Mustafa Abdullayev

USA Cars Assignment 2

AY6010 – Probability Theory and Introductory Statistics

Term: Fall 2019

Instructor: Tom Breur

10/13/2019

Introduction

This is Microsoft Word Report accompanying Microsoft Excel Workbook. In my workbook, I analyzed USA cars dataset. Data was provided by class instructor, Tom Breur, and derived from USA government database. It consisted of information about percentage of household without vehicle and average number of vehicles in 317 USA cities for years 2015 and 2016. Data was cleaned, organized and ready for analysis. With this dataset, I constructed various confidence intervals for population mean, proportion and variance. Moreover, I did hypotheses testing for various assumptions and significance levels. Also, I utilized powerful Excel built-in functions and graphs to dive deeper to observe hidden patterns and visually communicate my findings to audience.

Part 1 – Confidence Intervals

For Confidence Interval constructing I used Z values since had the information regarding population variance. But for second sheet of my workbook, I assumed I dont know population variance and since sample size was small (<30) I utilized t values. After calculating appropriate z and t values for first two cases, I used them to find margin of errors. Then combining with our sample means I constructed my confidence intervals for both cases. Theoretically t distribution has fatter tails. Indeed, when observing my confidence interval for both of t and z distribution I saw that it is wider in t distribution. For same confidence interval CI which created with t distribution values yield wider interval. Another interesting fact was that, in order to find population CI with 98% confidence interval we need less than 200 people. This is not a huge number to collect and confidence is high. On the other hand, to find proportion with 99% confidence we need almost 260 people. These are plausible number in order to have little cost and high confidence.

|  |  |  |
| --- | --- | --- |
| **Confidence Level CI** | **Desired Margin of Error** | **Minimum Sample Size Needed** |
| **92%** | **0.05** | **110** |
| **96%** | **0.05** | **151** |
| **98%** | **0.05** | **193** |
| **Confidence Level CI** | **Desired Margin of Error** | **Minimum Sample Size Needed** |
| 90% | 0.08 | **106** |
| 95% | 0.08 | **151** |
| 99% | 0.08 | **260** |

Part 2 – Hypothesis testing

In this part we tested various hypotheses. I usually used Z test because we knew the population variance. In order to test our hypotheses, I calculated the test statistics and our critical values. I need our significance value and also population size and population variance to calculate these numbers. When it comes to testing, I saw that number of cars owned by household is between 1.5 and 2. This is because I was unable to reject my hypothesis stating that :

1. M < 1.5
2. M > 2

,in respectively. Indeed, when checking the original data we see that actual number is 1.72. This shows that our calculations were correct

Conclusion

To conclude, I used Excel and its libraries to analyze a USA car ownership data. Data had information about 2015 and 2016 car ownership statistics for USA states and cities. Key finding was that, when we find confidence interval with same confidence, it will be wider for t distributions. This was theoretically true sine t distribution has fatter tails and here we observed it practically.