ECO220Y, Term Test #3: SOLUTIONS

January 19, 2018, 9:10 – 11:00 am

(1) (a) $n = 5$; $p = \frac{2,477,340}{6,858,075} = 0.36123$; Binomial probability: $P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$ for $x = 0,1,2, \dots, n$

$P(4) = \frac{5!}{4!(5-4)!} 0.36123^4 (1 - 0.36123)^{5-4} = 0.05438$

$P(5) = \frac{5!}{5!(5-5)!} 0.36123^5 (1 - 0.36123)^{5-5} = 0.00615$

$P(X > 3) = 0.05438 + 0.00615 = 0.0605$

(b) $n = 1,800$; $p = 0.474$; X is a Binomial random variable. We can use the Normal approximation because we expect at least 10 successes and at least 10 failures. $E[X] = np = 1,800 * 0.474 = 853.2$, $V[X] = 1,800 * 0.474(1 - 0.474) = 448.7832$, $SD[X] = \sqrt{448.7832} = 21.1845$.

$P(X > 825) = P\left(Z > \frac{825-853.2}{21.1845}\right) = P(Z > -1.33) = 0.5 + 0.4082 = 0.91$. Hence there is about a 91% chance that more than 825 live with their parents.

(Note: It is also acceptable to do the continuity correction to find $P(X > 825.5) = P(Z > -1.31) = 0.5 + 0.4049$.)

(Note: It is also acceptable to frame this as a sample proportion question to find $P\left(\hat{P} > \frac{825}{1,800}\right) = P(\hat{P} > 0.4583) = P\left(Z > \frac{0.4583-0.474}{\sqrt{\frac{0.474*(1-0.474)}{1,800}}}\right) = P(Z > -1.33)$, which gives the same answer as above.)

(c) Define X_V as the number living with parents in the Vancouver sample: $X_V \sim B(p = 0.386, n = 400)$ that can be approximated as Normal $X_V \sim N(\mu = 154.4, \sigma^2 = 94.8016)$. Define X_H as the number living with parents in the Hamilton sample: $X_H \sim B(p = 0.445, n = 400)$ that can be approximated as Normal $X_H \sim N(\mu = 178, \sigma^2 = 98.79)$. (Note that in both cases we easily pass any rule of thumb for Normal approximation to the Binomial.)

$P(X_V - X_H > 0) =?$ where $(X_V - X_H) \sim N(\mu = -23.6, \sigma^2 = 193.5916)$, which is obtained by using the laws of expected value and the laws of variance (for independent random variables) and the fact that a linear combination of independent Normal random variables is also Normal.

$P(X_V - X_H > 0) = P\left(Z > \frac{0 - -23.6}{\sqrt{193.5916}}\right) = P(Z > 1.70) = 0.5 - 0.4554 = 0.0446$

[*Note:* You could frame this a different in proportions question – i.e. $(\hat{P}_V - \hat{P}_H)$ – and obtain the exact same answer. This works because the sample sizes (400) are identical so asking about the count or the proportion is the same thing in this case.]

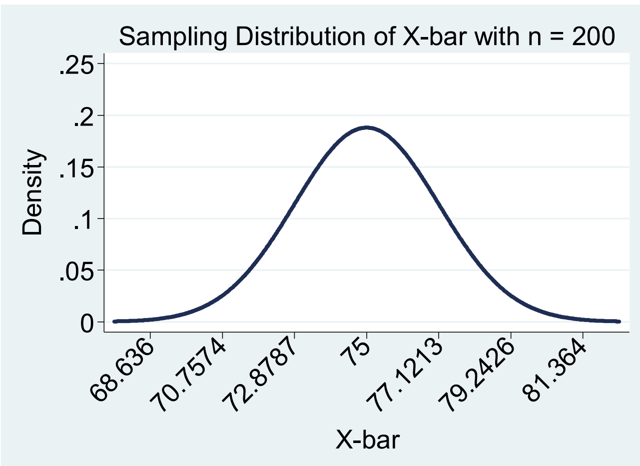
(2) (a) These are each sample statistics: the sample mean $\bar{X} = 74$ and the sample standard deviation $s = 29$.

(b) $59 \pm 2 * 35 = [-11, 129]$: given that replies are limited to 0 to 100, *all* of the observations lie within two standard deviations of the mean.

(c) $E[\bar{X}] = \mu = 75$

$SD[\bar{X}] = \frac{\sigma}{\sqrt{n}} = \frac{30}{\sqrt{200}} = 2.12$

The shape of the sampling distribution of the sample mean will be Normal (Bell shaped) according to the Central Limit Theorem as a sample size of 200 is surely sufficiently large (even though the population itself is clearly not Normal).



(d) The relevant formula for the CI estimate for a proportion: $\hat{P} \pm z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$. For the scenario where all 451 participants answered: $0.23 \pm 1.96\sqrt{\frac{0.23(1-0.23)}{451}} = 0.23 \pm 1.96 * 0.0198 = 0.23 \pm 0.0388$, which yields a LCL of 0.191 and a UCL of 0.269, which is narrower than shown in Figure 3.

[For the scenario where 90 participants answered: $0.23 \pm 1.96\sqrt{\frac{0.23(1-0.23)}{90}} = 0.23 \pm 1.96 * 0.0444 = 0.23 \pm 0.0869$, which yields a LCL of 0.143 and a UCL of 0.317, which is consistent with Figure 3.]

Other things equal, as the sample size rises the width of the CI gets narrower: we can make a more precise inference about the unknown population proportion if we have a bigger sample size because that lowers sampling error.

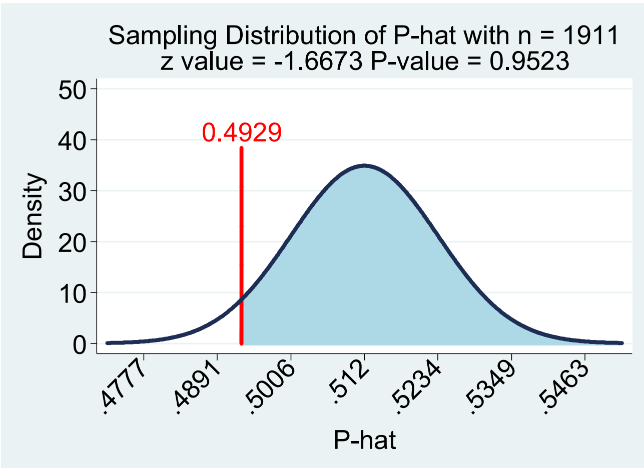
(e) We are 95 percent confident that among all people (in the sampled population), the proportion agreeing that an AV should sacrifice the life of the passenger to save 100 other lives is between 0.4 and 23.6 percentage points higher than the proportion agreeing with the sacrifice to save 20 other lives. The point estimate is 12 percentage points higher with a margin of error of 11.6 percentage points. While 12 percentage points higher (i.e. the difference between 86 percent agreeing versus 74 percent agreeing) is non-trivial, one may have expected a bigger jump given the possibility of saving five times as many lives (100 versus 20). Certainly the margin of error of 11.6 percentage points is huge and this results in a CI estimate of the difference that is very wide – ranging from virtually no difference at all in the fraction agreeing the passenger's life should be sacrificed to nearly 24 percentage points higher agreement.

(3) We can add variances in this case because it is reasonable to assume that the random variables recording the donation for each contact are independent. Also, it is important to notice that the question asks about a *total*, not a *mean*. $V[X_1 + X_2 + \cdots + X_6 + Y_1 + Y_2 + \cdots + Y_7 + W_1 + W_2 + \cdots + W_{12}] = V[X_1] + V[X_2] + \cdots + V[X_6] + V[Y_1] + V[Y_2] + \cdots + V[Y_7] + V[W_1] + V[W_2] + \cdots + V[W_{12}] = 6 * 55^2 + 7 * 49^2 + 12 * 31^2 = 46{,}489$, which then gives $SD[X_1 + X_2 + \cdots + X_6 + Y_1 + Y_2 + \cdots + Y_7 + W_1 + W_2 + \cdots + W_{12}] = \sqrt{46{,}489} = \$215.61$

(4) $H_0: p = 0.512$ versus $H_1: p > 0.512$.

A P-value of 0.952 means that the sample proportion $(\hat{P})$ of males for this particular subgroup came out *below* the natural proportion males. This large P-value means we have absolutely no evidence to support the research hypothesis in favor of sex selection in favor of males. Hence, we fail to reject the null hypothesis that there is no sex selection.

Below is a graph illustrating an example situation like this. (A graph was not required.)



(5) (a) $H_0: p = 0.5$ versus $H_1: p > 0.5$

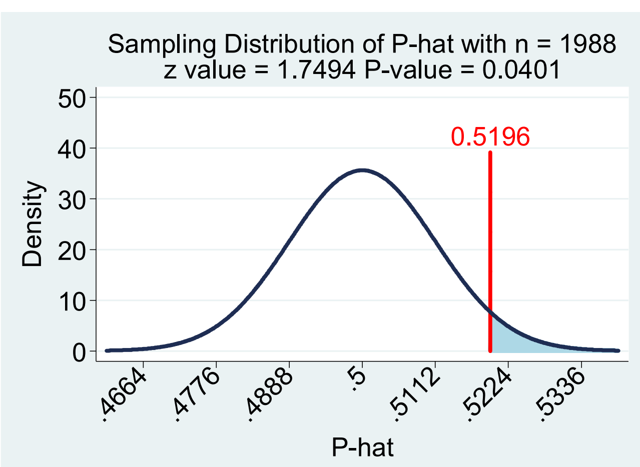
(b) Compute P-value for Scenario (1): $P-value = P\left(\hat{P} > \frac{1{,}033}{1{,}988} \mid p = 0.5, n = 1{,}988\right) = P\left(Z > \frac{0.5196-0.5}{\sqrt{\frac{0.5(1-0.5)}{1{,}988}}}\right) =$

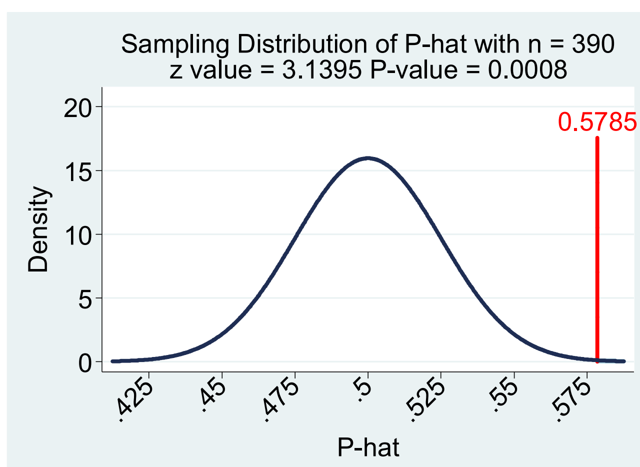
$P(Z > 1.7494) = 0.04$, which is illustrated with a graph below. (A graph was not required.)

Next, compute P-value for Scenario (2): $P-value = P\left(\hat{P} > \frac{226}{390} \mid p = 0.5, n = 390\right) = P\left(Z > \frac{0.5785-0.5}{\sqrt{\frac{0.5(1-0.5)}{390}}}\right) =$

$P(Z > 3.1395) = 0.0008$, which is illustrated with a graph below. (A graph was not required.)

Hence, Scenario (2), despite the smaller sample size, provides stronger statistical support for Mayor Tory's position because the P-value is substantially smaller.





(6) (a)

$P(Z < 2.33) = 0.99$

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

[Note: The correct values for μ and σ are provided with the histogram of the *population*, which is given in the *Supplement* for Question (6). Hence, there is no reason to use the simulated estimate of μ and the estimate of $\frac{\sigma}{\sqrt{n}}$ from the Monte Carlo simulation results. Doing so will not give you the requested *expected value* of the 99th percentile, but rather an estimate of it.]

$$2.33 = \frac{\bar{X} - 141859.79}{41434.96/\sqrt{30}}$$

$$2.33 = \frac{\bar{X} - 141859.79}{7564.954}$$

$\bar{X} = 159{,}486.1$

(b) A sample mean below \$130,000 would be more surprising. From Summary #1 we see that P($\bar{X}$ < \$130K) is somewhere between 0.01 and 0.05, whereas from Summary #2 we see that P(median > \$145K) is somewhere between 0.05 and 0.10. [Note: you must use the simulation results to answer this part because a sample size of n = 30 is not sufficiently large to use the CLT for the sample mean and you do not have any theory results for the sample median.]

(c) Summary #2: $P(median > \$170K) \cong \frac{2}{25{,}000} = 0.00008.$