ECO220Y1Y, Test #3, Prof. Murdock SOLUTIONS

(1) (a)

$H_0: (p_W - p_M) = 0$

$H_1: (p_W - p_M) \neq 0$

(b) A Type I error would be concluding that there is a difference between women and men in the fraction willing to tolerate high risk when there really is NO difference by gender.

(c) A Type II error would be when there really is a difference between women and men in the fraction willing to tolerate high risk, but the researcher is unable to prove that difference exists.

(2) (1) Imagine that among all citizens 25 percent support the government; (2) Increase the sample size to 2,000; (3) Prove that less than one-half of the population support the government; (4) Use a 10% significance level

(3) The relevant formula is $\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$ with degrees of freedom: $\nu = n - 1$. The sample size is NOT 11,133 because it is conditional on giving and only 2.1% gave in the treatment group that got a letter with a 1 to 1 match ratio. Hence the relevant sample size is approximately $n \approx 234$, meaning the degrees of freedom are 233. We can be conservative and use 2.601 from the Student t table (although 2.596 is also acceptable).

$45.143 \pm 2.601 * 3.099$ [Note that the table already reports the *standard error*: it is incorrect to divide it by root n.]

45.14 ± 8.06

$LCL = 37.1$ and $UCL = 53.2$

We are 99% confident that for those who receive a letter offering a 1-to-1 match ratio, the mean donation among all potential donors is between $37 and $53 for those that chose to donate (a non-zero amount).

(4)

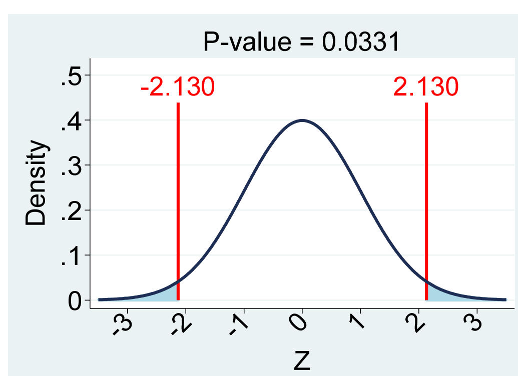
$H_0: (p_C - p_T) = 0$

$H_1: (p_C - p_T) \neq 0$

$$z = \frac{\hat{P}_2 - \hat{P}_1}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_1} + \frac{\bar{P}(1-\bar{P})}{n_2}}}$$

$$\bar{P} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{107 + 278}{6{,}648 + 13{,}594} = \frac{385}{20{,}242} = 0.019020$$

$$z = \frac{\frac{107}{6{,}648} - \frac{278}{13{,}594}}{\sqrt{\frac{0.01902(1-0.01902)}{6{,}648} + \frac{0.01902(1-0.01902)}{13{,}594}}} = \frac{-0.004355}{0.002044} = -2.13$$



$P - value = P(z > 2.13) + P(z < -2.13) = 2 * (0.5 - 0.4834) = 0.033$ [With this small P-value we *have* proven at a 5% significance level, but not a 1% level, that the fraction donating differs depending on the letter format.]

(5) (a) At Boise State University, compared to 2000-01, the percent of students choosing to take a course again (repeat it) is about 1.25 percentage points higher in 2010-11 (about 3.75 percent versus about 5 percent), which is a large increase of 33 percent corresponding to when grade forgiveness is in effect. Compared to 2000-01, average cumulative GPA out of 4.0 is about 0.2 higher in 2010-11 (about 2.58 versus about 2.78), which is a considerable increase of about 8 percent corresponding to when grade forgiveness is in effect.

(b)

$H_0: (\mu_{10/11} - \mu_{00/01}) = 0$

$H_1: (\mu_{10/11} - \mu_{00/01}) > 0$

(6)

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

$\frac{1{,}772}{2{,}407} \pm 1.960 \sqrt{\frac{\frac{1{,}772}{2{,}407}\left(1-\frac{1{,}772}{2{,}407}\right)}{2{,}407}}$

$0.7362 \pm 1.960 * 0.00898$

0.7362 ± 0.0176

$LCL = 0.7186$ and $UCL = 0.7538$

We are 95% confident that Human B, an unnamed archaeologist and expert, can correctly classify between 71.9% and 75.4% of all pottery fragments.

(7) (a) Across the 5,268 students in the 39 primary schools in Malawi, the standard deviation of the difference between their math and Chichewa (language) scores is 21.7 percentage points. There is a lot of variation: some students are doing much better in math and others much better in the local language even though on average they do worse in math.

(b) Use $\bar{X}_d \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$ with degrees of freedom: $\nu = n - 1$

$-2.0251 \pm 1.645 \frac{19.5205}{\sqrt{5{,}268}}$ with $\nu = 5{,}267$

$-2.0251 \pm 1.645 * 0.2689$

-2.0251 ± 0.4424

The margin of error of 0.4424 is small: it is less than one half of one percentage point. We can make a very accurate inference about the difference in believed math versus Chichewa scores given the huge sample size of 5,268 children.

(8) (a) Among those that continue to not have Medicaid health insurance, in the period from March 2008 through September 2009 in the state of Oregon, about 34.5 percent have at least one visit to the emergency department, which is quite high.

(b) This is the standard error of the difference in two sample means with independent samples. Its size depends on the standard deviation the number of visits in the control group, the standard deviation of the number of visits in the treatment group, and the sample size of each group.

(c) On average people who won Medicaid insurance coverage had about 0.41 more emergency department (ED) visits during the period from March 2008 to September 2009 compared to those without coverage. The first column says that those with coverage are 7 percentage points more likely to visit the ED one or more times. That means they are 20.3% more likely to visit (100*7.0/34.5), but the average number of visits is 39.9% higher (100*0.408/1.022). [Note that the standard errors and P-values, which assess statistical significance and not economic significance, are NOT relevant for a correct answer.]

(9) (a) $a = 22.8, b = 9.5, b_0 = 32.3, b_1 = -9.5$ [Note: We do not expect approximations that are this accurate: if you are within plus/minus 0.7 of these, that is a very good approximation.]

(b) The s_e measures the amount of scatter and the best estimate is 11.739 MMBTUs, which are the same units as the dependent variable (electricity usage in MMBTUs). It is an estimate of the standard deviation of the residuals around zero. Given the extensive scatter, the line underpredicts usage by nearly 50 MMBTUs in some cases and overpredicts by over 20 MMBTUs in other cases, 0.019 and 0.384 are far too small. The s_e is very large: while the size of the house is related with electricity usage, it is weak predictor. This simple regression far over and underpredicts electricity usage for many homes.