ECO220Y, Test #2, June 26, 2015, 2:00 – 5:00, SOLUTIONS/RUBRICS

(1) (a) RUBRIC:

0: No evidence of any familiarity with the paper
1: Very little familiarity with the paper
2: Familiarity with the paper, but trouble identifying the main claim
3: Does a good to excellent job in identifying the main claim

(b) RUBRIC:

+1 Recognizes that the natural log straightens the scatter plot
+1 Time series data
+3 Correctly interprets the OLS slope and is context-specific (i.e. about China during the relevant time period)
+2 Correctly interprets the R^2 (noting how high it is)
+1 Correctly notes that the OLS intercept has no interpretation in this case

(c) RUBRIC:

+2 Correctly points out the similarities: need for the log transformation and the high R^2
+2 Correctly notes that the growth rate in India is less than half as fast as China during this period

(d) RUBRIC:

+3 Correctly points out that Japan's growth rate is very different during these three consecutive time periods (fast, medium, and, most recently, non-existent)
+3 Correctly notes that, just like the earlier graphs, Japan's fast growth was very steady: i.e. notes the huge R^2 in the two earlier time periods just like we see today with India and China
+3 Pulls this together to correctly point out that despite huge R^2 values extrapolating beyond the range of the data is highly inadvisable: i.e. understands why Japan is a great example to support the author's main warning about forecasting future performance based on past performance

(e) RUBRIC:

+3 Correctly interprets the OLS slope (**+1** for understanding what the regression is, even if slope interpreted incorrectly)
+3 Correctly interprets the R^2 (correctly noting that it very small)
+2 Draws a valid conclusion overall

(f) RUBRIC:

+2 Correctly interprets the OLS slope
+3 Correctly interprets the R^2 (correctly noting how it compares in size: basically zero)
+3 Draws a valid conclusion overall (including a comparison with Panel A)

(2) (a) We cannot obtain the exact probability with the given STATA summary of the population of all ON employees making between \$100,000 and \$300,000, but we can say that the chance an employee makes more than \$135,000 is between 0.10 and 0.25 and is likely closer to 0.25 than 0.10. This is obtained by looking at the 75th percentile (which is couple thousand dollars lower than \$135,000) and the 90th percentile which is well over \$135,000. (Note: Standardizing and using the Normal table is an entirely unacceptable approach for this question, which involves an extremely skewed population.)

(b) If we could use the CLT, then $P(\bar{X} > ?) = 0.01$ can be found by noting $P(Z > 2.325) = 0.01$.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

$$2.325 = \frac{\bar{X} - 125.3419}{29.96436/\sqrt{30}}$$

$$\bar{X} = 138.0613$$

This does not match the STATA summary of the Monte Carlo simulation, where the value of the 99th percentile is 139.8059, because a sample size of 30 is NOT sufficiently large (in this example) and the sampling distribution of the sample mean is still a bit positively skewed and not Normal. The 99th percentile of the sampling distribution of the sample mean is interpreted as: a sample mean of $139,806 (with a sample of 30 employees) would be surprisingly high. There is only a 1% chance of such a high sample mean occurring because of sampling error.

(3)

$$(\hat{P}_2 - \hat{P}_1) \pm z_{\alpha/2}\sqrt{\frac{\hat{P}_2(1-\hat{P}_2)}{n_2} + \frac{\hat{P}_1(1-\hat{P}_1)}{n_1}}$$

$$(0.1082 - 0.0987) \pm 1.96\sqrt{\frac{0.1082(1-0.1082)}{998} + \frac{0.0987(1-0.0987)}{618}}$$

$$0.0095 \pm 1.96 * 0.0155$$

$$0.0095 \pm 0.0304$$

Hence, a point estimate is that the callback rate is almost 1 percentage point higher for applicants with Chinese names compared to Canadian-Chinese names: 10.82% called back versus 9.87% called back. With a 95% confidence level the margin of error (ME) on this point estimate is 3.04 percentage points, which is a huge margin of error compared to the point estimate. Hence, we are 95% confident that the callback rate in the population is between 2.09 *percentage points* lower to 3.99 *percentage* points higher for those with Chinese names compared to Canadian-Chinese names. Hence, we cannot say much about whether having Canadian-Chinese name versus a Chinese name makes any difference: the point estimate suggests some difference in call back rates BUT the margin of error is so large that it is possible that there is no difference or that one or the other has a substantially higher callback rate. We need bigger sample sizes for each of these two groups to make a more precise inference.

(4) (a) A one-tailed test makes sense in this context because researchers are concerned about sex-selection in favor of males (i.e. that the proportion of boys born would be higher than the natural rate): there is no discussion of sex-selection in favor of females.

$H_0: p = 0.512$

$H_1: p > 0.512$

(b)

<u>Parity = 0:</u> $P-value = P\left(\hat{P} > \frac{934}{1815}\right) = P(\hat{P} > 0.5146) = P\left(Z > \frac{0.5146-0.512}{\sqrt{\frac{0.512(1-0.512)}{1815}}}\right) = P\left(Z > \frac{0.0026}{0.0117}\right) = P(Z > 0.22) =$

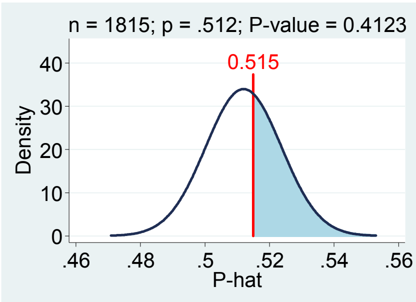
$0.5 - 0.0871 = 0.4129$

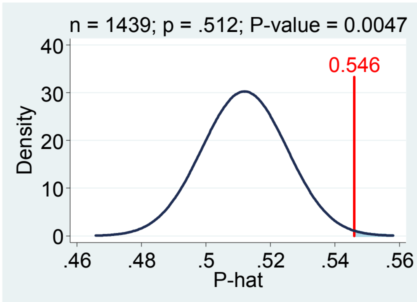
<u>Parity = 1:</u> $P-value = P\left(\hat{P} > \frac{786}{1439}\right) = P\left(Z > \frac{0.5462-0.512}{\sqrt{\frac{0.512(1-0.512)}{1439}}}\right) = P(Z > 2.60) = 0.5 - 0.4953 = 0.0047$

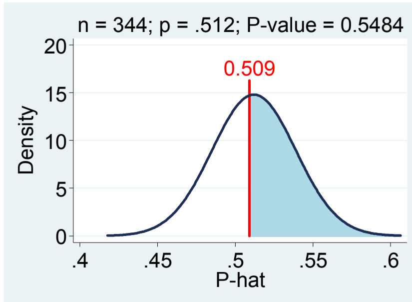
<u>Parity = 2:</u> $P-value = P\left(\hat{P} > \frac{175}{344}\right) = P\left(Z > \frac{0.5087-0.512}{\sqrt{\frac{0.512(1-0.512)}{344}}}\right) = P(Z > -0.12) = 0.5 + 0.0438 = 0.5438$

We have extremely weak evidence in favor of sex selection (the research hypothesis) for parity 0. We have strong evidence in favor of sex selection for parity 1. We have NO evidence of sex selection for parity 2 (the proportion *females* is actually a bit higher than the natural rate). Hence, overall the picture is mixed for South Korean moms: it is unclear whether or not sex selection is occurring.

Here are graphs that illustrate the answers. Note: Students are NOT required to draw graphs. Note: The numbers below are exact values from software: answers above differ (a bit) because of rounding and using the Normal table.







(5) Use the Normal approximation to the Binomial to find: $P(X > 853 \mid n = 900, p = 0.94) = ?$

$E[X] = np = 846$

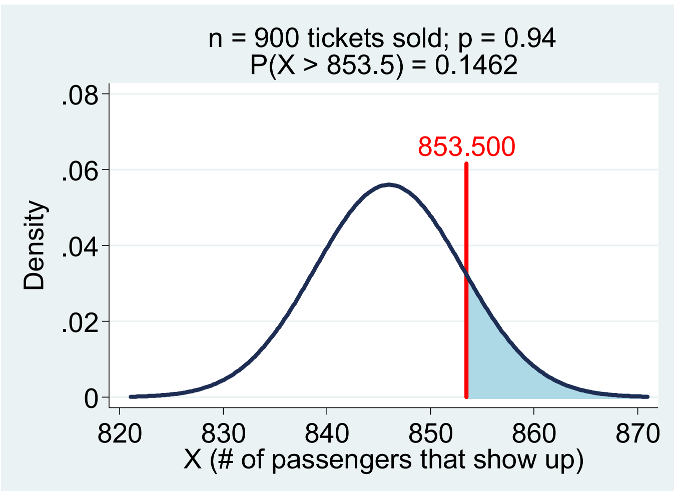
$SD[X] = \sqrt{np(1-p)} = 7.125$

$$P(X > 853.5) = P\left(Z > \frac{853.5 - 846}{7.125}\right) = P(Z > 1.05) = 0.15$$

Note: It is acceptable to ignore the continuity correction in this case, which gives a slightly less accurate answer:

$P(X > 853) = P\left(Z > \frac{853-846}{7.125}\right) = P(Z > 0.98) = 0.16.$

Note: Students may also have wrote above as a proportion question, which is fine, and yields an identical numeric answer.



There is about a 15% chance that the flight will be overbooked and one or more passengers will get bumped.

(Note: This question is adapted from an exercise in the textbook that was assigned for homework: Exercise 37 in Chapter 9.)