Analysing the Correlation Between U-Pass and T-Card Repurchases

Overview

In an era of ever-rising tuition fees, costs of university textbooks, and the general costs of living, the average student enrolled in a Canadian undergraduate program is faced with a financial struggle. At the University of Toronto Mississauga campus, this struggle is amplified by the replacement costs that some students have to bear when having to repurchase a T-Card (UofT identity card) or the Mississauga Transit U-Pass (a transit card that allows unlimited trips on the Mississauga transit system during the designated period of study). The cost to replace the U-Pass is significantly higher than the cost to replace a T-Card; with the cost of replacing a U-Pass being either $60 for the summer term, or $100 for the fall/winter term, while T-Card replacements have a fixed price of 12 dollars per card. This project will attempt to assess the correlation between the number of T-Cards and U-Passes re-purchased and estimate possible future outcomes.

Goals

The objective of our project is to analyze the proportion of U-Passes to T-cards repurchased by undergraduate students to determine the collective amount of finances inefficiently spent on repurchasing the transit passes along with the student identification cards

Specifications

- ➢ Repurchases will be defined as a time when a T-card or a U-Pass was repurchased by a student after it has been misplaced or stolen.
- ➢ The total amount of money spent on U-Pass distributions will be the sum of the amount spent on U-Pass repurchases during the summer study term and the fall/winter study term. The reason for the separation by the term is due to the difference in price.
- ➢ The sampling frame is not applicable for this project because obtaining a sampling frame for all students attending full-time or part-time studies at the UTM is not possible due to the confidentiality restrictions placed by the Office of the Registrar. Additionally, since our method of data collection involves a large degree of randomness, a frame for the students who will be present in the Davis building at our designated dates and times does not exist.
- ➢ The target population is all undergraduate UTM students currently enrolled in full-time or part-time studies.

➤ The sampling method employed is systematic random sampling. The extra statistical methods we will use to aid in data analysis will be Bootstrap estimation and multiple linear regression.

Sampling Procedure

In order to obtain our random sample, we will collect responses from students in the Davis Building during five consecutive days at random times from 10am to 8pm. The reason we have chosen the Davis building is that it provides a clear representative sample of the UTM population for a number of reasons. Firstly, the Davis building is in prime position to attract foot traffic from both the UTM Bus Stop and the largest parking garage at UTM. This will allow for a more representative sample because we would account for undergraduates who drive to school and those who travel by bus. The Davis building is also home to a number of different facilities and student services such as recreational facilities, the Bookstore, Disability Center, International Education Center and the largest food court on campus. Due to the building's wide variety of services, it proves itself to be the optimum location for sampling a representative population of UTM, allowing our sample to be diversified by the dichotomy of students who use these facilities daily.

Due to the stream of representative sampling units that will be available in the building, we will choose to use the Systematic Random Sampling (STRS) as our sampling method. Using an interval allow us to avoid obtaining sampling units that share many common characteristics, thus allowing for a random and representative sample. Considering our sampling units will be filling out the survey on their own, we have to account for the fields they leave blank. The bootstrapping method allows us to estimate values that weren't entered by our sampling units during the data collection phase.   Bootstrapping will allow for N/A fields to have numerical data, which will enable us to perform crucial statistical calculations to draw our inferences. In order to comply with student confidentiality policies, our survey will not require participants to disclose any personal information and will be conducted through the means of a web based application which will be run on the tablet provided by our group.

Survey Procedure

For systematic random sampling, we have to identify our interval between each sampling unit to be interviewed. This is calculated by dividing our total population (~14000) by our selected sample size (400). In our case, the interval is 35. The starting subject's number will have to be between 0 and 35, and after a random selection through the use of the statistical programming environment R, we arrived at 12. The systematic random sampling method
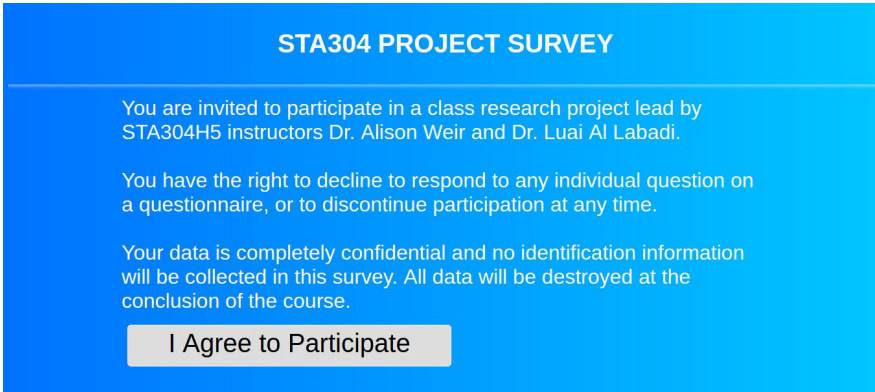
only works if the sampling units can be sorted into a sequence. Therefore, our interviews will take place close to the entrances to the Davis Building. Since our team consists of multiple members, we will divide ourselves into groups of 2 and cover the many entrances to the Davis Building in order to get a truly random representative sample. During the data collection phase, we will approach our sampling units with tablets running a web application developed by our team specifically for this project. After obtaining consent and having our units fill out the survey, the information will be saved automatically into a csv (Comma Separated Value) file. We will adjust for non-responses by simply moving on to the next sampling unit without waiting for the systematic random sampling interval.

Sophisticated Statistical Method

Multiple Linear Regression: attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data.

The multiple linear regression method will be used in order to estimate the number of T-Cards and U-Passes misplaced or stolen in forthcoming academic years at UTM. Independent variables to be used in the regression analysis include, but are not limited to, the student population at UTM in the past 4 years, the participant's year of study, program of study, means of transportation during the academic period, and his/her registration status at UTM (full-time vs part-time).

Appendix B

STA304 PROJECT SURVEY

You are invited to participate in a class research project lead by STA304H5 instructors Dr. Alison Weir and Dr. Luai Al Labadi.

You have the right to decline to respond to any individual question on a questionnaire, or to discontinue participation at any time.

Your data is completely confidential and no identification information will be collected in this survey. All data will be destroyed at the conclusion of the course.

I Agree to Participate

Appendix C

Note: Attached is a screenshot of the web-application at the time of the proposal's submission. The application's look and functionality is subject to change, but no changes that would involve collecting confidential data will be involved.

STA304 PROJECT SURVEY

Indicate Your Year of Undergraduate Study!

1	2	3	4	4+
●	●	●	●	●

Select the Number of Re-Purchased T-Cards and their Respective Years of Study

	0	1	2	3	3+
Year 1	●	●	●	●	●
Year 2	●	●	●	●	●
Year 3	●	●	●	●	●
Year 4	●	●	●	●	●
Year 4+	●	●	●	●	●

Select the Number of Re-Purchased UPASSes and their Respective Years of Study

	0	1	2	3	3+
Year 1	●	●	●	●	●
Year 2	●	●	●	●	●
Year 3	●	●	●	●	●
Year 4	●	●	●	●	●
Year 4+	●	●	●	●	●

Number of U-Passes Re-Purchased in the Fall/Winter

`e.g 0-9`

Number of U-Passes Re-Purchased in the Summer

`e.g 0-9`

[Submit Information]