ECO220Y1Y, Test #3, Prof. Murdock SOLUTIONS

(1) One example would be tracking a random sample of people who have had a change in their insurance status, where for each person you record their use of the emergency department when they have insurance and their use when they do not have insurance. Each person would be paired with themselves. [Alternatively, people could be paired by some other observable such as previous use of the emergency department: first person in Sample 1 used the ED four times in the previous period and first person in Sample 2 used the ED four times in the previous period and so on. Sample 1 gets insurance and Sample 2 does not.]

(2) (a)

$H_0: p = 0.10$ [this can also be correctly written as $H_0: p \geq 0.10$]

$H_1: p < 0.10$

$P(Z < -2.326) = 0.01$

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

$-2.326 = \frac{\widehat{P_{c.v.}} - 0.1}{\sqrt{\frac{0.1(1-0.1)}{1{,}000}}} => -2.326 = \frac{\widehat{P_{c.v.}} - 0.1}{0.009487} => \widehat{P_{c.v.}} = 0.078$

Hence, we would need to see 7.8% or less of the sample of black applicants get a job offer to prove discrimination at a 1% level.

(b) We would need to see an even more extreme low percent of the black applications being successful – for example, less than 6% instead of less than 8% – because sampling error would be higher.

(3) (a)

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

$\hat{P} = \frac{240}{83+32+240} = \frac{240}{355} = 0.676$

$0.676 \pm 1.645 \sqrt{\frac{\frac{240}{355}\left(1 - \frac{240}{355}\right)}{355}}$

$0.676 \pm 1.645 * 0.0248$

0.676 ± 0.041

$LCL = 0.635$

$UCL = 0.717$

(b)

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

$(0.80 - 0.55) \pm 2.576\sqrt{\frac{0.8(1-0.8)}{1{,}100} + \frac{0.55(1-0.55)}{2{,}200}}$

$0.25 \pm 2.576 * 0.0161$

0.25 ± 0.04

$LCL = 0.21$

$UCL = 0.29$

We are 99% confident that in 2019 international economic experts are between 21 and 29 percentage points more likely to think that a rise in government spending will lead to a rise in inflation compared to the general US population. [This is a very large difference that is quite precisely estimated.]

(4)

$H_0: \mu = 2.1$ [this can also be correctly written as $H_0: \mu \leq 2.1$]

$H_1: \mu > 2.1$

A Type 1 error – rejecting a true null hypothesis – happens if the fertility of this population of women is at or below an adequate replacement rate, which at best yields population stability (or yields population decline), but we incorrectly conclude that their fertility supports population growth.

A Type 2 error – failing to reject a false null hypothesis – happens when the fertility of this population of women is above an adequate replacement rate, but we are unable to prove it.

(5) (a)

$H_0: \mu = 1.6$ [this can also be correctly written as $H_0: \mu \leq 1.6$]
$H_1: \mu > 1.6$
$t = \frac{\bar{X}-1.6}{1.3474759/\sqrt{1{,}216}}$
$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{277*0+211*1+412*2+195*3+91*4+15*5+13*6+2*7}{1{,}216} = 1.768914$
$t = \frac{1.768914-1.6}{1.3474759/\sqrt{1{,}216}} = 4.37$

(b)

$H_0: \mu = 1.4$ [this can also be correctly written as $H_0: \mu \geq 1.4$]

$H_1: \mu < 1.4$

$t = \frac{1.2-1.4}{1.1/\sqrt{54}} = -1.34$

$\nu = 54 - 1 = 53$

Using the Student t table for the row for 53 degrees of freedom, the P-value is between 0.05 and 0.10 because the test statistic is between 1.298 and 1.675.

Hence, we have sufficient evidence for a 10% significance level to prove that the subpopulation of women aged 50 to 55 who completed law school have fertility that falls below the U.N.'s ultralow fertility line of 1.4 births per women. [We do not know the effects on the overall population decline/growth because we do not know what fraction of women fall into this subpopulation, but it is presumably small as not that many people complete law school.]

(c) The researchers seek to prove that fertility for women aged 50 to 55 with a PhD is less than 0.78, which is South Korea's fertility rate. Obtaining a P-value above 0.5 means that the evidence in the sample contradicts this. They <u>cannot</u> prove that this subgroup of women has fertility below 0.78 births per woman with a random sample that have a fertility rate *above* 0.78 births per woman.

(6)

$H_0: p = 0.5$ [this can also be correctly written as $H_0: p \leq 0.5$]

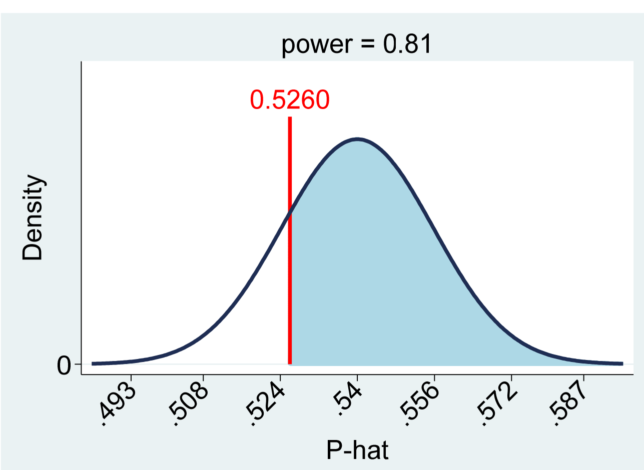
$H_1: p > 0.5$

$$\beta = P(Type\ II\ error) = P(\hat{P} < 0.5260 \mid n = 1{,}000, p = 0.54)$$

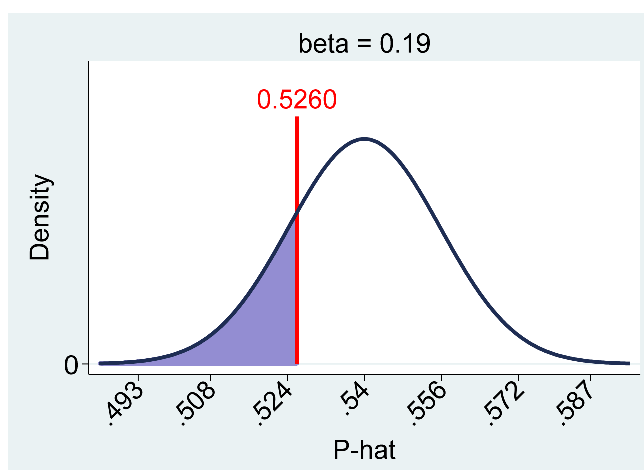
$$\beta = P\left(Z < \frac{0.5260-0.54}{\sqrt{\frac{0.54(1-0.54)}{1{,}000}}}\right) = P(Z < -0.89) = 0.5 - 0.3133 = 0.1867$$

$$Power = 1 - \beta = 0.8133$$

Hence, if 54% of this subpopulation have one child, there is an 82% chance that we will successfully be able to prove that a majority (at least 50%) have one child if we collect a random sample of 1,000 women and wish to achieve a 5% significance level.



OR



(7) (a) Are households that adopt solar more likely to be all-Democratic?

[It must be a yes/no question, and it must be about a difference in proportions (not means)] [Note: A closely related answer – "Are all-Democratic households more likely to adopt solar?" – is WRONG because that information is not provided by Table 1: it compares adopters and non-adopters not all-Democratic households with other types of households.]

(b) How large is the difference in average wealth (in $1,000s) between households that adopt solar versus households that do not?

[It must be a size of a difference question, and it must be about a difference in means (not proportions)]

(c)

$$SE[\bar{X}_{MW}] = \frac{s_{MW}}{\sqrt{n_{MW}}} = \frac{2.34}{\sqrt{0.42*7{,}244}} = 0.042$$

$$SE[\bar{X}] = \frac{s}{\sqrt{n}} = \frac{2.36}{\sqrt{7{,}244}} = 0.028$$

Even though the standard deviations of 2.36 kW and 2.34 kW are nearly identical, the standard error for the subgroup of medium wealth households (MW) will be much bigger than for households of all wealth levels combined because the latter has a much bigger sample size, which means less sampling error.

(d)

$$H_0: (\mu_A - \mu_N) = 0$$

$$H_1: (\mu_A - \mu_N) \neq 0$$

[where A stands for households that are solar adopters, and N stands for non-adopters]

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(2.16 - 1.77) - 0}{\sqrt{\frac{0.81^2}{7{,}244} + \frac{0.75^2}{176{,}423}}} = \frac{0.39}{0.00968} = 40.3$$

[Given the large sample sizes the degrees of freedom are approximately infinity.]

This t test statistic is huge, and the P-value equals 0.

There is an extremely statistically significant difference in the size of the house between adopters and non-adopters, meeting any significance level, which is not surprising given the huge sample sizes. Further, an average difference of 390 square feet is large: the house size of adopters is 22% (=100*390/1770) larger than non-adopters. Hence, there is a significant difference: the result is both statistically and economically significant.