Mustafa Abdullayev

USA Migration Project

AY6010 – Probability Theory and Introductory Statistics

Term: Fall 2019

Instructor: Tom Breur

20/10/2019

Introduction

This is Microsoft Word Report accompanying Microsoft Excel Workbook. Data to analyze was about migration rate in different counties of USA. Data was provided by class instructor, Tom Breur. It was cleaned and ready to analyze. Main aim of this workbook is to interpret meaningful patterns and insights in the data relevant to the given business question. At first part, I performed estimation of population parameters, I used confidence intervals based on both one sample and two sample. Afterward, I carried out various hypothesis tests for population parameters, difference between two population parameters and ANOVA. Also, I utilized powerful Excel built-in functions and graphs to dive deeper to observe hidden patterns and visually communicate my findings to audience.

Analysis

At first, I created a confidence interval for the mean of difference between domestic migration and international migration. Upon checking population, I observed that, despite having close means their variance was extremely different. Since I know population variance and my sample sizes were bigger than 30 (42 and 40,in respectively) I used Z-test and constructed confidence intervals. Crucial observation here is that, or 0.05 significance level , not both of upper and lower limits were smaller than 0. This fact suggest that we are 95% confident that difference between means ( domestic – international ) can be both of positive and negative values, therefore domestic migration mean is not smaller than that of international. We further validated this in our hypothesis testing. We tested if we can reject our hypothesis that mean difference is not smaller than 0, but I could not reject it. So, both findings support each other. So, status quo is to assume that mean of domestic migration is not smaller than that of international population.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Confidence Level CI** | **zc** | **Margin of Error** | **CI Lower Limit for the Means Difference** | **CI Upper Limit for the Means Difference** |
| 90% | 1.645 | 2.823 | **-2.345** | **3.301** |
| 95% | 1.960 | 3.364 | **-2.886** | **3.842** |
| 98% | 2.326 | 3.992 | **-3.515** | **4.470** |

Moreover, I created CI with t-test values by using smaller samples. I assumed that my confidence interval here would be wider since tails of t distribution is flatter. Indeed, that was the case. And, CI contains 0, which suggest that there can be case as means are equal. So, it is reasonable to assume that we could not reject the hypothesis staying that means are equal. To test that, I same samples and assumed no prior knowledge of population variance. In that case I had to calculate sample variances in lieu of population parameters. T-test contain something called degree of freedom which is the number of values in the final calculation of a [statistic](https://en.wikipedia.org/wiki/Statistic) that are free to vary. Since, width of tails in the t-test are bigger I assumed that I should get the same results as z-test, which is stricter. I did this test in two parts. First, I assumed population variances are unequal and secondly, I assumed that they are equal. In both cases I was unable to reject our null hypothesis staying that means are equal. So, both CI findings and hypothesis test supported each other.

Other than means, I constructed confidence intervals for population proportion. Having defined success as rate, which is bigger than 7, I constructed CI for proportion differences. The striking fact was that, for all confidence intervals I used, both lower and upper limits were bigger than zero. This show that difference in proportions does not contain zero and always bigger than 0. It shows domestic success is bigger than international success. To test that hypothesis , I did hypothesis testing in Q9. My null hypothesis was that success in domestic population is bigger than that of international. I could not reject that hypothesis because our p-value is bigger than significance level (0.05). So, default assumptions is to say that success rate in domestic migration is bigger than that of international population.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Confidence Level CI** | **zc** | **Margin of Error** | **CI Lower Limit** | **CI Upper Limit** |
| 92% | **1.751** | **0.062** | **0.180** | **0.303** |
| 95% | **1.960** | **0.069** | **0.173** | **0.311** |
| 99% | **2.576** | **0.091** | **0.151** | **0.332** |

Finally, I did analysis of variances. I tried to reject hypothesis stating that 2 population variances are same. In order to achieve that I need F test. In here, my p-value was extremely small, almost zero. P-value tells us that, what is the probability that our result is due to luck or randomness. So, since my p-value was extremely small (smaller than significance level) I can say that my finding is not due to any king of randomness. For example, that means, it is not due to selected samples but ,indeed, due to inherent difference between 2 population variances. In turn this mean our finding is statistically significant so I can reject the null hypothesis. After rejecting my null hypothesis, I can assume that variances of 2 population is not same.

Conclusion

To conclude, throughout my excel workbook I constructed confidence intervals and performed hypothesis testing. CI and hypothesis tests were used to compare two populations namely domestic migration rate and international migration rate in different USA counties. The most striking fact was the difference between population variances. Our p-value was almost 0, signaling that there is significant difference between two variances that can not be attributed to randomness or luck. Indeed, variances were 119.92 and 3.60 in respectively. This shows us that rate of international migration is somewhat systematic. But domestic migration rates were highly volatile and hard to predict.

Reference

Wikipedia contributors. (2019, October 13). Degrees of freedom (statistics). In *Wikipedia, The Free Encyclopedia*. Retrieved 13:53, October 20, 2019, from [Degree of Freedom](https://en.wikipedia.org/w/index.php?title=Degrees_of_freedom_(statistics)&oldid=921039554)