STA304 Final Report

Introduction

Our topic is Means of Transportation to UTM. The sampling frame includes all students enrolled in STA304 Fall 2019 Semester at UTM. We wanted to use a cluster random sampling method in our population of the STA304 students, however, we decided that simple random sampling would be better for our calculations. We used an online questionnaire to gather data about distances travelled to UTM and the transportation methods used by students. We aimed to have a minimum of 30 responses in our questionnaire so we would be able to apply the *central limit theorem* if $n \geq 30$, which would allow us to approximate our calculations.

Our objective throughout the duration of this survey was to evaluate the potential relationships between distances from campus and students methods of transportation. We deemed this research to be of importance as we wanted to draw conclusions from our analysis and propose changes to the university to better accommodate students. An example of an accommodation that can be made is more carpool permits and spaces, if it is shown that there is a substantial amount of carpoolers. The null hypothesis in our study was that the method a student uses to travel to school is not dependent on the distance they live from campus and the alternative was the distance a student lives from UTM affects their method of transportation. We anticipated the closer a student lives the more likely they are to use public transportation, consequently we aimed to reject the null hypothesis in this study.

Shortcomings & Analysis

We decided to make calculations based on *conditional probability* (the process of determining if a certain factor will occur given something has already taken place) since we believed it would be best suited for what our research consisted of. We failed to incorporate *confidence intervals* (defined range of values within a specified probability that the value parameter lies within it) and *p-values* (probability of observing the results obtained from the test) since mostly, if not all of our data was not continuous.

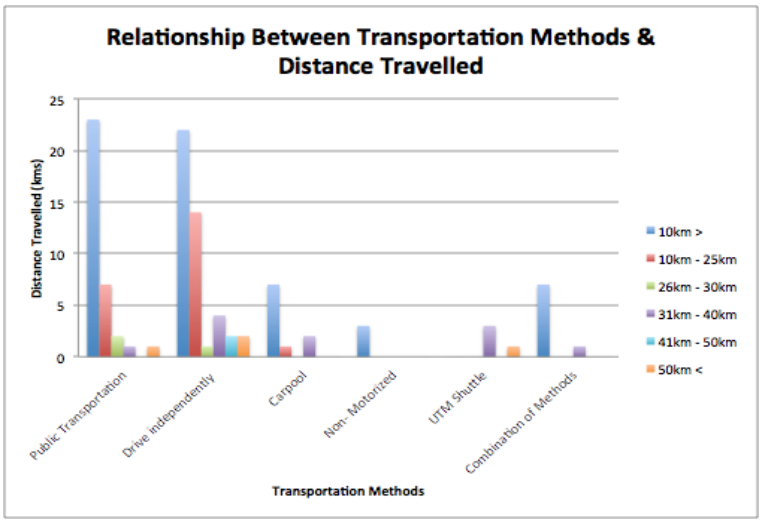
In future studies, we would improve upon our statistical analysis by taking our *contingency table* (our table showing the counts of the tests, with each row representing a range of distance travelled to get to UTM, and each column representing a method of transportation) and implementing a *chi-squared test* on it (a test that would allow us to see whether there is a relationship between distance travelled and transportation method, or whether any association with them is due to chance). Using this method we would be able to derive p-values and confidence intervals. One more way of improving upon our sampling is offering an incentive to respondents to encourage more people to respond (ex// A chance to win a $10 gift card). The increase in respondents that this would result in would further improve the accuracy of our report (by reducing variation).

Something we did do, however, is compare our study to that of an organization with larger resources and greater reach. We verified that our response rate was very high by

comparing it to the response rate achieved by UofT in a separate study published in July 2018. We were able to achieve a high response rate of 42.91% vs. a response rate of 31%.

We are unable to use *linear regression* to determine the extent to which there is a linear relationship between the dependent and the independent variable (distance from utm and method of transportation, respectively). Although distance itself is continuous, the division of distance into intervals makes it categorically discrete. Furthermore, methods of transportation is a discrete variable.

An overall limitation with this report is that we are generalizing our findings to the entire UTM student population, however we only sampled from a small population of students relative to the population. There are 14,544 undergraduate students at UTM. We conducted Simple Random Sampling on only 247 undergraduate students at UTM. Furthermore, the students that were sampled had the commonality that they were all taking the same course, with the lecture times for the course all taking place in the afternoon. Thus, it can be argued that sampling from this small subset of the undergraduate population is not truly representative of UTM students. In fact, our sampling of only STA304 Students can be described as a Sample of Convenience (which is often biased), in which we only sampled students that were easiest to select and survey (holds true even if there were legal barriers preventing us from sampling a larger subset of students).



It is easy to see that a larger portion of students who live farther away drive independently from the above chart. Likewise, a larger portion of students who live closer to UTM use public transportation. In the Appendix below, public transportation and driving independently are the two most common transportation methods chosen by the sampled STA304 students, and account for 76% of the transportation methods.

Conclusion

In conclusion, many statistical methods were used throughout our research to decide on failing to reject the null hypothesis or rejecting it entirely. We were able to determine more probabilities using a specific theorem, namely *Baye's Theorem*. This theorem allowed us to examine how students transportation behaviour changes as distances from campus increase. The use of linear regression models would have been a great help to the study as its purpose is to

outline the relationship between two variables. We aimed to draw conclusions from our study to pose suggestions to UTM for better commute methods for students. We did find the most commonly used methods of transportation to be driving independently or taking public transportation. Despite our shortcomings, we found success in our study by arriving to the conclusion that the further a student travels, the greater likelihood he/she will drive independently.

Appendix

Distance (in km	Public Transportation	Drive Ind.	Carpool	Non-Motorized	UTM Shuttle	Combination	Totals
10>	23	22	7	3	0	7	62
10-25	7	14	1	0	0	0	22
26-30	2	1	0	0	0	0	3
31-40	1	4	2	0	3	1	11
41-50	0	2	0	0	0	0	2
50<	1	2	0	0	1	0	4
TOTALS	34	45	10	3	4	8	104

Distance (in km	Public Transportation	Drive Ind.	Carpool	Non-Motorized	UTM Shuttle	Combination	Totals
10>	37%	35%	11%	5%	0%	11%	100%
10-25	32%	64%	5%	0%	0%	0%	100%
26-30	67%	33%	0%	0%	0%	0%	100%
31-40	9%	36%	18%	0%	27%	9%	100%
41-50	0%	100%	0%	0%	0%	0%	100%
50<	25%	50%	0%	0%	25%	0%	100%
TOTALS	33%	43%	10%	3%	4%	8%	100%