ECO220Y1Y, Test #3, Prof. Murdock: SOLUTIONS

January 17, 2020, 9:10 – 11:00 am

NOTE: The parts of the solutions [in brackets] are extra explanations and are not required parts of your answer.

(1) (a) $P(OH \mid RL) = \frac{0.18}{0.18+0.30} = 0.375$

(b) $P(RH) = 0.35 + 0.17 = 0.52$

(c) $P(OL' \,\&\, RL') = P(OH \,\&\, RH) = 0.35$

(2) (a) $P(X > 50) = P\left(Z > \frac{50-58.53}{14.44}\right) = P(Z > -0.59) = 0.5 + 0.2224 = 0.7224$

(b) $P(X > ?) = 0.90$; $P(Z > -1.28) = 0.90$; $Z = -1.28 = \frac{X-58.53}{14.44}$; Hence, $X = 40.05$ $(= 58.53 - 1.28 * 14.44)$

(3) The probability should be small because the chance of upwardly mobility is *higher* in Sample 1 (60.3%) than Sample 2 (47.0%) and we're finding the probability that there are more upwardly mobile people in Sample 2, which would be surprising. In addition to being affected by the relative probabilities of upward mobility, the bigger the sample sizes the smaller the chance of this weird result: a sample size of 100 from each is fairly large so there shouldn't be a huge amount of sampling error that could lead to this surprising situation.

$X_1 \sim B(n = 100, p = 0.603)$ and $X_2 \sim B(n = 100, p = 0.470)$.

[In each case the success/failure condition passes (expect at least 10 successes and at least 10 failures): use the Normal approximation.] [Also, because $n_1 = n_2$, finding $P\left((\hat{P}_2 - \hat{P}_1) > 0\right)$ leads to the same final answer.]

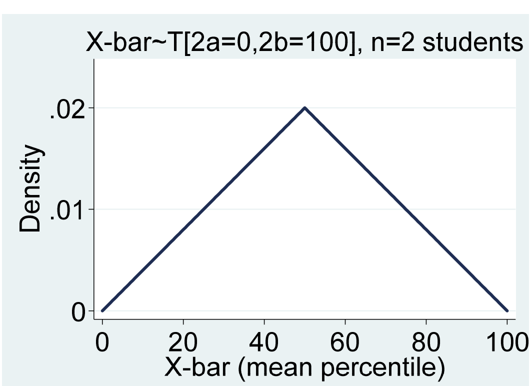
$X_1 \sim N(\mu = 60.3, \sigma^2 = 23.9391)$ and $X_2 \sim N(\mu = 47.0, \sigma^2 = 24.91)$.

$(X_2 - X_1) \sim N(\mu = -13.3, \sigma^2 = 48.8491)$.

$P\left((X_2 - X_1) > 0\right) = P\left(Z > \frac{0 - -13.3}{\sqrt{48.8491}}\right) = P(Z > 1.903) = 0.5 - 0.4713 = 0.03$

[Hence, there is about a 3% chance that there are more upwardly mobile people in Sample 2 than in Sample 1.]

(4)



(5) (a) $\frac{452*0.83+128*0.87}{452+128} = \frac{486.52}{580} = 0.84$

(b) Use a CI estimate for a proportion: $\hat{P} \pm z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$.

$0.83 \pm 1.15\sqrt{\frac{0.83(1-0.83)}{452}} = 0.83 \pm 1.15 * 0.018 = 0.83 \pm 0.02$, which yields a LCL of 0.81 and a UCL of 0.85.

(c) We need to make an inference about the difference in population proportions.

Define Group 2 to be "University": $\hat{P}_2 = 0.87$

Define Group 1 to be "Non-University PhD": $\hat{P}_1 = 0.83$

The point estimate of the difference is $\left(\hat{P}_2 - \hat{P}_1\right) = 0.04$. Next, obtain the CI estimate of the difference:

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

$$0.04 \pm 2.576\sqrt{\frac{0.87(0.13)}{128} + \frac{0.83(0.17)}{452}}$$

$$0.04 \pm 2.576 * 0.0346$$

$$0.04 \pm 0.089$$

The lower confidence limit (LCL) is -0.05 and the upper confidence limit (UCL) is 0.13. [Note that if you haven't memorized 2.576 as the critical value for a 99% confidence level, with the Normal table you can see that it is somewhere between 2.57 and 2.58. Either of those is also acceptable.]

We are 99% confident that among all start-up firms in Norway between 2000 and 2007 that include a founder with a PhD, the percent surviving at 5 years is between 5 percentage points *lower* to 13 percentage points *higher* for those founded by a PhD employed at a university compared to those start-ups founded by someone with a PhD but not a fulltime university employee. The point estimate of being 4 percentage points more likely to survive is quite modest: the difference between 87 percent surviving and 83 percent surviving. The margin of error is huge: there is a big difference between being 5 percentage points *less* likely to survive versus 13 percentage points *more* likely to survive. There is no clear advantage in the survival prospects of having a fulltime university employee involved in the founding of the start-up compared to someone with just a PhD (without the university link).

[Note: It is valid to do the entire analysis switching the definition of groups 1 and 2, so long as you are consistent.]

(6) Given that the population is Uniform, which is perfectly symmetric, a sample size of 22 is sufficiently large to apply the CLT. Further, for $X \sim U[0,30]$ we know $E[X] = \mu = \frac{0+30}{2} = 15$ and $V[X] = \sigma^2 = \frac{(30-0)^2}{12} = 75$, which means $\sigma = 8.660254$. Hence, we standardize and use the Normal table: $P(\bar{X} > 14) = P\left(Z > \frac{14-15}{8.660254/\sqrt{22}}\right) = P(Z > -0.54) = 0.5 + 0.2054 = 0.71$.

(7) (a) For a much larger sample size, we would expect it to be even closer to the population 75th percentile, which is \$161,838. Because of sampling error, the sample in Summary #2 happens to have a 75th percentile that is about \$1,680 too small: \$160,157.8 (sample statistic) versus \$161,838 (population parameter).

(b) $P(Z < 2.33) = 0.99$

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

$\frac{\bar{X}-141859.8}{41435.85/\sqrt{25}} = 2.33$

$\bar{X} = 141859.8 + 2.33 * 8287.2 = 161{,}168.5$

[Hence, the 99th percentile of the sampling distribution of X-bar would be \$161,168.5 if the shape were Normal rather than \$163,964.3.]

(c) Answering requires looking at the simulated sampling distribution of X-bar for a sample size of 25 given in **<u>Summary #3</u>**. A sample mean of \$146,968.9 ($\bar{X}$), which is more than \$5,000 above the population mean of \$141,859.8 (μ), *can be plausibly explained* by sampling error. In Summary #3, the 75th percentile is \$147,051.5, which means that there is over a 25 percent chance of a sample mean of \$146,968.9 or higher: $P(\bar{X} \geq 146{,}968.9) > 0.25$. In other words, there is more than a 25% chance of observing such a big sample mean (that big or even bigger) simply because of sampling error, which makes sampling error an entirely *plausible* explanation for such a large sample mean as observed in Summary #2.

[Note: You <u>cannot</u> make a valid argument based on the standard deviation in Summary #3 and the Empirical Rule because, in this case, the sampling distribution of $\bar{X}$ is not Normal. Recall from part (b) that a sample size of 25 is <u>not</u> sufficiently large in this case so we cannot use the CLT.]

(8) $\hat{P} = \frac{2{,}081}{2{,}081+1{,}926} = \frac{2{,}081}{4{,}007} = 0.519341$

$$P(\hat{P} > 0.519341) = P\left(Z > \frac{0.519341 - 0.512}{\sqrt{\frac{0.512(1-0.512)}{4{,}007}}}\right) = P\left(Z > \frac{0.007341}{0.007897}\right) = P(Z > 0.93) = 0.5 - 0.3238 = 0.18$$

A P-value of 0.18 provides weak evidence for the research hypothesis (sex selection) and it is not sufficient to reject the null (no sex selection). Hence, we have an inconclusive result: we can neither conclude that there is or is not sex selection for males in Ontario for the third child of moms born in Pakistan.

(9) (a) larger than

(b) smaller than

(c) smaller than

(d) larger than

(e) equal to