Term Test #3, ECO220Y1Y, January 30, 2015, Prof. Murdock: SUGGESTED SOLUTIONS

(1) (a) As the sample size rises from 20 to 100 the sampling error (aka sampling variability or sampling noise) on the sample median, a sample statistic, will decrease. Hence the sampling distribution, which the Monte Carlo results simulate, should be less spread out. Hence the 5th percentile will increase (be less extreme in the left tail), which means (A) is the correct choice. We would not expect any change in the 50th percentile (the center of the sampling distribution) as the change in sample size affects the spread not the central tendency, which is why (B) is incorrect. We would expect the 95th percentile, the standard deviation and the interquartile range in the STATA summary to all DECREASE as the sampling distribution would be less spread out, which is why (C) – (E) are incorrect answers to the question.

(b) $P(\bar{X} < 4) = ?$

$X \sim U[0, 10]$

$$E[X] = \mu = \frac{a+b}{2} = \frac{0+10}{2} = 5$$

$$V[X] = \sigma^2 = \frac{(b-a)^2}{12} = \frac{(10-0)^2}{12} = 8.3333$$

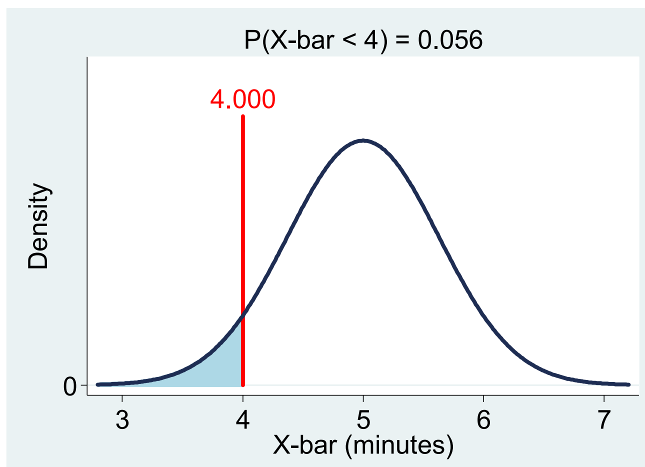
$$E[\bar{X}] = \mu = 5$$

$$V[\bar{X}] = \frac{\sigma^2}{n} = \frac{8.3333}{21} = 0.3968$$

$$SD[\bar{X}] = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{8.3333}}{\sqrt{21}} = 0.6299$$

Given that the population is Uniformly distributed, which is not too far from Normal, a sample size of 21 is sufficiently large to apply the Central Limit Theorem (CLT): the sampling distribution of $\bar{X}$ will be Normal.

$$P(\bar{X} < 4) = P\left(Z < \frac{4-5}{\sqrt{0.3968}}\right) = P(Z < -1.59)$$
$$= 0.5 - 0.4441 = 0.0559$$



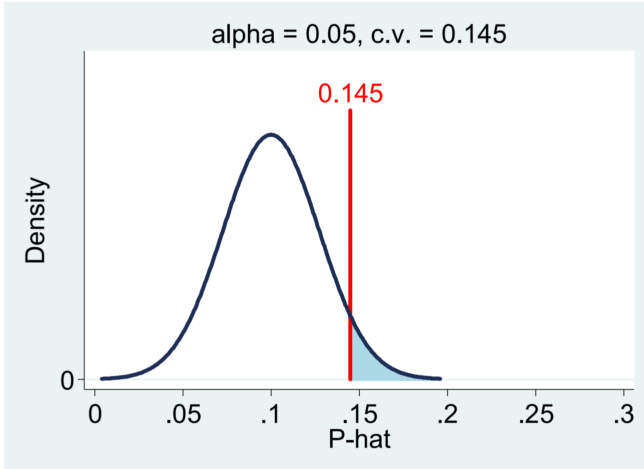
(2) (a)

$H_0: p = 0.10$

$H_1: p > 0.10$

(b) The significance level is the maximum chance of a Type I error I will tolerate, which in this case would be wrongfully terminating a grader whose error rate is not above 10% (i.e. inferring that the grader has an error rate above 10% when in fact s/he does not). This would suggest a low α. However, this must be weighed against a Type II, which would mean failing to terminate an error-prone grader. The chance of a Type II error increases as α is decreased. The choice of the significance level reflects a personal decision in light of these competing concerns. Personally, I would probably choose a significance level of $\alpha = 0.05$ as I am concerned about both Type I errors and Type II errors as I would want to use the most accurate graders to ensure fairness to the students (which means not firing good graders and not failing to fire bad graders).

(c) Find the rejection region:



$P(Z > z_\alpha) = \alpha$

$P(Z > 1.645) = 0.05$

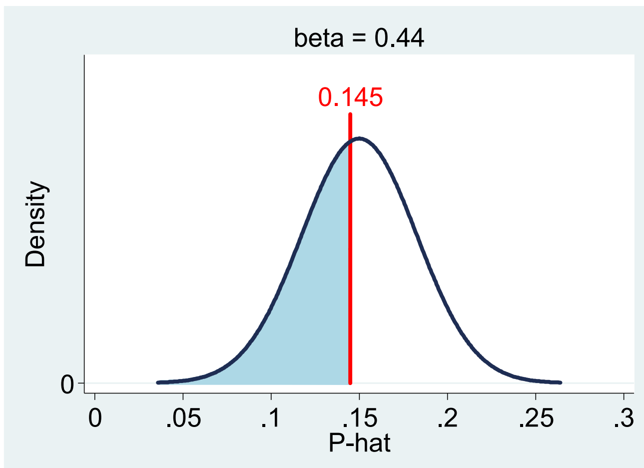
$$Z = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$1.645 = \frac{\hat{P} - 0.10}{\sqrt{\frac{0.10(1-0.10)}{120}}} \xrightarrow{yields} c.v. = 0.145$$

$$\left(Note: SD[\hat{P} \mid H_0] = \sqrt{\frac{0.10(1-0.10)}{120}} = 0.027\right)$$

Hence, the unstandardized rejection region is $(0.145, \infty)$ (which could also be written as $(0.145, 1)$).

Next, find the probability of making a Type II error (failing to reject the false null hypothesis). In other words, what is the chance the sample proportion of marking errors for a grader with an overall error rate of 15% does not end up in the rejection region?



$P(\hat{P} < 0.145 \mid p = 0.15, n = 120) = ?$

$$P(\hat{P} < 0.145) = P\left(Z < \frac{0.145 - 0.15}{\sqrt{\frac{0.15(1-0.15)}{120}}}\right)$$
$$= P(Z < -0.1534) = 0.5 - 0.06 = 0.44 = \beta$$

$$\left(Note: SD[\hat{P} \mid p = 0.15] = \sqrt{\frac{0.15(1-0.15)}{120}} = 0.033\right)$$

Hence, the chance that the coordinator will fail to dismiss the incompetent grader is 0.44, which is a pretty high chance of a Type II error. In other words, the power $(0.56 = 1 - \beta)$ is not particularly high. (Note: If $\alpha = 0.01$ is chosen, then the c.v. 0.164 is, β is 0.67 and power is 0.33. If $\alpha = 0.10$ is chosen, then the c.v. is 0.135, β is 0.32 and power is 0.68.)

Term Test #3, ECO220Y1Y, January 30, 2015, Prof. Murdock: SUGGESTED SOLUTIONS

(3) (a)

$H_0: p_1 - p_2 = 0$

$H_1: p_1 - p_2 \neq 0$

$\hat{P}_1 = 0.28; \hat{P}_2 = 0.55$

$$\bar{P} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{0.28 * 400 + 0.55 * 613}{400 + 613} = 0.443$$

$$z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_1} + \frac{\bar{P}(1-\bar{P})}{n_2}}} = \frac{0.28 - 0.55}{\sqrt{0.443(1-0.443)\left(\frac{1}{400} + \frac{1}{613}\right)}} = -8.46$$

This meets any plausible significance level (including $\alpha = 0.01, 0.05, 0.10$) as the P-value ($= 2 * P(Z < -8.46)$) is so small that it is basically zero. Further, the difference is definitely significant as it is both statistically significant *and* a large difference that people would certainly care about: a 27 percentage point difference.

(b) Start by showing where -2.45 came from (NOTE: students do NOT need to show this part):

$H_0: p_1 - p_2 = 0$

$H_1: p_1 - p_2 \neq 0$

$\hat{P}_1 = 0.28; \hat{P}_2 = 0.55$

$$\bar{P} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{0.28 * 40 + 0.55 * 40}{40 + 40} = 0.415$$

$$z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_1} + \frac{\bar{P}(1-\bar{P})}{n_2}}} = \frac{0.28 - 0.55}{\sqrt{0.415(1-0.415)\left(\frac{1}{40} + \frac{1}{40}\right)}} = -2.45$$

Next, evaluate $z = -2.45$. (NOTE: students DO need to show this part):

Option 1: P-value approach:

$$P - value = P(Z < -2.45) + P(Z > 2.45) = 2 * (0.5 - 0.4929) = 0.0142$$

This P-value is smaller than $\alpha = 0.05$ (and $\alpha = 0.10$) but it does not meet the burden of proof of $\alpha = 0.01$.

Option 2: Rejection region approach:

For $\alpha = 0.01$ the (standardized) rejection region is $(-\infty, -2.575)$ and $(2.575, \infty)$ (or can use 2.57 or 2.58)

For $\alpha = 0.05$ the (standardized) rejection region is $(-\infty, -1.96)$ and $(1.96, \infty)$

For $\alpha = 0.10$ the (standardized) rejection region is $(-\infty, -1.645)$ and $(1.645, \infty)$ (or can use 1.64 or 1.65)

The test statistic (-2.45) is in the rejection region for $\alpha = 0.05$ (and $\alpha = 0.10$) but it does not meet the burden of proof of $\alpha = 0.01$.

This meets a 5% and 10% significance level but not a 1% significance level. The difference remains economically significant (large). However, with the smaller sample sizes the evidence in favor of the research hypothesis is weaker (because there is more sampling error). Most people would still call this a significant difference as it is large and is statistically significant at the conventional 5% level. (The difference is so huge that even with the small sample sizes we still obtain a statistically significant difference at a 5% level.)

(4) (a)

$H_0: p_1 - p_0 = 0$

$H_1: p_1 - p_0 < 0$

$\hat{P}_0 = \frac{429}{3{,}026} = 0.1418; \hat{P}_1 = \frac{363}{3{,}615} = 0.1004$

$$\bar{P} = \frac{X_0 + X_1}{n_0 + n_1} = \frac{429 + 363}{3{,}026 + 3{,}615} = 0.1193$$

$$z = \frac{\hat{P}_1 - \hat{P}_0}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_1} + \frac{\bar{P}(1-\bar{P})}{n_0}}} = \frac{0.1004 - 0.1418}{\sqrt{0.1193(1-0.1193)\left(\frac{1}{3{,}615} + \frac{1}{3{,}026}\right)}} = -5.18$$

Use the P-value approach to give a quantitative measure of the strength of the evidence.

$P - value = P(Z < -5.18) \cong 0$

The evidence is overwhelming that callback rates are lower for applicants with non-English names at any plausible significance level.

(b) Because this is a randomized field experiment, we can conclude that it is the non-English name that is causing the lower callback rate. Other factors observable from the resume were randomly assigned and hence are not correlated with the applicant's name and cannot explain the lower callback rate. For example, it could *not* be that there were systematic differences in education or work experience that explain the discrepancy. If these had been observational data then these alternate explanations would have been possible.

(c)

$$(\hat{P}_1 - \hat{P}_0) \pm z_{\alpha/2}\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_0(1-\hat{P}_0)}{n_0}}$$

$$(0.1004 - 0.1418) \pm 1.96\sqrt{\frac{0.1004(1-0.1004)}{3{,}615} + \frac{0.1418(1-0.1418)}{3{,}026}}$$

$$-0.0414 \pm 1.96 * 0.0081$$

$$-0.0414 \pm 0.0158$$

Hence, a point estimate is that the callback rate is 4.14 percentage points lower for applicants with non-English names other things equal: 10.04% called back versus 14.18% called back. With a 95% confidence level the margin of error (ME) on this point estimate is 1.58 percentage points. Hence, we are 95% confident that the callback rate in the population is between 2.26 and 5.72 *percentage points* lower for those with non-English names, which is a huge difference.

(5) $\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$

$$5.511 \pm 1.680\frac{1.502}{\sqrt{45}}$$

$$5.511 \pm 1.680 * 0.224$$

$$5.511 \pm 0.376$$

$$LCL = 5.135$$

$$UCL = 5.887$$

We are 90% confident that the mean rating of this driver on a 10 point scale among ALL passengers (we are making an inference about the *population* mean) lies between 5.135 and 5.887.