by Mateusk Oskarbski for the University of Minnesota First-Year Writing Program under the Department of Writing Studies

T-Test vs. Z-Test: Statistical Significance?

Recently in Frontier Hall, there was a spread of a highly contagious norovirus between several of the students in that hall. The students are proposing that the norovirus must have started because of improper cooking in the dining hall because all that were originally infected also went to Late Night several nights ago. Perhaps that soft shell beef taco was not the best idea. Suppose a worker at the Department of Health randomly contacts 80 students in Frontier Hall and asks them two questions: "Were you at Late Night at any time in the last 72 hours and been experiencing vomiting or diarrhea in the last 72 hours?" Out of the 80 students, 60 of them went to Late Night and have the symptoms of norovirus. Now what if there was no change in the dining hall food and another sample of 80 different students answer that same question. This time only 40 people of the 80 went to Late Night and have the symptoms of norovirus. Is this statistically significant? For a University of Minnesota student completing STAT 3021, they would easily be able to conduct a 2-sample t-test for a sample proportion. However, a student could also conduct a 2-sample z-test for a sample proportion as well. Both z-procedures and t-procedures are used to calculate statistical significance depending on the size of the sample, Sample Size (n), and whether or not a population standard deviation is available (Witte 268). Since these students are deprived of the knowledge of how to perform a proper z-test, STAT 3021 class should teach the z-procedures along with their current curriculum to improve student's statistical knowledge and performance.

Now, automatically the first argument the opposition thinks of is the time needed to teach both the z-test and the t-test for the statistics course. The students are already expected to learn the "basic methods of statistics, including descriptive statistics, elementary probability ideas and random variables, distributions of sample averages, one and two sample t-procedures, simple

linear regression, [and] basic ANOVA" (ClassInfo - Spring 2016 Statistics Classes). There is already the need to spend an ample amount of time on understanding the difference between one and two sample t-procedures, so why devote even more time into learning another set of procedures? Time should not be the problem, because the z-test and the t-test share many similarities. In order to perform a t-test or z-test, students need to make sure that they are not taking more than ten percent of the population size when sampling (Ten Percent Rule), and they must prove that the distribution is Normal (Witte 214). The Ten Percent Rule states that if too many individuals within a population are sampled (>10%), then the test is no longer statistically valid (214). In order to prove a sample is Normal, there is the assumption that it is Normal if the Sample Size (n) is greater than or equal to 30 (Central Limit Theorem)(202), or the population can be Normal.

The mathematics to conduct the significance tests for both *t and z* are similar as well. To do a z-test, you have to find a z-score by subtracting the raw score by the mean and dividing that difference by the population standard deviation (Witte 213). To do a t-test, you have to find a t-score by the subtracting the raw score by the sample mean and dividing that difference by the sample standard deviation (272). Both values can be found on a chart or table of z-scores or t-scores in order to determine the P-Value, the probability of the null hypothesis being true (292-295). The only differences between a t-test and a z-test is that a z-test requires the population standard deviation, and a t-test requires the degrees of freedom (n-1 or n-2) when looking up values or doing calculations (270-271). So when students are in class learning about how to properly conduct a 2-sample t-test, students could save time without the need to find the t-score and look for a z-score if the population standard deviation is listed. If the math is almost the exact same for both tests, the same assumptions need to be made, and all the students need to

know is the only time that the z-test will not apply is when a population standard deviation is not available. Thus, there is no problem in integrating such a small fix into the STAT 3021 curriculum.

There is also the idea opponents would argue that the STAT 3021 course only teaches t-procedures for one and two samples because the students are going into fields where t-procedures would be used more often. T and z-tests show many similarities, but one of the advantages of using a t-test is that the population standard deviation does not need to be known. Instead, the sample standard deviation is used for the t-procedures (Witte 272). Many fields would benefit from using t-procedures instead of z-procedures: scientists, healthcare professionals, people in business, researchers, and many other occupations. These occupations usually will not have the population standard deviation, and they would never need to use z-procedures to check for statistical significance. However, in the field of science and discovery, repeating experiments is almost as important as doing the experiments in the first place. Scientists are rewarded for completing original work instead of repeating their own experiments because original works get publications (Resnik 101). The only time where repeating an experiment gets any publication is when the repeated experiment draws controversial data that will repeal the original hypothesis (101-102). If this is the case, then it is more likely that a scientist would put his focus on making something original rather than checking work. Now replace "scientist" with any occupation, and the previous statements sound right. Why does this matter? Well, if a person is going to repeat an experiment, there is a very good likelihood that previous researchers have estimated a population standard deviation. There is a population standard deviation, but why would a researcher want to use a z-test instead of a t-test? Ideally, one would use either a population standard deviation or a sample standard deviation to calculate

a t-score, but there are two advantages of using a population standard deviation in a z-test whenever applicable. For statistical theory, using a Normalized z-score is better than using a corrected t-score more like using a Porsche is better than using a Prius in a drag race. Both get the job done and are reliable vehicles, but one will do the job better than the other. For the student's sake, they do not have to waste time calculating Degrees of Freedom (n-1 or n-2) for a t-test and incorporating t-score to determine statistical significance. But because Degrees of freedom (n-1 or n-2) are determined by Sample Size (n), it would have an effect on the t-test, the Sample Size (n) argument.

The size of the sample is an important aspect of statistics that affects the outcome of both the z-procedures and the t-procedures. As stated previously, opponents argue that proper z-procedures need to make sure a Sample Size (n) is big enough to use the population standard deviation (n>30). Let us go back to the previous example about the norovirus. Know that a sample was created from contacting Frontier residents, t-procedures could be used to calculate statistical significance. Now, suppose that the sample standard deviation for the sample is 5 students for both samples. Now if we calculate the statistical significance using a 2-sample t-test for proportions, we get a P-value of $1.926858*10^{-57}$ for this t-test. This value is essentially 0, and there is no chance that the kids from the first sample could have gotten the infection from the dining hall food. Now suppose that the Department of Health could only call 8 residents of Frontier with 6 students ill in the first call and 4 students ill in the second call with a standard deviation of 0.5 students for both of the samples. Doing the same test, we get a P-value of $1.3694567*10^{-6}$. That might sound like a small difference, but the quotient of $10^{-6}/10^{-57}$ is 10^{51}! The change in the degrees of freedom (n-1 or n-2) change the t distribution and this makes a dramatic difference in P-values where a borderline difference could turn into a statistical

difference all by a change in the size of the sample. Standard deviation is also affected from the Sample Size (n) because correcting the standard deviation for the Sample Size (n) only makes the standard deviation smaller. Dividing the standard deviation by the Sample Size (n) for a 1-sample t-test would decrease the spread of data, but would not change the mean or the raw score making a t-score or z-score larger. A larger score results in a larger statistical difference. What about a 2-sample t-test. The correction of the standard deviation for a 2-sample t-test is a little more complicated, but the resulting standard deviation would be smaller if the Sample Size (n) is significantly larger. So why complete two separate t-tests? T-tests are as dependent on the size of the sample as much as z-procedures are. Making a big enough Sample Size (n) is costly and time consuming for both parties. So why do we need to discriminate the t-test over the z-test if both require a large Sample Size (n) in order to make good statistical conclusions?

If the t-test requires a large Sample Size like the z-test, is harder to do in comparison to the z-test, could be applicable for all professions, and could be integrated easily into the STAT 3021 curriculum, then for what reason is the University not teaching this concept? The z-test is something easier for students to understand before introducing t-tests. bell hooks once mentioned in her essay regarding the education process that "the most important learning experience that could happen in [the] classrooms was that students would learn to think critically and analytically, not just about the required books ,but about the world they live in" (hooks 102). And if the University wants to make students who are going to use their statistical knowledge to real life situations, then it is best to say that they should be teaching the students everything the student needs to know before they become expert statisticians.

[1717 Words]

WORKS REFERENCED

"ClassInfo - Spring 2016 Statistics Classes." Web. 21 Apr. 2016. http://classinfo.umn.edu/?term=1163&level=&subject=STAT&catalog_nbr=3021

hooks, bell. "Pedagogy and Political Commitment: A Comment." *Talking Back: Thinking Feminist, Thinking Black.* Boston: South End Press, 1989. 98-104

Resnik, David B. The Ethics of Science: An Introduction. London: Routledge, 1998.

Witte, Robert S., and John S. Witte. Statistics, 8th Edition. John Wiley & Sons, 2006.

SUCCES MEMO

Simple (Compact): For Core, my central message is that STAT 3021 classes should teach both the z-procedures and t-procedures to make students better statisticians. Compact is shown where I compare a z-test to a Porsche and a t-test to a Prius to demonstrate the statistical power of the z-procedures vs. t-procedures.

Unexpected: A Knowledge Gap was opened when I posed the question in the first paragraph "Is this statistically significant?" when referring to the norovirus example. Knowledge Gap was then closed in the fifth paragraph when I concluded that the dataset was statistically significant and that the school dining halls in the surrounding area are not responsible for the norovirus outbreak.

Concrete: The first paragraph shows Concrete by stating "a soft shell beef taco." Abstract idea represented is the possibility of improper cooking by the dining hall staff. The audience could visualize and taste a soft shell beef taco. The fourth paragraph shows a Porsche. A person could visualize and, for the car enthusiast, even imagine the sound of the engine. The abstract idea represented is the ideal nature of the z-test in comparison to the t-test (Prius).

Credible: Statistics with Human Scale is shown in the Introduction when setting up parameters for how many students are sick for each of the two studies conducted. Eighty college students could be imagined. If that is not small enough, the Sample Size (n) is reduced by a factor of ten down to eight college students. There is also a Testable Credentials in the fourth paragraph where I justify the why the z-procedures should be used with other occupations. So not many occupations would have a population standard deviation, but good scientists repeat experiments. If these scientists repeat experiments, then they might get enough data to get a population standard deviation. Then if there is a population standard deviation is available, then z-test should be used.

Emotional: The Conclusion focuses on how students would benefit from having a curriculum that has both t-procedures and z-procedures. The Appeal to Self-Interest (Maslow's Needs) focuses on a higher level of critical and analytical thinking satisfying the Learning desire of Maslow's Needs. Because the Learning desire is not a physical need, we also satisfy the Avoiding Maslow's Basement appeal as well.

Stories: At the beginning of the paper, the Norovirus Experimental Design shows statistic students how to properly design a statistical test and determining which type of procedures that would be used. The problem is determining whether or not the University dining halls are responsible for spreading norovirus, and the problem is resolved by setting up a significance test. The experiment is formed, the experiment is preformed, and the significance test to be used is determined.