

**ASSIGNMENT FRONT SHEET**

**Course Name: ALY6010 71904 Prob Theory and Intro Stats**

**Professor Name: Dr. Morteza Ziyadi**

**Student Name: Dong Quoc Tuong (Lukas)**

**Student Class: Fall 2019 CPS Term: A. 2019**

|  |
| --- |
| **Module 4: One-sample Confidence Intervals & Hypothesis Testing**  **Completion Date: October 15st Word Count: 1018 Due Time:12:00am** |

**Statement of Authorship**

*I confirm that this work is my own. Additionally, I confirm that no part of this coursework, except where clearly quoted and referenced, has been copied from material belonging to any other person e.g. from a book, handout, another student. I am aware that it is a breach of Northeastern University’s regulations to copy the work of another without clear acknowledgement and that attempting to do so renders me liable to disciplinary procedures. To this effect, I have uploaded my work onto Turnitin and have ensured that I have made any relevant corrections to my work prior to submission.*

**Tick here** to confirm that your paper version is identical to the version submitted through Turnitin

1. **Introduction**

Thanks to statistical software in data analysis, this report allows us to understand how we can perform an estimation of the population parameters with the confidence level. From a small sample, we will be able to answer the hypothesis questions that were proposed with . This report also includes ANOVA and Chi-square testing

1. **Analysis**

**Part 2**





Because this is a big sample size (>30) which is why we will find the z value using the norm.s.inv formula in excel with 92%,96%,98%. Then we find the margin of error by taking the Z value multiply with the sample’s standard deviation and divide it to square root of the sample size (40). The interval will be between the Mean plus or minus the margin of error. For example, for a sample of 40 variables, we can predict that with 92% confidence level that the mean of the population will be around 0.79- 1.06. The population (0.92) means lies inside of the interval so our prediction seems to be correct.





Then to find the minimum sample size that we as a given confidence level, we will multiply the Z value with sample’s standard deviation, divide by that confidence level’s margin of error. Then we square that number and we will have the required sample size. For example, we will need a sample size of 286 people to carry out a reliable prediction of the population’s mean with confidence level at 92%.

**Part 3**





Because this is a small sample size (<30) which is why we will find the t value using the t.inv formula in excel with 92%,96%,98% and the degree of freedom (23). Then we find the margin of error by taking the t value multiply with the sample’s standard deviation and divide it to square root of the sample size (30). The interval will be calculated the same way like above. However, the results are different as we see that only with 98% confidence level, the population’s mean lies inside the interval

**Part 4**





The success rate is calculated by countif function with the condition that the variables are less than 1 and then divide them to the sample size. The failure proportion equals to 1 minus the success proportion. Sampling error is multiply the two proportions together and divides that number to sample size. Use Norm.S.inv to find the z value. Margin of error equals to z value multiply the square root of sample error. Find the interval like above. Minimum sample size needed would be multiplication of the two proportions and the square of z value that divide by 0.05, round up the nearest number. We find the population’s success rate lies inside the interval for all three confidence levels.





**Part 5**







Use formula chisq.inv with the probability equals to 1 minus/plus the confidence level and then divide by 2 with the degree of freedom to find the variance left and right. Then find the Variance interval and then the Standard deviation interval. The population‘s standard deviation is inside of the all interval. Our prediction is highly correct

**Part 6**





Since we need to find out if the mean is different from one, we will put the null hypothesis as μ=1 and the alternative hypothesis would be μ not equal to 1. This is the two tailed problem with big sample so we will start finding Test statistic z and use 2\*norm.s.dist to find P value because the z value is less than 0. Find two critical values. The z value lies between the two critical level, so we cannot reject the null hypothesis because we do not have enough evidence to say that the population mean is different from 1.





**Part 7**





Since we need to find out if the mean is more than one, we will put the null hypothesis as μ<=1 and the alternative hypothesis would be μ > 1. This is the right tailed problem with small sample so we will start finding Test statistic t and use 1-t.dist to find P value. Find critical value. The t value is less than the critical value, so we cannot reject the null hypothesis because we do not have enough evidence to say that the population mean is more than 1.





**Part 8**





We will use Chi-Square test with this one. (Ugoni & Walker, 1995). Since we need to find out if the proportion for success is less than 65% or not, we will put the null hypothesis as p>=65% and the alternative hypothesis would be p < 65%. This is the left tailed problem. We start finding the z value by taking the test success proportion minus the 65% then divide it to the square root of the multiplication of the success proportion that divide the sample size. Use Norm.s.dist to find the P value and norm.s.inv to find the critical value. We can reject the null hypothesis because z value is less than the critical value. So we have enough evidences to say that p value is less than 65%,





**Part 9**





We will use Anova test with this one (Lane, 2013). Since we need to find out if the variance is different from 1, we will put the null hypothesis as σ2=1 and then the alternative σ2 not equal to 1. This is a two tailed problem, so we will find the z value by multiply the sample’s total minus 1 with sample ‘s variance and then divide by 1. P value are 2\* ( 1 – chisq.dist and 2\* chisq.dist and Critical Values are calculated through the chisq.inv formula.The z lies between two critical values, it means that we cannot reject the null hypothesis because there is we do not have enough evidences to know if the variance is different from 1





1. **Conclusion**

The exercise perfectly encapsulated the importance of choosing the correct testing methods to yield the desired result. Bearing in my that this is just prediction so if anyone decides to take a different sample group, the result for the analysis will be completely different and we will be able to reject more null hypothesizes.

**References**

Lane, D. M. (2013). Introduction to Statistics (Chapter 15), 517–598. Retrieved from http://onlinestatbook.com/mobile/analysis\_of\_variance/anova.pdf

Ugoni, A., & Walker, B. F. (1995). The Chi square test: an introduction. *COMSIG Review*, *4*(3), 61–64. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17989754%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2050386