

**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 5: Two-sample Confidence Intervals & Hypothesis Testing**  **Completion Date: October 22st Word Count:1171 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**

The report illustrates how researchers can perform estimation of a population base on the confidence intervals and the two samples sizes. Then from that on, we can test the hypotheses to see whether we can accept or reject them. This report also includes ANOVA and other testing methods. The occupational data regarding New York City and Los Angeles serves as the data foundation for our analysis.

1. **Analysis**

**Part 2**





We first compute the means of population one and two, find the population means difference. Due to the size of the sample is bigger than 30, we find the z value with the norm.s.inv method with the confidence CI at 90%,95%,98%. Henceforth, we can know the margin of error by multiplying the unchanged sampling error to the Z values that we just discovered earlier. The lower/upper limit for the means difference will be equaled to the sum or difference between the sample means difference and the margin of error. For instance, with a confidence level at 90%, I can predict that the means difference of these two population will fall between -0.58 and -0.09. From a respective point of view, our estimation for the means difference of two populations are correct regardless of the confidence level.



**Part 3**





.Due to the sizes of the samples are smaller than 30 (22 and 20), we find the t value with the t.inv method with the confidence CI at 90%, 95%,98%. We also calculate the Sampling Error and Degrees of Freedom as well through complicated formula like in the picture above. After that,. we can know the margin of error by multiplying the unchanged sampling error to the t values that we just discovered earlier. The lower/upper limit for the means difference will be equaled to the sum or difference between the samples means difference and the margin of error. For instance, with a confidence level at 90%, I can predict that the means difference of these two population will fall between -0.83 and 0.28. From a respective point of view, our estimation for the means difference of two populations are correct regardless of the confidence level.



**Part 4**





Through the countif function, we can calculate the number of success cases from the sample with the condition that they have to be more than 2. Then we find the proportion of success and the portion of failure. We also calculate the Sampling Standard error like this here. We will find the z value through the norm.s.inv. Then compute the margin of error and interval limit like above. For instance, with a confidence level at 90%, I can predict that the difference of population success proportion is will fall between -0.0463 and 0.0340. From a respective point of view, our estimation for the difference of population success proportion is correct regardless of the confidence level.



**Part 5**





Since we need to prove that the standard deviation of the two population is different. We will put the null hypothesis σ12 /σ22  equals to 1. It is a two-tailed problem and the sample size are small to do the Z test, so we will do the t test instead. (Chieh, 2018) Test statistic in this case is taking the two sample variance divided to each other, bearing in mind that the first one has to have larger sample variance then the second. P value is calculated by 2\*(1-F.DIST formula. There is only one significant value only. We can reject the Null hypothesis because the test statistic is greater than the critical value, meaning that we have enough evidence to do so. After that we use Data Analysis function to yield the same result.



**Part 6**





Since we need to prove that the difference means of the two populations is less than 0. We will put the null hypothesis µ1 - µ2 >0. It is a left-tailed problem and the sample size are big enough for us to do the Z test. Test statistic in this case is taking the sample means difference dividing the sampling error. P value is calculated by Norm.S.DIST formula. There is only one significant value only. We cannot reject the Null hypothesis because the test statistic is less than the critical greater than the critical value, meaning that we do not have enough evidence to do so. After that we use Data Analysis function to yield the same result.



**Part 7**

Since we need to prove that the difference means of the two populations is less than 0. We will put the null hypothesis µ1 - µ2 >0. It is a left-tailed problem but we have the unequal variance. So the formula to calculate Sampling Standard error and degree of freedom are totally different from the Equal variance. We find the test statistic like above but use the T methods instead. As we saw that the Test Statistics are different but the Critical values are somewhat similar between these the Equal and unequal variances. Nevertheless, at the end, we still cannot reject the Null hypothesis because the Test Statistic is greater than the critical values, meaning that we do not have enough evidence to reject H0 .After that we use Data Analysis function, it yield somewhat the same result.

* **Unequal Variance**





* **Equal Variance**





**Part 8**





Since we need to prove that the difference means of the two dependent populations is different from 0. We will put the null hypothesis µ1 - µ2 =0. The degree of freedom equals to the sample size minus 1. It is a two tailed problem and we use the 2\*T.DIST formula to find the P value after the test statistic. There are two critical values 2.0930 and -2.0930. The Test statistic lies between the two critical values, that means we cannot reject the null hypothesis because we do not evidence s to suggest that the means of the population are not equal to each other. After that we use Data Analysis function to yield the same result.

**Part 9**





Since we need to prove that p1 >p2 that means that the null hypothesis will be p1 <= p2. This is a

This is the right tail problem and we use the 2\*T.DIST formula to find the P value after the test statistic. Because the Test statistic is less than the critical value, we cannot reject the null hypothesis because we do not have enough evidences to do so. The result is different from the Q4’s result where the p1 (7.83%) is clearly greater than p2 (5.76%).

1. **Conclusion**

This report perfectly demonstrates how important it is when it comes to selecting the right testing methods for the desired outcome. Additionally, the fact that the formulas vary from one to another means we need to pay extra attention when it comes to these exercise like this. If not, the predictions would be inaccurate. (Zax, 2013) However, Question 9 and 4 have shown us that these are just estimations and might be incorrect. So it is much better that we take different samples to test again and again.

**References**

Chieh, C. J. (2018). Making sense of the two-sample t-test. Retrieved from https://www.isixsigma.com/tools-templates/hypothesis-testing/making-sense-two-sample-t-test/

Zax, J. (2013). Confidence intervals and hypothesis tests. *Introductory Econometrics*, *1*, 191–231. https://doi.org/10.11126/stanford/9780804772624.003.0006