part (c), interpret the interval estimate.

(e) Repeat parts (b) – (d) but this time comparing the most generous matching scheme (3 to 1) with the least generous matching scheme (1 to 1) for Red States.

(4) In an academic journal article titled "Educational imposters and fake degrees," Attewell and Domina (2011) analyze data from the *National Education Longitudinal Study (NELS)*. This major database offers a representative sample of U.S. 8th graders in 1988. It reconnects with those same people again in 1994 and 2000 and asks about educational achievement. Further, the *NELS Postsecondary Education Transcript Study* obtains the higher education transcripts, directly from the institution (e.g. University of Maryland) for everyone in the *NELS* database that claims to have a higher education degree (~9,600 of the ~12,100 participants). The study requested transcripts from 3,200 institutions and obtained an institutional response rate of approximately 88%. Of the 3,343 survey respondents claiming a BA (a Bachelor of Arts, which is a four-year university degree), 185 turned out to be fake. Hence, a surprisingly high 5.53% of respondents lied (even though they had nothing to gain in a survey response) and said they had a degree from a university that never gave her/him such a degree. Below are calculations of three different confidence intervals.

$$0.0553 \pm 1.645\sqrt{\frac{0.0553*0.9447}{3{,}343}} \quad LCL = 0.049 \; and \; UCL = 0.062$$

$$0.0553 \pm 1.96\sqrt{\frac{0.0553*0.9447}{3{,}343}} \quad LCL = 0.048 \; and \; UCL = 0.063$$

$$0.0553 \pm 2.576\sqrt{\frac{0.0553*0.9447}{3{,}343}} \quad LCL = 0.\ 045 \; and \; UCL = 0.065$$

(a) What is the difference between these three confidence interval estimates?

(b) What happens to precision as the confidence level is increased? Is there a trade-off?

(c) Considering the first CI given, which of these probability statements is CORRECT? Explain.

- **(A)** P(0.049 < p < 0.062) = 0.10
- **(B)** P(0.049 < $\hat{P}$ < 0.062) = 0.10
- **(C)** P(0.049 < p < 0.062) = 0.90
- **(D)** P(0.049 < $\hat{P}$ < 0.062) = 0.90
- **(E)** None of the above

(d) Considering the third CI given at the start of this question, we are 99% confident that of_______, the proportion who have a fake BA will ________ the interval from 0.045 to 0.065. Explain.

- **(A)** the sample of 3,343; be captured by
- **(B)** the sample of 3,343; happen to land in
- **(C)** all Americans around 25 years old in 2000 claiming to have a BA; be captured by
- **(D)** all Americans around 25 years old in 2000 claiming to have a BA; happened to land in

(5) Consider Figure 11.3 from page 340 of the textbook. It is reproduced to the right.

(a) What does Figure 11.3 show?

(a) How should you expect this figure to change if the confidence level is raised? (e.g. go from 95% to 99% confidence)

(c) How should you expect figure to change if the sample size were larger?

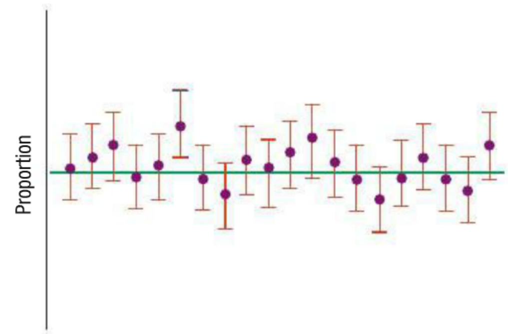

Figure 11.3 The horizontal green line shows the true proportion of people in January 2008 who thought the economy was improving. Most of the 20 simulated samples shown here produced 95% confidence intervals that captured the true value, but one missed.