Homework 15: ECO220Y – SOLUTIONS

Required Problems:

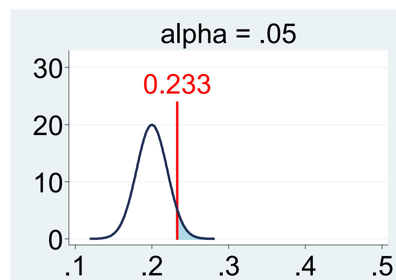
(1) FALSE. The idea that there is a trade-off between the P-value and Type II error is a common misconception based on seriously flawed logic. There is a trade-off between Type I and Type II error but that refers to the significance level (α, the predetermined threshold for the maximum Type I error we are willing to tolerate) and not the P-value (measure of the strength of our actual evidence). It is true that the P-value is the probability of making a Type I error if the null were true. However, it does NOT follow that a smaller P-value means a larger Type II error. First, the trade-off refers to how big the burden of proof is: the choice of a significance level (α). If you choose a large significance level, such as $\alpha = 0.10$, then you would raise the chance of sending an innocent person to jail (Type I error) and you would decrease the chance of letting a guilty person go free (Type II error). If you choose a smaller significance level, such as $\alpha = 0.01$ (closer to the "beyond a reasonable doubt" standard), then you would lower the chance of sending an innocent person to jail (Type I error) and you would increase the chance of letting a guilty person go free (Type II error). BUT, while you choose the significance level, you do NOT choose the P-value. Hence there is a difference between the significance level, which is the burden of proof, and the P-value, which is the proof/evidence you actually have. We would never say that because we have a huge amount of evidence of a defendant's guilt (very small P-value) that we think that there is an increased chance of letting a guilty person go free. That doesn't make sense. We could say that if we generally REQUIRE a huge amount of evidence to convict people (a very small significance level) then we will increase the chance of letting guilty people go free. Another way to think about this is that there are many factors that affect BOTH the P-value and the chance of making a Type II error: sample size, the standard deviation of the sample proportion, whether the research hypothesis is one or two directional, and the value specified in the null hypothesis. Changing any of these underlying factors will change both the P-value and the probability of making a Type II error. Hence we cannot say there is a causal relationship between the P-value and the probability of making a Type II error, which means that we cannot say that changing one would result in a change in the other.

(2) (a) Find the critical value of the hypothesis test.

$$P(Z > 1.645) = 0.05$$

$$P\left(\hat{P} > 1.645 * \sqrt{\frac{0.20(1-0.20)}{400}} + 0.20 \mid H_0 \; is \; true, n = 400\right) = 0.05$$

$$P(\hat{P} > 0.233 \mid H_0 \; is \; true, n = 400) = 0.05$$



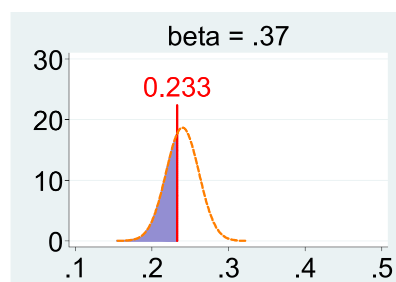
Hence the critical value (unstandardized) is $p^* = 0.233$. The rejection region is (0.233, ∞): if our test statistic, $\hat{P}$, lies in this region then we (correctly) reject the false null hypothesis.

If our test statistic, $\hat{P}$, lies outside this region then we (incorrectly) fail to reject the false null hypothesis: make a Type II error. Find the probability of a Type II error.

$$\beta = P(\hat{P} < 0.233 \mid p = 0.24, \; n = 400)$$

$$\beta = P\left(\frac{\hat{P} - 0.24}{\sqrt{\frac{0.24(1-0.24)}{400}}} < \frac{0.233 - 0.24}{\sqrt{\frac{0.24(1-0.24)}{400}}}\right)$$

$$\beta = P\left(Z < \frac{0.233 - 0.24}{\sqrt{\frac{0.24(1-0.24)}{400}}}\right)$$

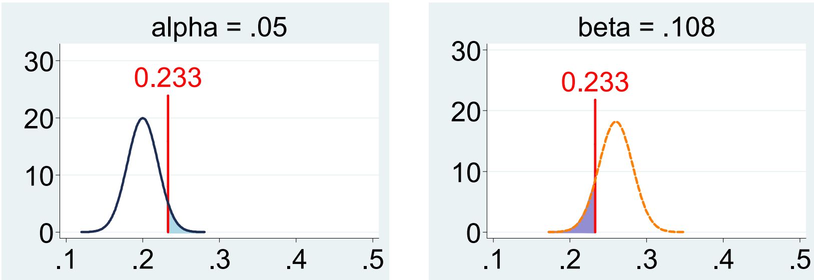


$$\beta = P(Z < -0.33) = 0.37$$

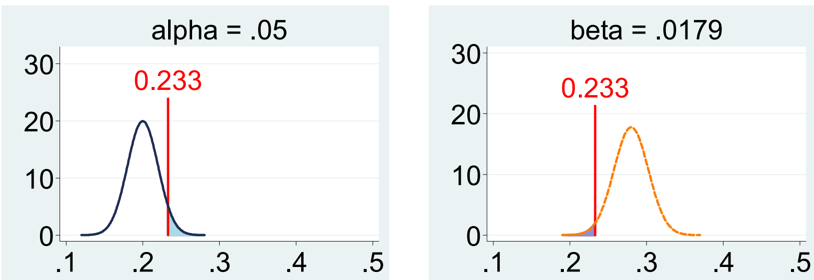
Hence the power of the statistical test is 0.63 (=1 – 0.37). This means that if we randomly sample 400 people and ask if they recall the product there is a 63% chance that the statistical test will lead us to conclude correctly that at least 20%

recall the product if in fact 24% of the population do. On the dark side, there is a 37% chance that we will fail to find sufficient evidence to support our research hypothesis even though it is in fact true.

(b)



(c)

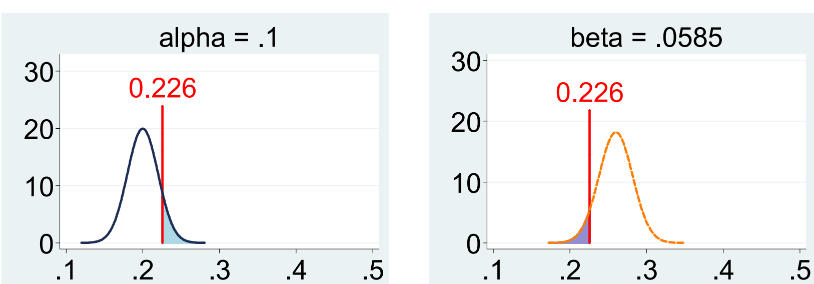


(d) The power increases from (a) to (c) because as the effectiveness of the ad improves – as a higher and higher fraction of the population recall the product – the chance that our random sample will contain a high fraction recalling the ad improves. As the sample proportion increases, our ability to reject the false null hypothesis (which says the fraction is small) improves.

(e) If half of the population recalls the ad and we sample 400 people we will almost surely obtain a high sample proportion that will fall deep into the rejection region and allow us to reject the false null hypothesis that the proportion is only 0.2. With the table you'd approximate the power as 1: the probability of a Type II error in this case is virtually zero (out past the 25th decimal place!).

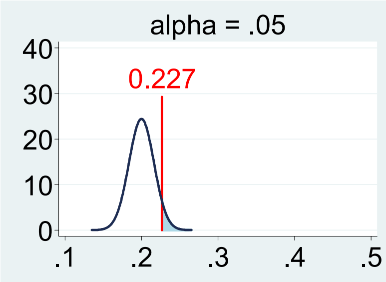
(f) If the proportion of people recalling the ad in the population is only slightly better than the null hypothesis – 20.5% versus 20% -- it is very likely that our sample statistic will be fairly small and provide insufficient proof of our research hypothesis. Power is only 0.083: there is a low chance that we will obtain a sample that will allow us to infer the research hypothesis is true (even though it is in fact true!).

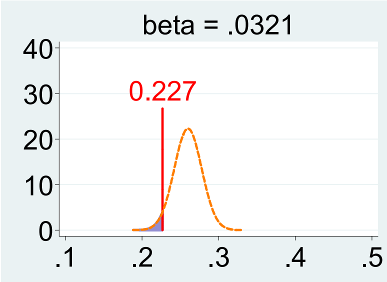
(g)



(h) See exercise 58 solutions.

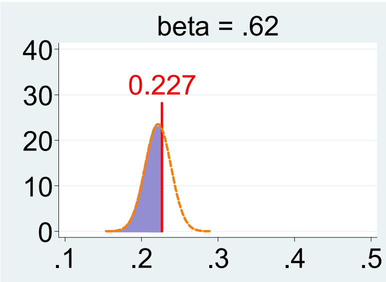
(i)





(j) See exercise 58 solutions.

(k) There *is* evidence to suggest that the radio station has in fact *exceeded* expectations. The sample proportion, $\hat{P}$, is 22.2% (0.222=133/600), which of course is greater than 20% (expectations). Unfortunately, there is not enough evidence at a 5% significance level to conclude that our sample statistic of 22.2% may not reflect sampling error: that is, getting lucky with 22.2% when in fact the population proportion is only 20%. The problem is that *if* the population proportion is only a bit higher than expectations (i.e. 22.2% versus 20%) our hypothesis test will have very low power: it is unlikely that we will be able to reject the null hypothesis even though it is wrong and our research hypothesis is correct. To quantify our argument we calculate the power of the test if *we assume* that p = 0.222 and we see the power would be only 0.38 in that case: there is only a 38% chance we'll be able to reject the false null.



(3) Choose a significance level. It was not specified and the situation presented is generic so we may go with the conventional significance level: α = 0.05. Find the rejection region for this one-sided hypothesis test.

Standardized rejection region:

$$P(Z > 1.645) = 0.05$$

Un-standardized region:

$$P(\hat{p} > 1.645 * \sqrt{\frac{0.6(1-0.6)}{100}} + 0.60 \mid H_0, n = 100) = 0.05$$

$$P(\hat{p} > 0.6806) = 0.05$$

Find probability of Type II error (β) if p = 0.58:

$$\beta = P(\hat{p} < 0.6806 \mid p = 0.58, n = 100) = P\left(Z < \frac{0.6806 - 0.58}{\sqrt{\frac{0.58(1-0.58)}{100}}} \right) = P(Z < 2.0383) = 0.9792$$

Find probability of Type II error (β) if p = 0.60:

$$\beta = P(\hat{p} < 0.6806 \mid p = 0.60, n = 100) = P\left(Z < \frac{0.6806 - 0.60}{\sqrt{\frac{0.60(1-0.60)}{100}}} \right) = P(Z < 1.645) = 0.95$$

*Note: This is a bit strange because in this case the null hypothesis is NOT false. To understand it, think about a value very close to 0.60 (such as 0.6001). In this case the probability of making a Type II error will be 1 - α.

Find probability of Type II error (β) if p = 0.62:

$$\beta = P(\hat{p} < 0.6806 \mid p = 0.62, n = 100) = P\left(Z < \frac{0.6806 - 0.62}{\sqrt{\frac{0.62(1-0.62)}{100}}} \right) = P(Z < 1.2485) = 0.8941$$

Find probability of Type II error (β) if p = 0.64:

$$\beta = P(\hat{p} < 0.6806 \mid p = 0.64, n = 100) = P\left(Z < \frac{0.6806 - 0.64}{\sqrt{\frac{0.64(1-0.64)}{100}}} \right) = P(Z < 0.8458) = 0.8012$$

Find probability of Type II error (β) if p = 0.66:

$$\beta = P(\hat{p} < 0.6806 \mid p = 0.66, n = 100) = P\left(Z < \frac{0.6806 - 0.66}{\sqrt{\frac{0.66(1-0.66)}{100}}} \right) = P(Z < 0.4349) = 0.6682$$

Find probability of Type II error (β) if p = 0.70:

$$\beta = P(\hat{p} < 0.6806 \mid p = 0.70, n = 100) = P\left(Z < \frac{0.6806 - 0.70}{\sqrt{\frac{0.70(1-0.70)}{100}}} \right) = P(Z < -0.4233) = 0.3360$$

Find probability of Type II error (β) if p = 0.72:

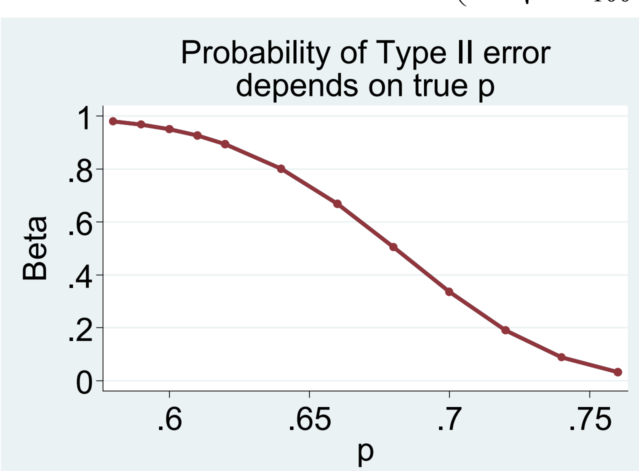
$$\beta = P(\hat{p} < 0.6806 \mid p = 0.72, n = 100) = P\left(Z < \frac{0.6806 - 0.72}{\sqrt{\frac{0.72(1-0.72)}{100}}} \right) = P(Z < -0.8775) = 0.1901$$

Find probability of Type II error (β) if p = 0.74:

$$\beta = P(\hat{p} < 0.6806 \mid p = 0.74, n = 100) = P\left(Z < \frac{0.6806 - 0.74}{\sqrt{\frac{0.74(1-0.74)}{100}}} \right) = P(Z < -1.3542) = 0.0878$$

Find probability of Type II error (β) if p = 0.76:

$$\beta = P(\hat{p} < 0.6806 \mid p = 0.76, n = 100) = P\left(Z < \frac{0.6806 - 0.76}{\sqrt{\frac{0.76(1-0.76)}{100}}} \right) = P(Z < -1.8591) = 0.0315$$



(4) eBay is clearly starting with the presumption that the objects for sale are not fake and will only infer that they are fake if overwhelming evidence is presented.

H_0: An object is authentic (not a fake)
H_1: An object is fake

eBay clearly places the burden of proof on proving that an item is fake (and not in proving that it is authentic). eBay has argued that collecting data (evidence) is expensive and is not its responsibility. Hence, there is nothing to reject its presumption. eBay is clearly very concerned about making a Type I error (rejecting a true null hypothesis), which in this case means wrongly labeling an authentic item as fake. Mr. Durzy [an eBay employee] argued that "if we began to automatically pull listings for things reported to us as fake, we could be pulling listings that are legitimate." This means that it has set the burden of proof very high: picked a significance level very close to zero. This is evidenced by the fact that it still fails to reject the null in the face of substantial evidence from trademark holders like Tiffany and upset buyers who've actually went to the trouble of having their item appraised (such as Ms. Pollack). The source of controversy is Type II errors: failing to reject a false null hypothesis. Because eBay has expressed the complete intolerance for Type I error (picked a very very small significance level), the magnitude of Type II error is huge. In this example a Type II error is failing to reject the presumption that an item is authentic when in fact it is a fake. This means that there will be many fakes being traded. This causes controversy because buyers and trademark holders are negatively affected by Type II errors and want the frequency of such errors reduced. On the other hand, the article points out that Type II errors *benefit* eBay because it collects fees no matter whether the object is authentic or fake. Hence we can guess why eBay is not very concerned about this.