# ENSC 280
# PHASE III
# GROUP 04 PROJECT
# SUMMER 2018

---

## A Statistical Study on the Relationship between Residential Distance from SFU Burnaby and Methods of Travel

---

| | |
|---|---|
| **Sirpreet Kaur Dhillon** | **301318486** |
| **Ece Eskikurt** | **301269853** |
| **Tatendaishe Jakaza** | **301278282** |
| **Reem Mustafa** | **301245064** |

# Table of Contents



# List of Tables



# List of Figures

# ABSTRACT

The objective of this project is to collect data on the mode of transport that students use to get to Simon Fraser University, and how it is related to the distance they live from the university. In short, we are conducting a commuter survey to understand how people commute to campus, and if their distance from campus affects their preferred means of transportation. The survey collected 62 responses from SFU students, the results below attempt to further understand commuting behavior.

Understanding how students commute to school can help policy makers understand how students get to school, future transportation trends, and the environmental footprint created from commuter travel. This data can be used to analyse whether better public transportation routes need to be devised and provide information for whether a student ride sharing service would be feasible and beneficial for students who live far from campus.

# INTRODUCTION

This report discusses the trend between mode of transportation to Simon Fraser University's Burnaby campus and the distance of a student's lodging from the campus. We also discuss reasons other than distance from SFU that may lead to a trend.

We use Google Forms to conduct a survey among our peers of their preferred mode of transport to get to campus and their distance from SFU. We choose these two parameters as we predicted that a person who lives further away from SFU (say 15-20 km) and has access to a car would prefer to drive where as a student who lives closer would use public transit.

SFU's student U-Pass plan makes it easier and more affordable for students to use public transit and hence we made the above predictions.

In this report, we come up with a statistical hypothesis for our predictions based on the data that we collected and the trends that we observe. A statistical hypothesis, sometimes called confirmatory data analysis, is a hypothesis that is testable on the basis of observing a process that is modeled via a set of random variables.

Then, we use data analysis methods to employ the seven-step method for accepting or rejecting our initial hypothesis. And this report presents our results and our analysis for those results.

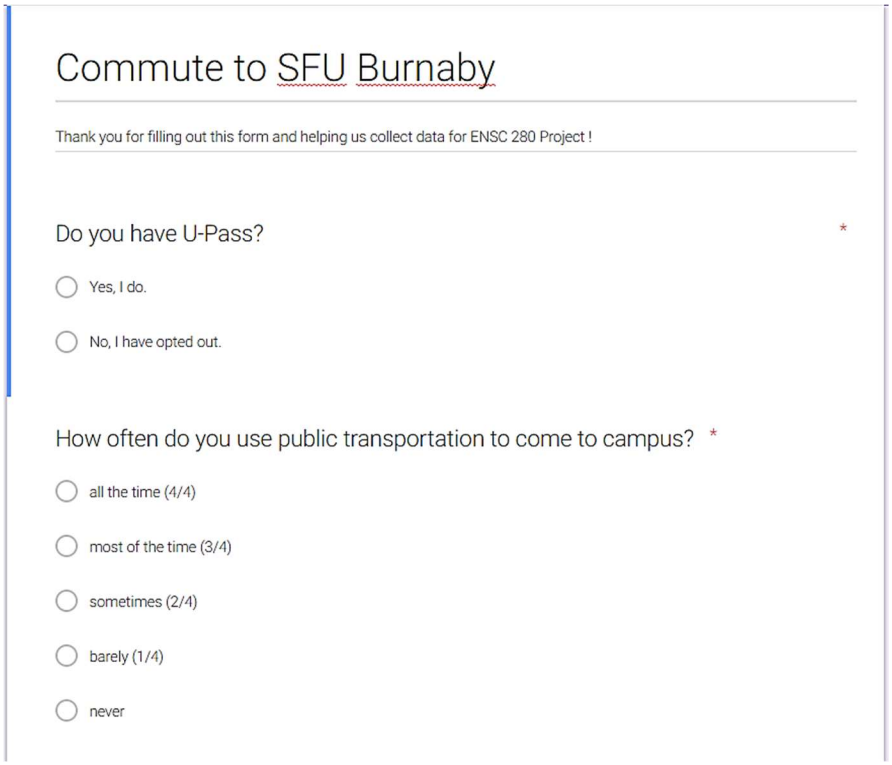
# DATA COLLECTION

For the data collection, we created a survey displayed in **Figure 1.** below through Google Forms. Each of us posted the survey on a social media platform that we are active on and ensured that at least 24 samples were collected. We used a public social media platform to randomize the data we would get rather than sending the survey to one group of people e.g. people who live near campus that we know.

We also made use of Google Maps to collect information on how far the compass card users/ drivers live from school.

| | |
|---|---|
| *Dependent Variable* | • Mode of transportation used to get to school |
| *Independent Variables* | • Distance from school (here SFU Burnaby campus)<br>• Availability of resources such as:<br>  ○ U-pass<br>  ○ Car |

***Table 1. Experiment Variables***



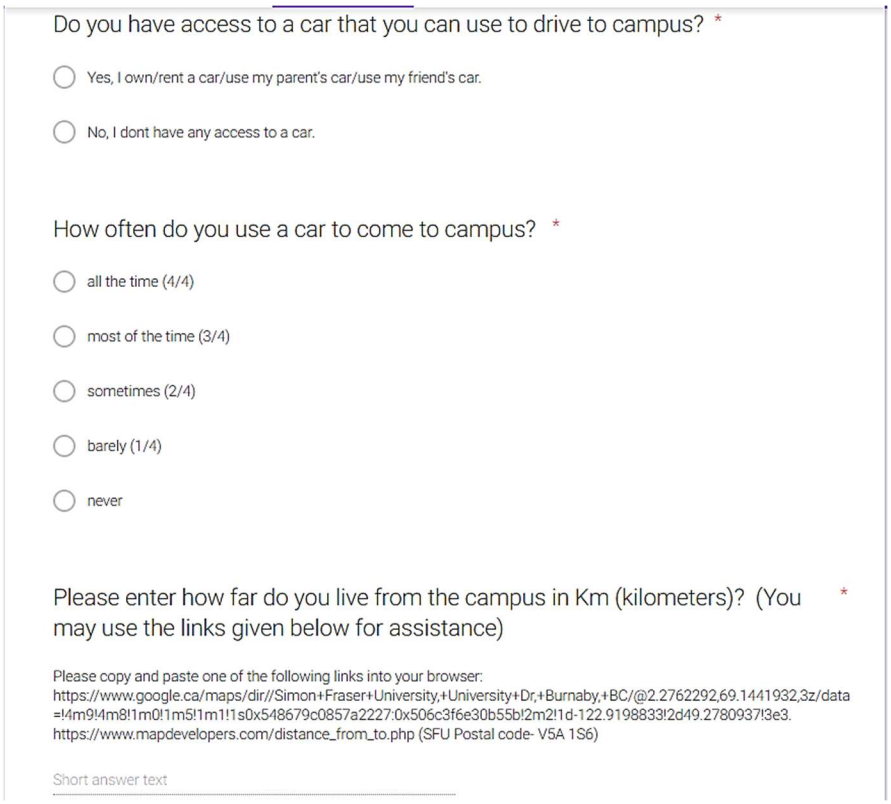



***Figure 1. Commute to SFU Burnaby Survey***

## *Raw Data*

** 0 = Never, 0.25 = barely, 0.5 = sometimes, 0.75 = most of the time, 1 = all the time.*

| # | Status of U-Pass possession, 1=yes, 0 = no | Student's tendency to use public transportation * | Status of having access to a car (0 = negative, 1 = positive) | Student's tendency to use a car to come to campus * | Distance from student's home to SFU Burnaby (km) |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 0 | 7.9 |
| 2 | 1 | 1 | 0 | 0 | 5 |
| 3 | 1 | 0.75 | 1 | 0.25 | 5 |
| 4 | 1 | 0.25 | 1 | 0.75 | 15 |
| 5 | 1 | 0.75 | 0 | 0.5 | 22.3 |
| 6 | 1 | 0.75 | 1 | 0.25 | 22 |
| 7 | 1 | 0.75 | 1 | 0.5 | 17.2 |
| 8 | 1 | 0.5 | 1 | 0.5 | 29 |
| 9 | 1 | 0.25 | 1 | 0.75 | 9.6 |
| 10 | 1 | 0.75 | 1 | 0.25 | 2 |
| 11 | 1 | 1 | 0 | 0 | 35 |
| 12 | 1 | 1 | 0 | 0 | 4.4 |

| | | | | | |
|---|---|---|---|---|---|
| 13 | 1 | 1 | 0 | 0 | 5 |
| 14 | 1 | 0 | 0 | 0 | 1.4 |
| 15 | 1 | 1 | 0 | 0 | 1.2 |
| 16 | 1 | 0.25 | 1 | 1 | 25 |
| 17 | 1 | 0.75 | 1 | 0 | 3 |
| 18 | 1 | 1 | 0 | 0.25 | 6 |
| 19 | 1 | 1 | 1 | 0.25 | 18 |
| 20 | 1 | 0.75 | 1 | 0.5 | 40 |
| 21 | 1 | 1 | 0 | 0 | 8.3 |
| 22 | 1 | 1 | 0 | 0 | 50 |
| 23 | 1 | 1 | 0 | 0 | 26 |
| 24 | 1 | 1 | 0 | 0 | 5.7 |
| 25 | 1 | 1 | 0 | 0 | 0.7 |
| 26 | 1 | 0 | 1 | 1 | 29.1 |
| 27 | 1 | 0.75 | 1 | 0.25 | 18 |
| 28 | 1 | 0.75 | 1 | 0.25 | 35 |
| 29 | 1 | 1 | 0 | 0 | 29 |
| 30 | 1 | 0.25 | 0 | 0 | 0.6 |
| 31 | 1 | 1 | 0 | 1 | 10 |
| 32 | 1 | 1 | 0 | 0 | 21.7 |
| 33 | 1 | 1 | 0 | 0 | 35.8 |
| 34 | 1 | 0.75 | 1 | 0.25 | 12.9 |
| 35 | 1 | 1 | 1 | 1 | 20 |
| 36 | 1 | 1 | 0 | 0 | 12.9 |
| 37 | 1 | 1 | 0 | 0 | 23.64 |
| 38 | 1 | 0.25 | 1 | 1 | 26.9 |
| 39 | 1 | 0 | 1 | 1 | 15 |
| 40 | 1 | 1 | 0 | 0.25 | 6 |
| 41 | 1 | 0.25 | 1 | 1 | 8 |
| 42 | 1 | 1 | 1 | 0.25 | 6 |
| 43 | 1 | 0.75 | 1 | 0.25 | 27.9 |
| 44 | 1 | 1 | 1 | 0.25 | 4 |
| 45 | 1 | 1 | 0 | 0 | 2.71 |

| | | | | | |
|---|---|---|---|---|---|
| 46 | 1 | 1 | 0 | 0.25 | 1 |
| 47 | 1 | 1 | 1 | 0.25 | 32.7 |
| 48 | 1 | 1 | 1 | 1 | 17 |
| 49 | 1 | 1 | 1 | 1 | 41 |
| 50 | 1 | 0 | 1 | 1 | 11 |
| 51 | 1 | 1 | 1 | 0 | 25 |
| 52 | 1 | 1 | 0 | 0.25 | 7.5 |
| 53 | 1 | 0.75 | 1 | 0.5 | 5 |
| 54 | 1 | 1 | 0 | 0 | 10 |
| 55 | 1 | 1 | 0 | 0 | 2 |
| 56 | 1 | 0.75 | 1 | 0.5 | 34 |
| 57 | 1 | 0 | 1 | 1 | 18 |
| 58 | 1 | 0 | 1 | 1 | 2 |
| 59 | 1 | 1 | 1 | 0 | 45 |
| 60 | 1 | 1 | 0 | 0 | 30 |
| 61 | 0 | 0 | 1 | 0 | 0.7 |

***Table 2. Raw Data***

## *Hypothesis*

If the student lives less than 15 km to the campus, the students prefers to take the public transportation rather than car.

The further away someone lived, the higher tendency they would have to drive in comparison to taking public transit.

In terms of data representation, we initially believed that there would be more people driving to campus in comparison to transiting when the distance was greater than 15 km.

# DATA REPRESENTATION

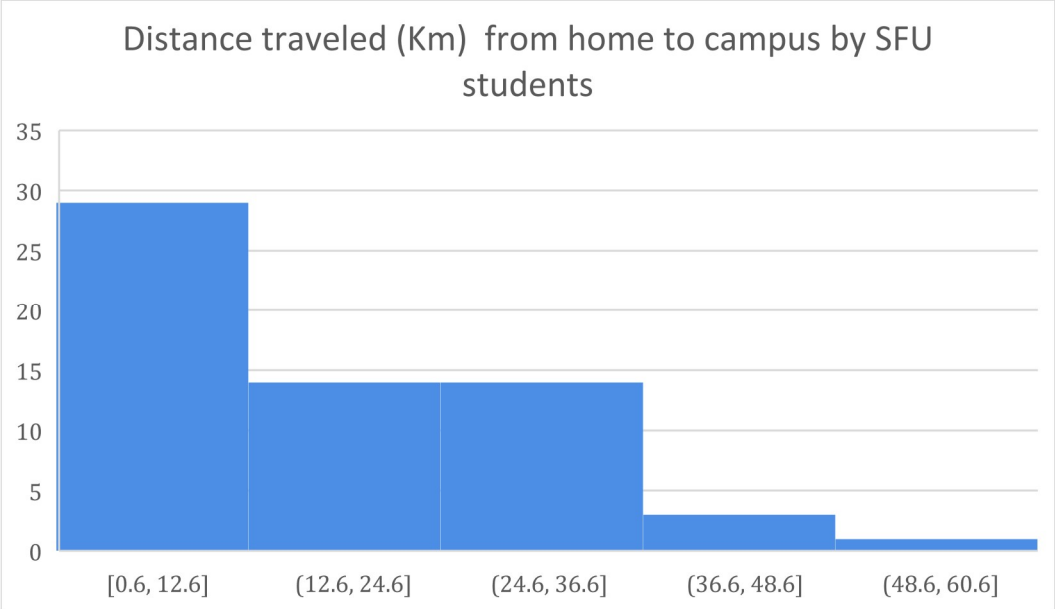



*Figure 2. Distance traveled by both commuters and drivers from home (n=65)*

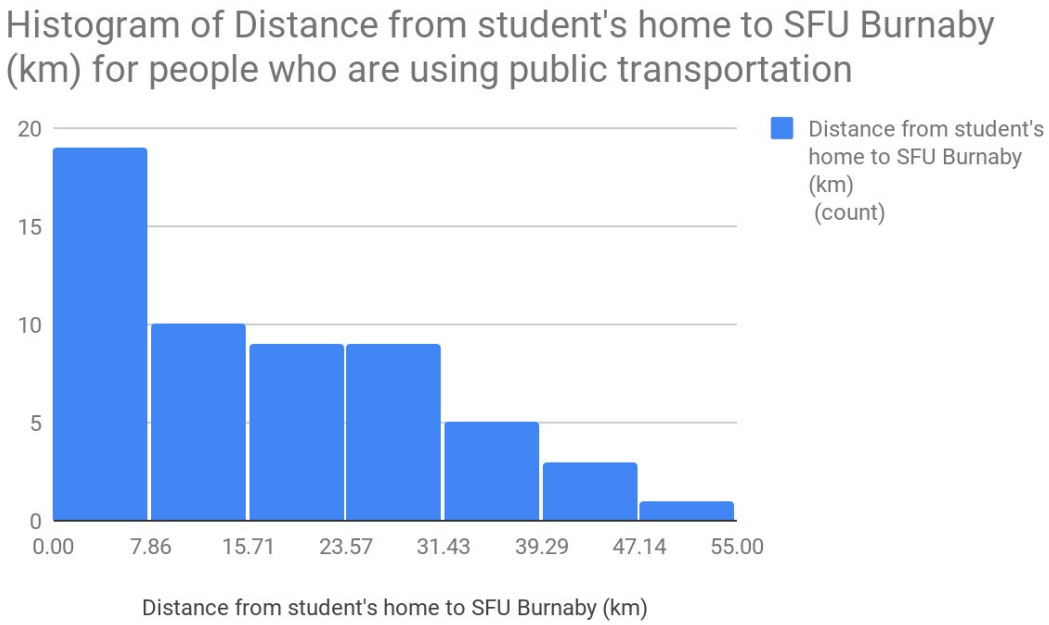



*Figure 3. Histogram representing distance (km) for students using transit to get to SFU*

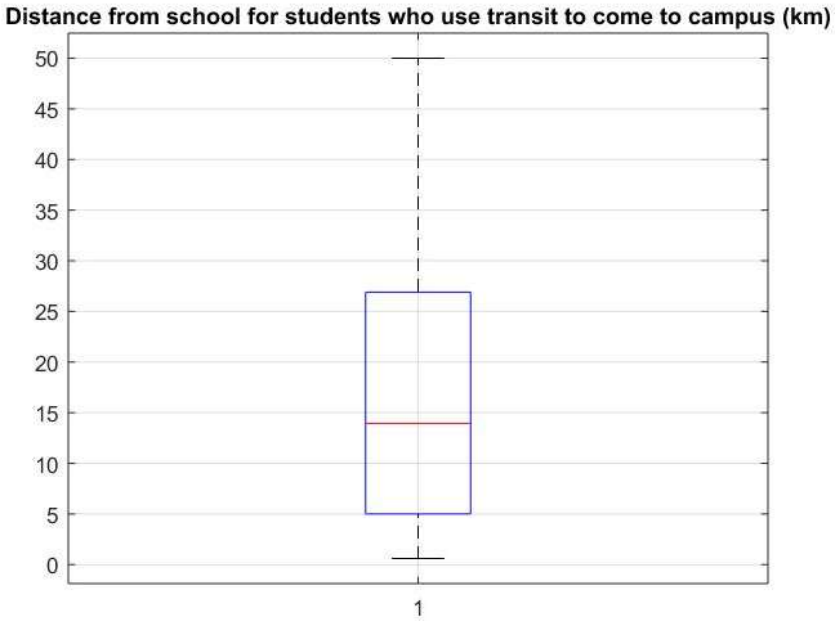



*Figure 4. Boxplot representing distance (km) for students using transit to get to SFU*

The quartiles for the box plot above are as follows:
Q1 = 5 k; Q2 = 13.95km; Q3 = 26.9 km
Minimum =0.6; Maximum = 50

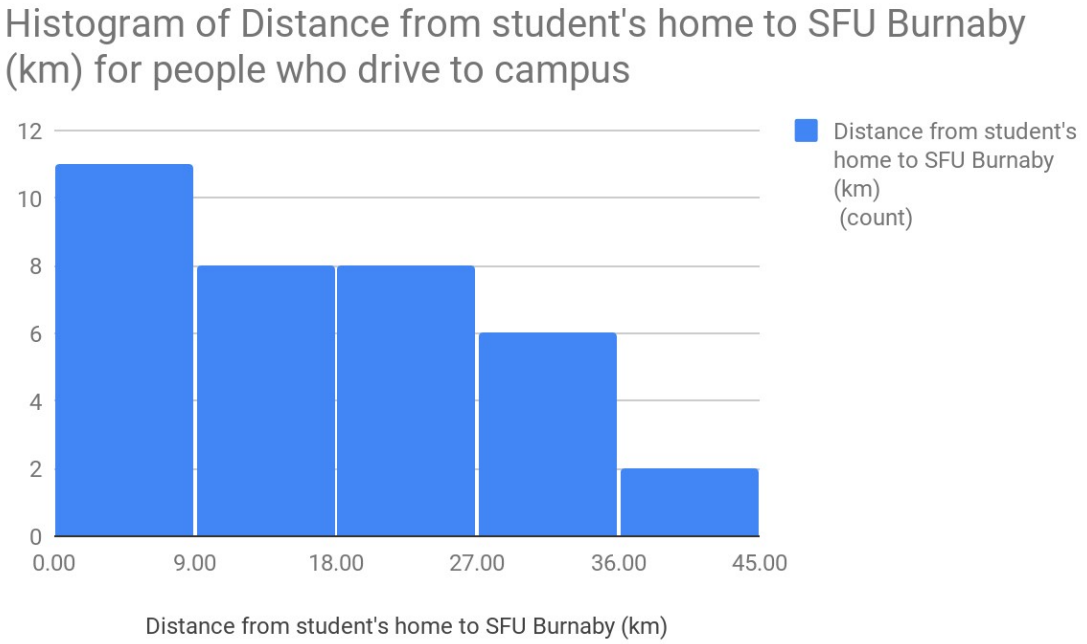



*Figure 5. Histogram representing distance (km) for students using car to get to SFU*

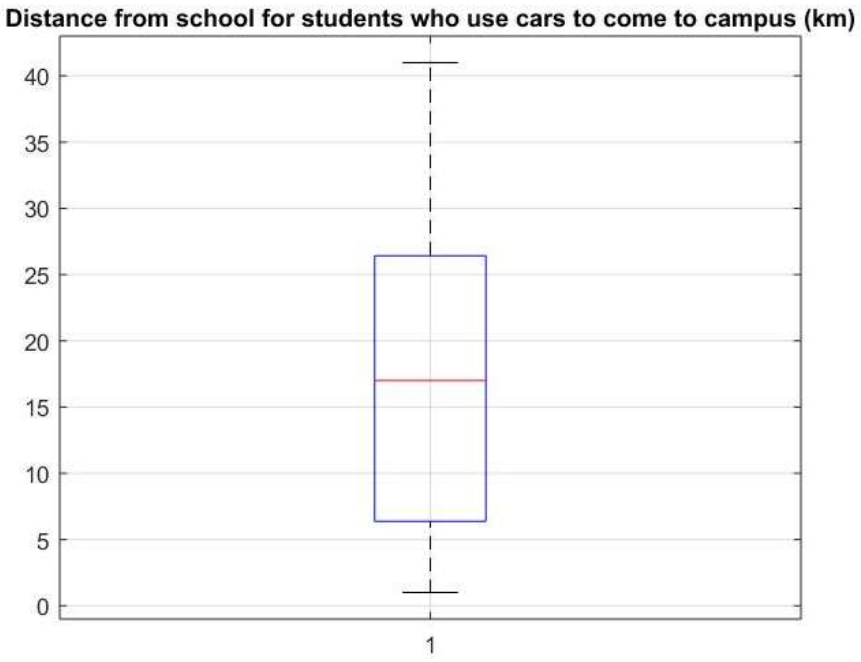



*Figure 6. Boxplot representing distance (km) for students using car to get to SFU*

The quartiles for the box plot above are as follows:
Q1 = 6.375 km Q2 = 17 km Q3 = 26.425 km
Minimum = 1; Maximum = 41

# DATA SUMMARY

## *Calculations of Data:*

Mean and standard deviation were calculated using EXCEL AVERAGE () and STDEV () functions respectively and tabulated below.

| Data | Mean | STDEV | Notes: |
|---|---|---|---|
| Distance from home | 16.26 | 13.05 | Most students live around 15 km from campus with the wider ranges still falling between 28km |
| Tendency to drive | 0.33 | 0.38 | Most students tend to drive a lot less than transit (3 out of 10 times) |
| Tendency to transit | 0.75 | 0.36 | Students tend to transit to school 75/100 of the time, with a few transiting to school with a slight standard deviation |
| U-Pass | 0.98 | 0.13 | Results show that almost all students have U-Pass |
| Access to a car | 0.56 | 0.50 | Half amount of the students has access to the car to transit to campus |

*Table 3. Mean and STDEV for the complete data set and for different parts of the data*

## *Experimental Data and Category Description:*

The distance from the school (SFU Burnaby campus) and home is the only numerical variable that is being considered.

We use this data (distance from SFU) and the answers about whether a person is using U-Pass or a car (personal or rented) to commute to analyses from what distance onwards the students switch between the two modes of transportation.

Category 1: Distance from home to SFU Burnaby campus for students who are transiting to SFU via public transportation

Category 2: Distance from home to SFU Burnaby campus for students who are driving to SFU using a car (personal or rented or car pooled)

# DATA ANALYSIS AND HYPOTHESIS

## *Suspicious Data*

Chauvenet's criterion was applied independently to our main data categories to determine whether data points identified to be furthest from the respective means should be rejected, and the mean and standard deviation recalculated.

Category 1- Public transport users:
For the students who use public transport, the distance that was identified to be furthest from the mean was 50 km for an N = 54 sample.

$\underline{x} = 16.936\ \sigma = 13.270$

$t_{sus} = \frac{|50 \quad .936|}{13.270} = 2.4916$

$$P_{sus} = 2 * P(Z \geq t_{sus})$$

$= 2(1 - \phi(t_{sus}))$

$= 2(1 - 0.993613) = 0.01277$

N*$P_{sus} = 0.6898$

The data point (50 km) should not be rejected since the expected number of such measurements is greater than 0.5. Therefore, the mean and standard deviation of this category remain the same.

Category 2- Car users:
For the students who drive to campus, the distance that was identified to be furthest from the mean was 41 km for an N = 35 sample.

$\underline{x} = 17.12\ \sigma = 11.456$

$t_{sus} = \frac{|41-17.12|}{11.456} = 2.084$

$$P_{sus} = 2 * P(Z \geq t_{sus})$$

$= 2(1 - \phi(t_{sus}))$

$= 2(1 - 0.981237) = 0.0375$

N*$P_{sus} = 1.31$

The data point (41 km) should not be rejected since the expected number of such measurements is greater than 0.5. Therefore, the mean and standard deviation of this category also remain the same.

## *Comparison of Actual Results and Expectations:*

- In terms of population with a U-Pass card, the results are as we expected, since SFU provides a student deal for U-Passes that all students qualify for, unless a student opts out. This is reflected in our survey's results, with one student out of 66 opting out of the U-Pass program.
- With regards to the responses for those who own their own car, we see around half the students own a car. This is correlated with the distance from school. We expected this result, as cars come with many expenses we expected that around half the students would own a car of their own.

The two results above allow us to further understand students' commuting behavior.

From our survey we noticed the following:

- The closer a student lived to campus, the higher their tendency to transit to school.
- Students who own their own car had a higher tendency of driving to school, than those who did not own their own car. Nevertheless, more than half of students who own their own car still do NOT drive to school. This aligns with our predictions due to the increased gas and insurance prices. Refer to plots #1 and #2 below.
- Looking at Plot #3 below, we notice that most students with a U-Pass transit to school daily, if not at least one a week. The 7 students who do not transit to school yet pay for a U-Pass are students included in Plot #1, i.e.: they chose to drive the cars they own to school over transiting to school.
- Comparing the distance from home and frequency of using a U-Pass, we did not find a distinct trend between distance from home and choosing to use a U-Pass.

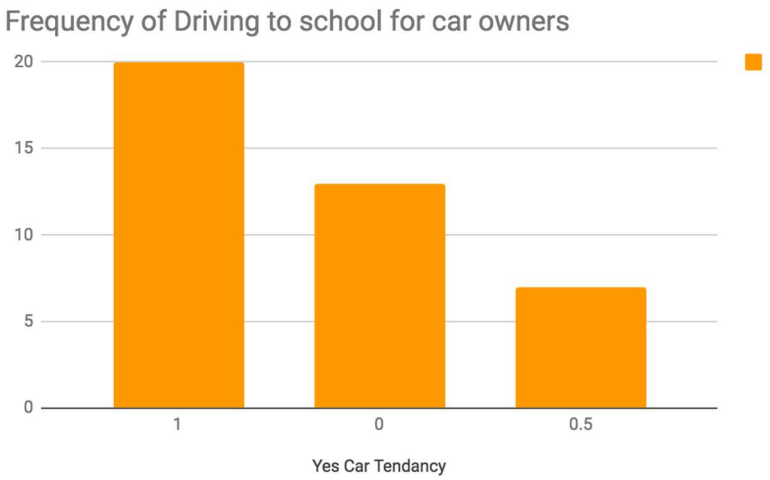



*Figure 7. Frequency of driving to school for students who own their own cars*

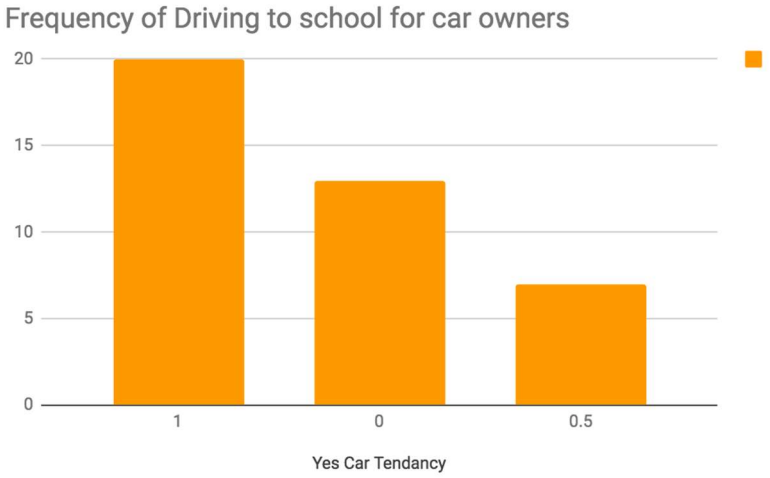



*Figure 8. Frequency of driving to school for students who do not own their own cars*

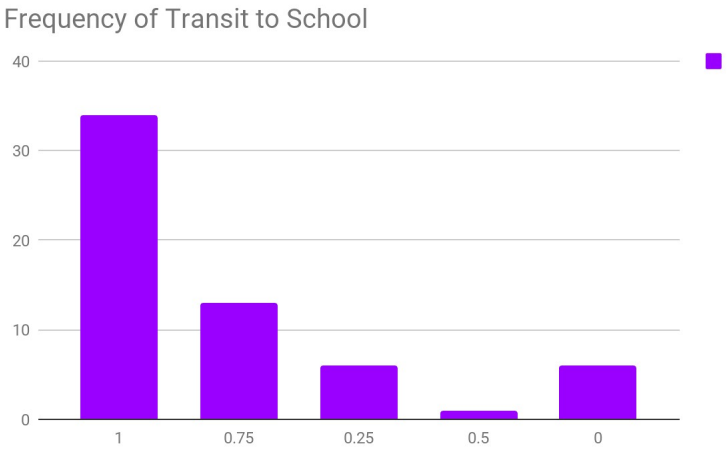



***Figure 9. Frequency of Transiting to school for all students who have U-Pass (n=65)***

## *Probability Analysis*

I. **Conditional Probability Question:** How often do students drive to school if they are living more than 15 km away?
P (drive | live more than 15km away) = 21/29

This probability shows that for the students, who live 15 or more km away from campus, probability of choosing to commute by car higher than choosing to commute by public transit

II. **Unconditional Probability Question**: What is the probability of driving to campus?
P(drive) = (35/62)

This probability shows that more than half of the students drive to campus. There is an overlap in the data collected between students who commute and those who drive because some students will drive on some days and commute on others. Therefore, this probability is higher than would have been expected for students.

# HYPOTHESIS TESTING

## *Part 1: Single Sample Test*

Parameter chosen:
Mean for distance traveled by all students from home

$H_o$: $\mu_o = 19.31$
$H_1$: $\mu_o < 19.31$

$\bar{X} = 16.26$
$\mu_o = 19.31$
$s = 13.05$
$n = 62$

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$=> \frac{(16.26-19.31)}{\frac{13.05}{\sqrt{62}}} = -1.833$$

$t_o$ = -1.833 from normal distribution table, p-value = 0.0329

Our criteria to reject is if $t_o < -t_{\alpha, n-1}$ :

$t_o$ = -1.833 $t_{0.05, 61} = 2.00$ but -1.833 > -2.000

- Fail to reject $H_o$ : $\mu_o = 19.31$
- Deductions: There is not enough evidence to suggest that the average distance traveled by SFU students to campus is less than 19.31 Km

## *Part 2: Single Sample Test ("Fail to reject $H_o$")*

$H_o$: $\mu_o = 18.75$
$H_1$: $\mu_o < 18.75$

$\mu_o = 18.75$
$s = 13.05$
$\bar{X} = 16.26$
$n = 62$

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$=> \frac{(16.26-18.75)}{\frac{13.05}{\sqrt{62}}} = -1.5$$

$t_o$ = -1.5 from normal distribution table, p-value = 0.0668

Our criteria to reject is if $t_o < -t_{\alpha, n-1}$ :

$t_o$ = -1.5 $t_{0.05, 61} = 2.00$ but -1.5 > -2.000

- Fail to reject $H_o$ : $\mu_o = 18.75$
- Deductions: There is not enough evidence to suggest that the average distance traveled by SFU students to campus is less than 18.75 Km

## *Part 3: 2-Sample Test A*

Test Hypothesis: $H_o : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$
Known means: $\overline{x_1}$ = 17,000 $\overline{x_2}$ = 13,950
Known sample sizes: $n_1 = 35$ $n_2 = 54$
Known standard deviations: $s_1 = 11456$ $s_2 = 13270$
Assuming variances unknown and $\sigma_1^2 \neq \sigma_2^2$

$$t_0 = \frac{17{,}000 - 13{,}950}{\sqrt{\frac{11{,}456^2}{35} + \frac{13{,}270^2}{54}}} \quad t_0 = 1.15191$$

P-value (probability above $|t_0|$ and below $-|t_0|$:
$= 2 \times \Phi(1.152) - 1$
$= 2 \times (0.874928) - 1$
P-value =0.749856

0.749856 > (α=0.05); fail to reject $H_0$. There is insufficient evidence to conclude that the means of the two sets are different at α=0.05.

## *Part 3: 2-Sample Test B*

$H_0 : \sigma_1^2 = \sigma_2^2 \quad H_1 : \sigma_1^2 \neq \sigma_2^2$

$$f_0 = \frac{s_1^2}{s_2^2} = \frac{11{,}456^2}{13{,}270^2} = 0.74529$$

P-value: $2 \times \Phi(0.745) = 2 \times (0.773373) = 1.548$
Criteria for rejection: $f_0 > f_{\propto, n_{1-1}, n_{2-1}}$
$f_0 > f_{0.025,34,53} \approx f_{0.025,30,60} \qquad f_{0.025,34,53} \approx 1.940$

$0.745 < 1.940$ ; fail to reject null hypothesis. There is insufficient evidence to support the claim that $\sigma_1^2 \neq \sigma_2^2$ .

# CONCLUSION

Due to SFU's student U-Pass plan, most of the students at SFU Burnaby campus commute to campus via public transit. Among these students, only a small number drive to campus if they have their own car. For students who own their own cars, 50% drive to campus. This driving trend is observed due to a person choosing to drive to campus on some days even if they have a U-Pass. This is made possible by the various car rental services that are available on campus such as Evo, Modo, Zip Cars, etc. Among people who own their own cars, only a small number chooses to drive all the time to get to campus because of the high prices of gas. Students who live more than 15-20 km away from campus, drive to campus if they own or can rent a car but if they cannot do that then they use public transport.

In summary, upon comparing the distance from home and frequency of using a U-Pass, even though we did find out that 75% of the students prefer to transit to campus on a normal day and 50% drive to campus, we did not find a distinct trend between distance from home and choosing to use a U-Pass as some people still preferred to drive if they had easy access to a car even if they live closer. While people living further away than 15-20 km would transit if they don't have a car or cannot drive.

# AMENDMENTS

## *Changes made to Phase I*

- Removed “Probability” from the title
- A Table of Contents was added, and the rest of the document was better organized
- The variables made clearer and more emphasized that we are using the type of commute the students prefer depending on how far they live to the campus
- Raw Data was linked to the excel file that is submitted along with this document
- An explanation or description was added for the independent, dependent and numerical data being used

## *Changes made to Phase II*

- Added histogram showing distance traveled by SFU students (both those who transit and drive)
- Corrected unconditional probability
- Corrected formatting

# MEMBERS' INVOLVEMENT

| | Phase I | Phase II | Phase III | Points Given |
|---|---|---|---|---|
| Sirpreet Kaur Dhillon | - Helped to shape the idea<br>- Edited the report<br>- Edited the survey | - Formatting Phase I<br>- Raw Data changes<br>- Excel spreadsheet<br>- Histograms<br>- Experimental Data and Category Description<br>- Hypothesis | - Introduction<br>- Conclusion<br>- Abstract | 1 |
| Ece Eskikurt | - Come up with the raw idea<br>- Created the draft of the survey<br>- Shared survey on social media for data collection | - Formatting Phase I<br>- Hypothesis<br>- Probability analysis<br>- Table of contents | - Part 2<br>- Formatting | 1 |
| Tatendaishe Jakaza | - Helped to shape the idea<br>- Created the draft of the report<br>- Edited the survey | - Formatting<br>- Made changes to Chauvenet's analysis<br>- Changed mean and standard deviation<br>- Boxplot<br>- Deleted data points that had 0 km distance | - Part 3 (a) and (b)<br>- Formatting<br>- Corrections for Phase 2 | 1 |
| Reem Mustafa | - Helped to shape the idea<br>- Edited the survey | - Formatting<br>- Initial Chauvenet's criterion analysis<br>- Mean and standard deviation<br>- Frequency plots | Part 1; edited abstract | 1 |

*Table 4: Member Involvement*