ECO220Y1Y, Test #2, Prof. Murdock SOLUTIONS

(1) (a) Define Event U as being unemployed, Event H as being a high school graduate, and Event N as having no degree.

$P(U \mid H) = \frac{P(U \,\&\, H)}{P(H)} = \frac{0.0075}{0.1544} = 0.049$

$P(U \mid N) = \frac{P(U \,\&\, N)}{P(N)} = \frac{0.0036}{0.0579} = 0.062$

For those aged 25 to 54 years in Canada in 2022, among HS graduates 4.9% are unemployed whereas among those with no degree 6.2% are unemployed, which is a considerable 1.3 percentage points more for the least educated group.

(b) If they were *independent*, we use the multiplication rule for independent events, $P(N \,\&\, U) = P(N) * P(U) = 0.0579 * 0.0384 = 0.0022$ (instead of 0.0036).

(2) (a) $P(X > 6.4) = P\left(Z > \frac{6.4-5.2}{1.6}\right) = P(Z > 0.75) = 0.5 - 0.2734 = 0.2266$, which means 22.7%.

(b) $P(X < ?) = 0.10$

$P(Z < -1.28) = 0.10$

$\frac{? - 5.2}{1.6} = -1.28$

$? = 3.152$, which is the 10th percentile of risk seeking

(c) First, compute the Summer 2023 sample mean:

$\bar{X} = \frac{0*7 + 2*2 + 3*5 + 4*4 + 5*8 + 6*10 + 7*14 + 8*6 + 9*2 + 10*2}{60} = 5.3167$

$P(\bar{X} > 5.3167 \mid \mu_{\bar{X}} = 5.2, \sigma_{\bar{X}} = \frac{1.6}{\sqrt{60}} = 0.2066) = P\left(Z > \frac{5.3167-5.2}{0.2066}\right) \approx P(Z > 0.56) \approx 0.5 - 0.2123 = 0.29$

Hence, sampling error is a plausible explanation because there is about a 30% chance of such a high sample mean.

(3) $V\left[\frac{R+C}{2}\right] = V[0.5R + 0.5C] = 0.25V[R] + 0.25V[C] + 2 * 0.5 * 0.5 * SD[R] * SD[C] * CORR[R, C]$

$= 0.25(2.1073)^2 + 0.25(2.2093)^2 + 0.5 * 2.1073 * 2.2093 * 0.5137 = 3.5262$

$SD[NewVar] = \sqrt{3.5262} = 1.88$

(4) (a) $P(A3 \mid E) = \frac{P(A3\ \&\ E)}{P(E)} = \frac{P(E \mid A3)*P(A3)}{P(E)} = \frac{0.303*0.395}{0.583} = 0.205$

(b) Use Binomial probability formula: $p(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$

$P(X \geq 2 \mid n = 120,\ p = 0.005) = 1 - P(X = 0) - P(X = 1)$

$P(X = 0) = \frac{120!}{0!(120-0)!} 0.005^0 (1 - 0.005)^{120-0} = (0.995)^{120} = 0.5478$

$P(X = 1) = \frac{120!}{1!(120-1)!} 0.005^1 (1 - 0.005)^{120-1} = 120 * 0.005^1 (1 - 0.005)^{120-1} = 0.3304$

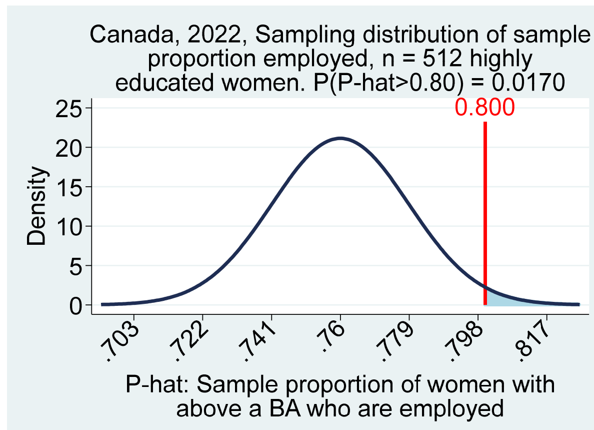
$P(X \geq 2) = 1 - 0.5478 - 0.3304 = 0.1218$

(c) $P(\hat{P} > 0.80 \mid n = 512,\ p = 0.760) =$

$P\left(Z > \frac{0.80-0.760}{\sqrt{\frac{0.760(1-0.760)}{512}}}\right) = P(Z > 2.12) = 0.5 - 0.4830 =$

0.017

Hence, there is only a 1.7% chance of observing such a high percent employed in the sample, which is fairly unlikely so sampling error is not a particularly plausible explanation, but it could happen.



(5) (a) Use $\hat{P} \pm z_{\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$

$0.37 \pm 2.576 \sqrt{\frac{0.37(1-0.37)}{24{,}599}}$

$0.37 \pm 2.576 * 0.0031$

0.37 ± 0.008

We are 99% confident that between 36 and 38 percent of *all* people living in these 12 high-income countries (Australia to the US) are substantially willing to limit driving as of 2021/22. This is an extremely precise estimate – the margin of error is less than 1 percentage point – because of the huge sample size of nearly 25,000 people living in these high income countries.

(b) The CI estimate from Part (a) is for all people in all twelve countries overall. But, as Table 7 in the *Supplement* shows, there are substantial differences across countries if we do not pool all responses together, which means there is no reason to expect 99% of the sample proportions across countries to be in that interval.

(6) (a) For the charitable giving, direct-mail study of Karlan and List (2007), we are 90% confident that offering a match with a low example amount causes between a 0.05 to 0.6 percentage point increase in the response rate compared to the standard letter with no match. About 1.8% of people receiving the standard letter respond by donating to the charity, so a 0.05 percentage point increase is quite small, but a 0.6 percentage point increase is quite big. We are not able to make a precise inference about the size of the effect.

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

$$(0.02273 - 0.02120) \pm 1.44\sqrt{\frac{0.02273(1-0.02273)}{11{,}129} + \frac{0.02120(1-0.02120)}{11{,}134}}$$

$$0.00152 \pm 1.44 * 0.00196$$

$$0.0015 \pm 0.0028$$

$LCL = -0.001$ and $UCL = 0.004$

(c) While offering a match does cause a higher response rate than no match (i.e. the ordinary letter), the various strategies around how the match is offered – whether a maximum grant threshold for matches is specified and how big it is and what kind of donation is used to illustrate the match (i.e. less generous versus more generous) – does not seem to have any measurable impact on the response rate. While Column (1) differs from the others, Columns (2) through (8) are very similar to each other.

(7) (a) In formal notation, 124.9381 is **$\underline{\mu}$**, and 35.81888 is **$\underline{\sigma}$**, and 119.9348 is **$\underline{\bar{X}}$**, and 31.2389 is **$\underline{s}$**. Compared to the random sample with n = 30 in the *Supplement*, if we drew another random sample with n = 30, we expect the sample mean to **<u>be higher</u>**, the sample standard deviation to **<u>be higher</u>**, and the sample 75th percentile to **<u>be higher</u>**.

(b) For n = 30, the probability of getting a sample mean below $107,000 is about **<u>3/200,000</u>** and the probability of getting a sample mean above $140,000 is somewhere between **<u>0.01</u>** and **<u>0.05</u>**. Next, **if n = 10 instead of n = 30**, then we expect the value in the spot 117.303 to **<u>decrease</u>**, the value in the spot 187.3988 to **<u>increase</u>**, the value in the spot 42.74376 to **<u>increase</u>**, and the value in the spot 124.9463 to **<u>barely change</u>**.