

Various Demographic Factors Effect on Students' Participation In ECA's*

Data Collected from STA304 Class

Vanshika Vanshika Navya Hooda Shea Alexia Chloe

November 11, 2024

It is known that for university students to have a well-rounded university experience, participation in extra-curricular activities is vital, not only for academic but personal growth. However, student participation in extracurricular activities depends on a number of factors. Some include scheduling and time-commitment factors, commute factors and the diversity of student activities/clubs. This study aims to explore this by investigating the reasons that students select certain extracurricular activities at the University of Toronto Mississauga (UTM) campus. An anonymous online survey on Google Forms was created for students in a third year Statistics course to complete. Findings showed that there was no significant relationship between campus distance and program of study with extra-curricular involvement. From Figure 2, the boxplot shows that there was some gender influence on the participation of extra-curricular activity, however this variation is not statistically relevant as the p-value was 1. Although results did not produce any significant findings, we believe that this is a valiant step in the right direction towards investigating student participation at UTM.

Table of contents

1	Introduction	2
1.1	Research Questions and Hypotheses	2
2	Data	3
2.1	Methodology	3

*Code and data are available at: <https://github.com/vanshikav2/Immigration>.

3	Analysis	3
3.1	Research Question 1	6
3.2	Research Question 2	9
3.2.1	Campus Distance and Activity Count	10
3.3	Research Question 3	13
3.3.1	Time Commitment and Student Involvement	13
3.3.2	Time Commitment and Activity Count	15
4	Discussion	16
5	Limitations	17
6	References	18

1 Introduction

In our study, we were interested in the participation of STA304 students in extracurricular activities (ECAs). In particular, we analyzed if various demographic factors affected students' participation in ECAs. During the first three weeks of October 2024, we distributed a survey on the STA304 Piazza discussion board in order to collect data on student demographic factors and information on individual participation in ECAs. Our study consisted of three research questions:

1.1 Research Questions and Hypotheses

1. **Research Question 1:** What is the most preferred type of extracurricular activity, and do demographic factors influence student preferences for certain extracurricular activities?
 - **Ho:** Student demographic factors have no effect on participation in extracurricular activities.
 - **Ha:** Participation in extracurricular activities is affected by demographic factors.
2. **Research Question 2:** Does proximity to campus affect participation in extracurricular activities?
 - **Ho:** Proximity to campus has no effect on participation in extracurricular activities.
 - **Ha:** Students living further from campus will have lower participation rates.
3. **Research Question 3:** Does timing/scheduling of extracurricular activities impact student participation in extracurricular activities?
 - **Ho:** Timing/scheduling of extracurricular activities has no impact on student participation.

- **Ha:** Participation in extracurricular activities is affected by their timing/scheduling.

The remainder of the research paper will be structured as follows: Section 2 covers data collection and methodology; Section 3 presents survey findings and data analysis; Section 4 goes over the interpretation of our results; Section 5 discusses the study limitations; ?@sec-data4 concludes the paper; ?@sec-data5 contains an appendix.

2 Data

2.1 Methodology

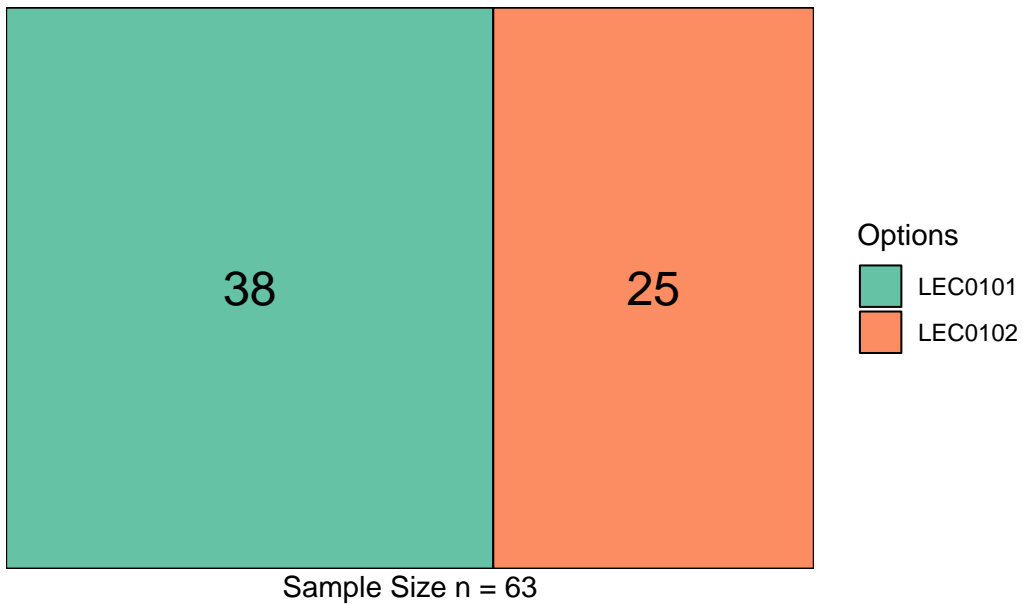
Data collection began in the first three weeks of October 2024, an online anonymous survey on Google Forms was created for students to complete. This survey was created to investigate the reasons that students select and participate in certain ECAs. This online survey was posted to a Piazza thread where students can easily locate the survey link. The sampling method used was Stratified Random Sampling where the strata were the STA304 lecture sections (ie. LEC0101 and LEC0102). Randomness was ensured as the link to the survey was posted to Piazza where everyone had the equal opportunity to complete the survey and be chosen for the study. We collected a sample of 200 students. The survey consisted of 10 questions asking for details on their lecture section, demographic information, program of study, and distance from the UTM campus. We also included questions on the number, types of activities and the time commitment of the ECAs they are involved in. To encourage students to complete our online survey, we asked participants to share a link to their group's survey, which we would complete in return.

3 Analysis

The majority of the analysis for our research questions is based on the various student demographic factors. As such, it is worthwhile to go over, in a preliminary fashion, the results of our survey to give further background on the study before continuing into the analysis of the research questions.

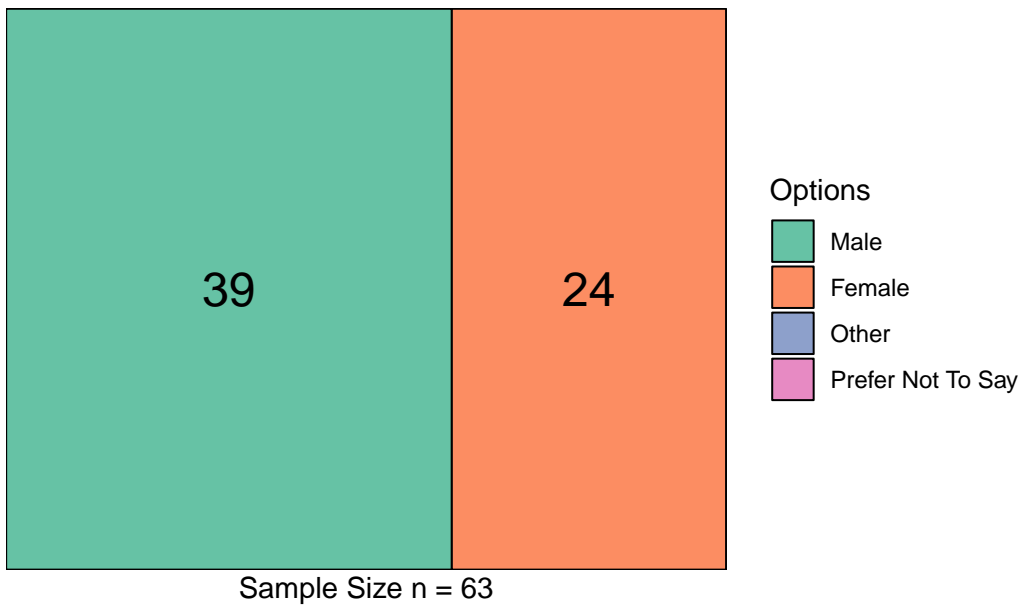
There were slightly more responses from LEC0101, approximately a 6:4 ratio LEC0101 to LEC0102.

Lecture Section Breakdown

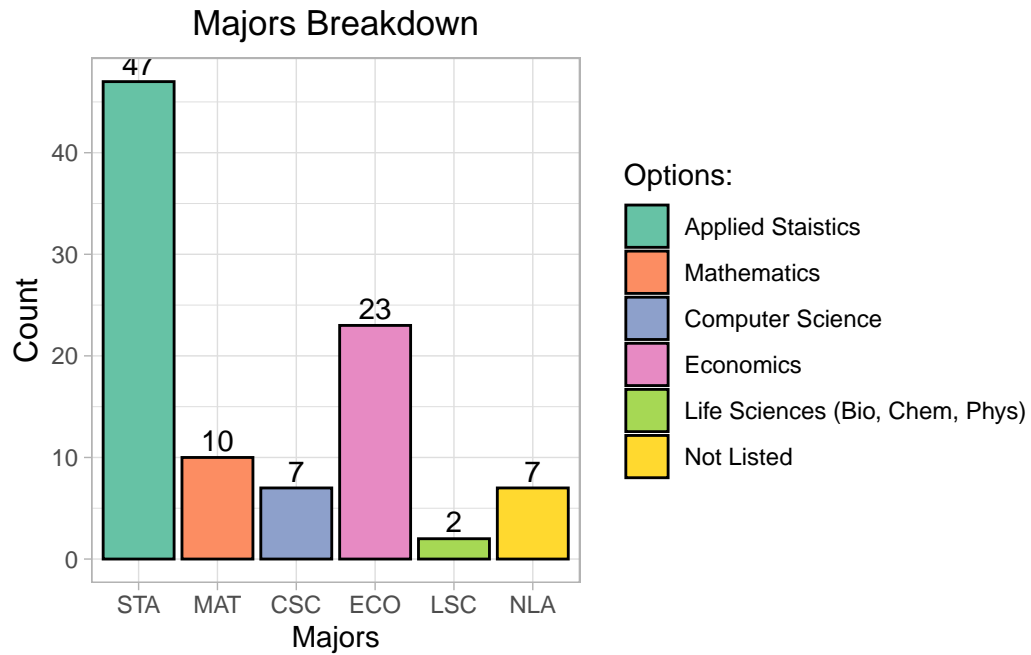


There was a higher percentage of males in the class, approximately a 6:4 ratio, with no students responding Other or Prefer Not to Answer.

Gender Identity Breakdown

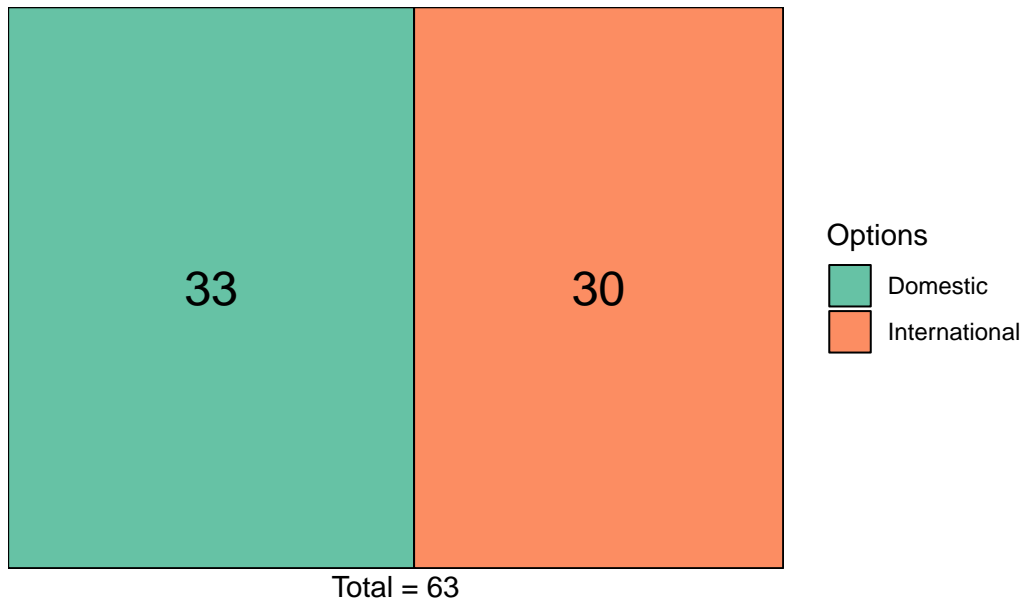


Students picked up to 2 majors, with STA seeing a significant share.

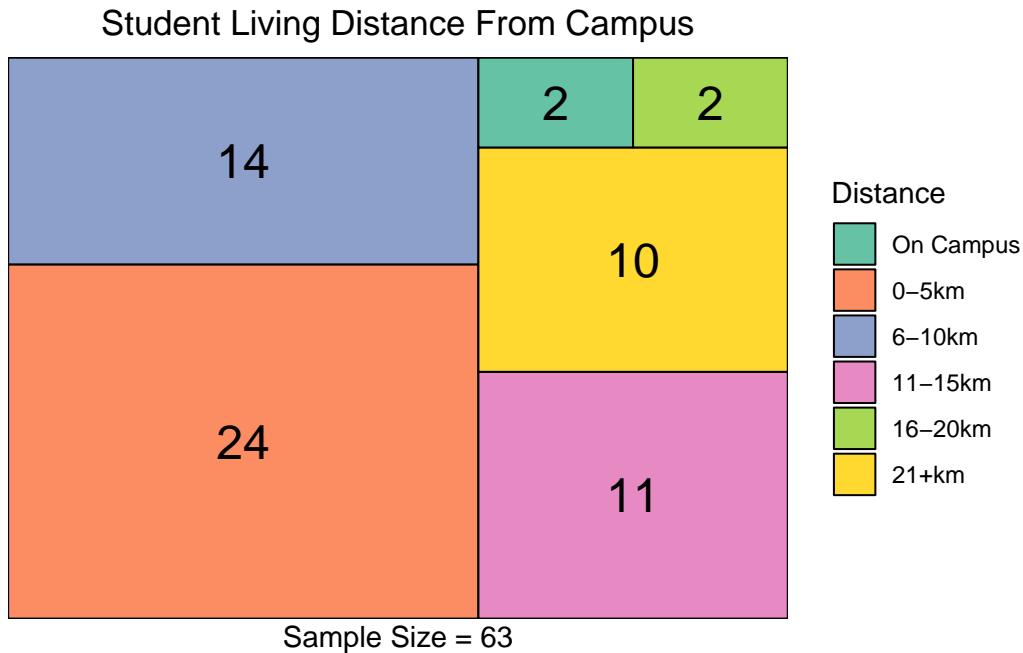


We had almost an even balance of Domestic and International students.

Student Status in Canada



Finally, we recorded students living distance from campus.



3.1 Research Question 1

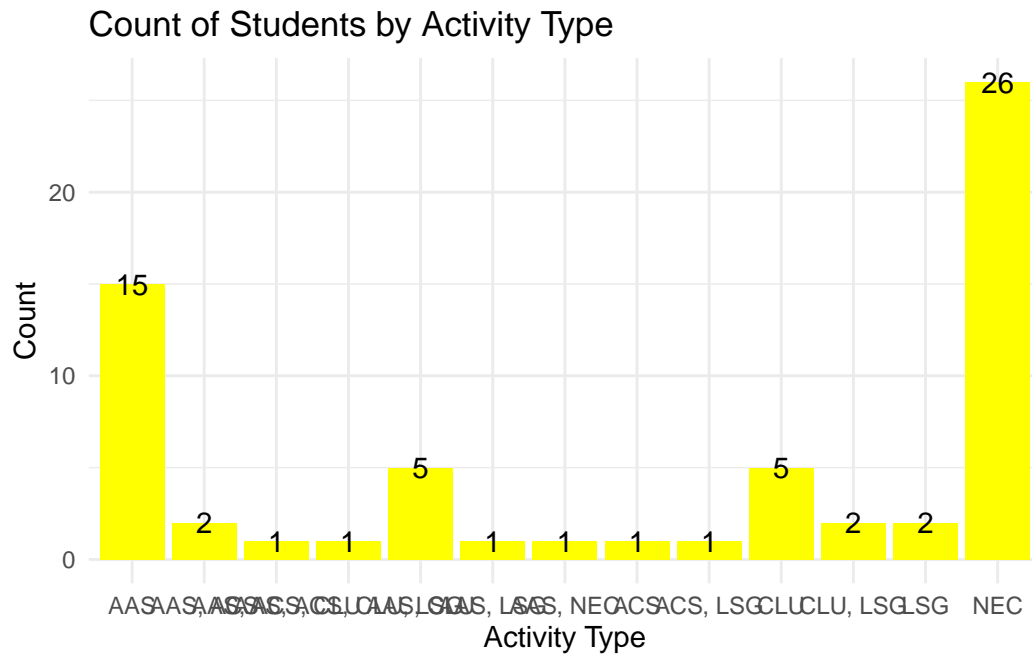
In RQ1, we analysed what is the most preferred type of extracurricular activity and do demographic factors influence people's selection or preference for certain extracurricular activities.

NOTE: For easy analysis and ability to come up with clearer conclusions, the Major column and the activity column were split into different rows.

For the most preferred type of extracurricular activity, we were able to get the count of students in the different extracurricular activities and display a graph showing the counts of students in each extracurricular activity.

```
# A tibble: 13 x 2
  activityType count
  <chr>      <int>
1 NEC          26
2 AAS          15
3 AAS, CLU       5
4 CLU           5
5 AAS, ACS       2
6 CLU, LSG       2
7 LSG           2
8 AAS, ACS, CLU  1
```

9	AAS, ACS, CLU, LSG	1
10	AAS, LSG	1
11	AAS, NEC	1
12	ACS	1
13	ACS, LSG	1



The most popular extracurricular activity among students was Athletics and Sports (AAS) and the least popular extracurricular activity is Academic Societies (ACS). It was also particularly interesting to observe that most students do not participate in Extracurricular Activities, making the NEC category second in popularity.

To investigate the proportion of students participating in each activity by demographic factors like major, studentStatus and genderIdentity. These diagrams are able to provide an initial view of how activity participation varies across the different groups.

```
# A tibble: 32 x 4
# Groups:   major [10]
  major activityType count proportion
<chr>   <chr>      <int>      <dbl>
1 ECO    AAS           1      0.125
2 ECO    ACS           1      0.125
3 ECO    LSG           2      0.25
4 ECO    NEC           4      0.5
```

5	MAT	NEC	2	1
6	MAT, CSC	NEC	1	1
7	NLA	AAS	3	0.6
8	NLA	NEC	2	0.4
9	STA	AAS	2	0.133
10	STA	AAS, ACS	1	0.0667
11	STA	AAS, ACS, CLU, LSG	1	0.0667
12	STA	AAS, CLU	2	0.133
13	STA	CLU	2	0.133
14	STA	NEC	7	0.467
15	STA, CSC	AAS	3	0.5
16	STA, CSC	AAS, ACS, CLU	1	0.167
17	STA, CSC	NEC	2	0.333
18	STA, ECO	AAS	5	0.333
19	STA, ECO	AAS, ACS	1	0.0667
20	STA, ECO	AAS, CLU	1	0.0667
21	STA, ECO	AAS, LSG	1	0.0667
22	STA, ECO	ACS, LSG	1	0.0667
23	STA, ECO	CLU	1	0.0667
24	STA, ECO	CLU, LSG	1	0.0667

i 8 more rows

From the visual representation we can see that there is no significant difference in activity participation across majors.

We also conducted an ANOVA test, and we can see that the boxplot aligns with our results. At a significance level of 0.05, We had an F-Value of 0.646 and a p-value of 0.668 indicating that there was no variation in the activity participation in the different majors to be considered significant.

```
# A tibble: 18 x 4
# Groups:   genderIdentity [2]
  genderIdentity activityType count proportion
  <chr>          <chr>      <int>      <dbl>
1 F            AAS          4        0.167
2 F            AAS, CLU      1        0.0417
3 F            AAS, LSG      1        0.0417
4 F            AAS, NEC      1        0.0417
5 F            ACS          1        0.0417
6 F            ACS, LSG      1        0.0417
7 F            CLU          4        0.167
8 F            CLU, LSG      2        0.0833
9 F            LSG          1        0.0417
```



```
10 F          NEC          8      0.333
# i 8 more rows
```

From the Boxplot, we can see that the boxplot for the females is relatively narrow whilst that of the males is wide. We can see that males have a wider range of participation in different activities compared to women. This variation can suggest that there is some gender influence on activity participation.

After conducting the T-test, at 0.05 significance level, we can conclude that this variation is not statistically significant as the p value is 1 indicating that we accept the null hypothesis that there is no significant difference in the mean proportion of activity participation between females and males. Both genders have the same average participation of 0.2.

```
# A tibble: 18 x 4
# Groups:   studentStatus [2]
  studentStatus activityType      count proportion
  <chr>         <chr>         <int>      <dbl>
1 D           AAS              10      0.303
2 D           AAS, ACS           2      0.0606
3 D           AAS, ACS, CLU       1      0.0303
4 D           AAS, ACS, CLU, LSG   1      0.0303
5 D           AAS, CLU           3      0.0909
6 D           ACS              1      0.0303
7 D           CLU              3      0.0909
8 D           CLU, LSG           1      0.0303
9 D           NEC              11      0.333
10 I          AAS              5       0.167
# i 8 more rows
```

The t-test results show a t value of 0 and a p-value of 1 indicating that there is no significant difference in the mean proportions of activity participation between the two groups. We can also see that the two groups have the same mean proportion of 0.2.

With this, we have observed that there is no significant influence of major, student status and gender on activity participation. It suggests that the choice of extracurricular activities is independent of one's demographics.

3.2 Research Question 2

In RQ2, we analysed if the proximity to campus affects one's ability to participate in extracurricular activities.

- Null Hypothesis (H_o): Proximity to campus (campusDistance) does not affect participation in extracurricular activities.
- Alternative Hypothesis (H_1): Proximity to campus (campusDistance) has an effect on participation in extracurricular activities.

3.2.1 Campus Distance and Activity Count

Statistical Test: To evaluate this, we used the correlation analysis test between campus distance and the number of extracurricular activities using Pearson's correlation test. Prior to conducting this test, we will verify the assumptions for the test first.

Table 1: Correlation Analysis

Statistic	CampusDistance	ActivityCount
Min	0.000000	0.000000
1st Quartile	1.000000	0.000000
Median	2.000000	1.000000
Mean	2.269841	1.047619
3rd Quartile	3.000000	2.000000
Max	5.000000	4.000000

The descriptive statistics suggest that most students live fairly close to campus (within 2-3 units) and participate in 1-2 extracurricular activities. This variability in both campusDistance and activityCount supports further analysis to see if there's a relationship between proximity and participation.

Moving on we inspect the relationship between the variables. As it can be seen from Figure 1, there seems to be no linear relation between the variables.

The Q-Q plot displays how the residuals (differences between observed and fitted values) align with the normal distribution. In this case, residuals do not fully align along the red diagonal line, indicating that normality is somewhat violated. The observed points deviate in a stepwise manner, particularly at the tails, which suggests potential non-normality in residuals.

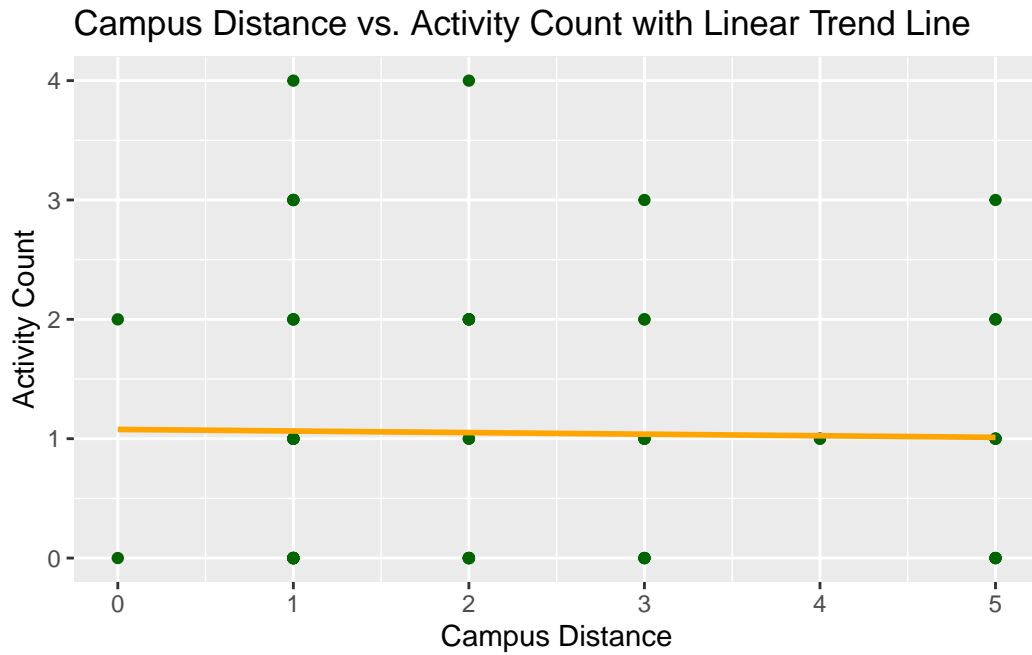


Figure 1: Scatter Plot

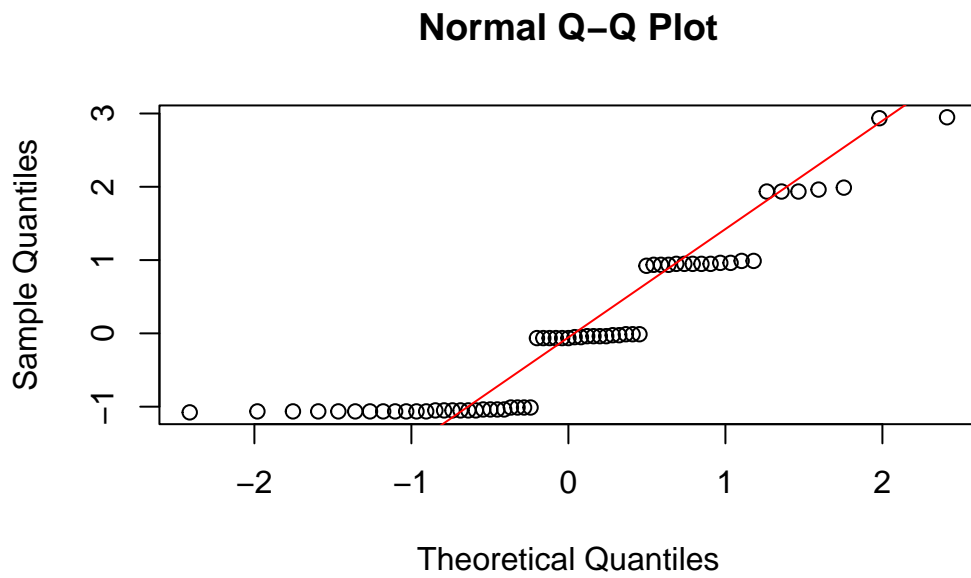
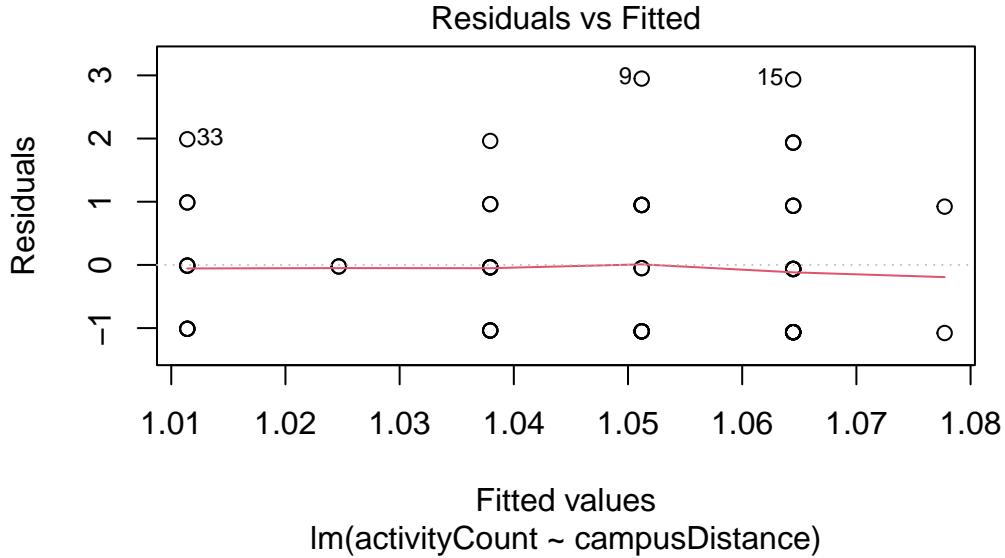


Figure 2: Scatter Plot



This plot helps in identifying any non-linearity, unequal variance (heteroscedasticity), or outliers. Here, residuals seem randomly scattered around the fitted line with no obvious pattern, suggesting homoscedasticity (constant variance). However, a few points like 33, 9, and 15 appear as potential outliers, which might influence the results.

Since there is non-normality between these variables, we proceed with Spearman's rank correlation to analyze the association between campus distance and activity count.

Table 2: Correlation Test Results

Test	Correlation_Coefficient	p_value	Significance
Pearson's Correlation	-0.018	0.891	Not Significant
Spearman's Correlation	0.025	0.847	Not Significant

The Spearman's rank correlation test results are as follows:

- Spearman correlation coefficient: 0.025
- p-value: 0.847

This result indicates a very weak positive correlation between `campusDistance` and `activityCount`, but it is not statistically significant (p-value > 0.05). Therefore, we fail to reject the null hypothesis, suggesting no significant association between proximity to campus and participation in extracurricular activities based on this data.

3.3 Research Question 3

In RQ3, we analyzed if the the timing, the time spent on extra curricular activity affected one's level of participation in extracurricular activities.

- Null Hypothesis (H_0): Time Commitment (timeCommitment) does not affect participation levels in extracurricular activities.
- Alternative Hypothesis (H_1): Time Commitment (timeCommitment) has an effect on participation levels in extracurricular activities.

We used the variables of student involvement (studentInvolvement) and activity count (activityCount) to represent the participation levels.

3.3.1 Time Commitment and Student Involvement

Statistical Test: To evaluate this, we used the correlation analysis test between time commitment and student involvement variables using Spearman's correlation test. Before proceeding with the test, we will verify the assumptions for the test.

Normality of Variables: In our exploration we found that the variables of timeCommitment (TC), and studentInvolvement (SI) are not normally distributed as seen by the QQ plots in Figure 3 as they deviate from the normality line. Hence we proceeded to apply the correlation test using the Spearman's rank coefficient instead of Pearson's coefficient (requiring normality).

Monotonic Association: Next we confirmed the assumption of Spearman's rank coefficient test requiring a monotonic relationship between the variables (timeCommitment and studentInvolvement). We found that the relationship between these variables was monotonic, specifically increasing as seen in Figure 4.

Results: As seen in Table 3, the Spearman's rank correlation coefficient is 0.808, which indicates a strong positive correlation, pointing to a significant monotonic relationship between timeCommitment and studentInvolvement. p-value is < 0.05 , hence the correlation is statistically significant, and we can reject the null-hypothesis.

Table 3: Correlation Test Results

Statistic	Value
Spearman Coefficient (rho)	0.808
p-value	1.144e-15

Hence students who spend more time in the activities have higher participation in extra curricular activities measured by the activity count, as suggested by the strong association based on our data.

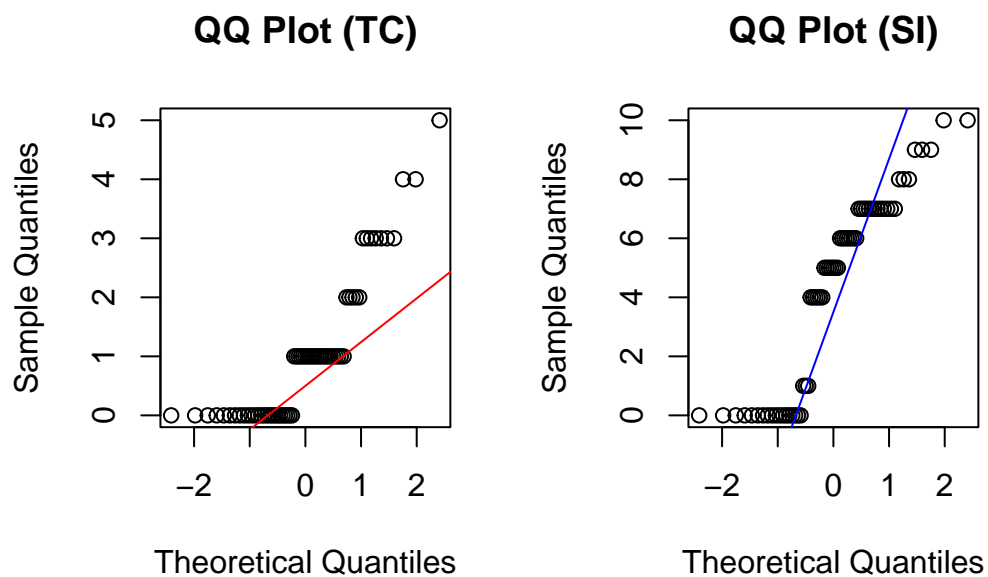


Figure 3: Normality of Variables

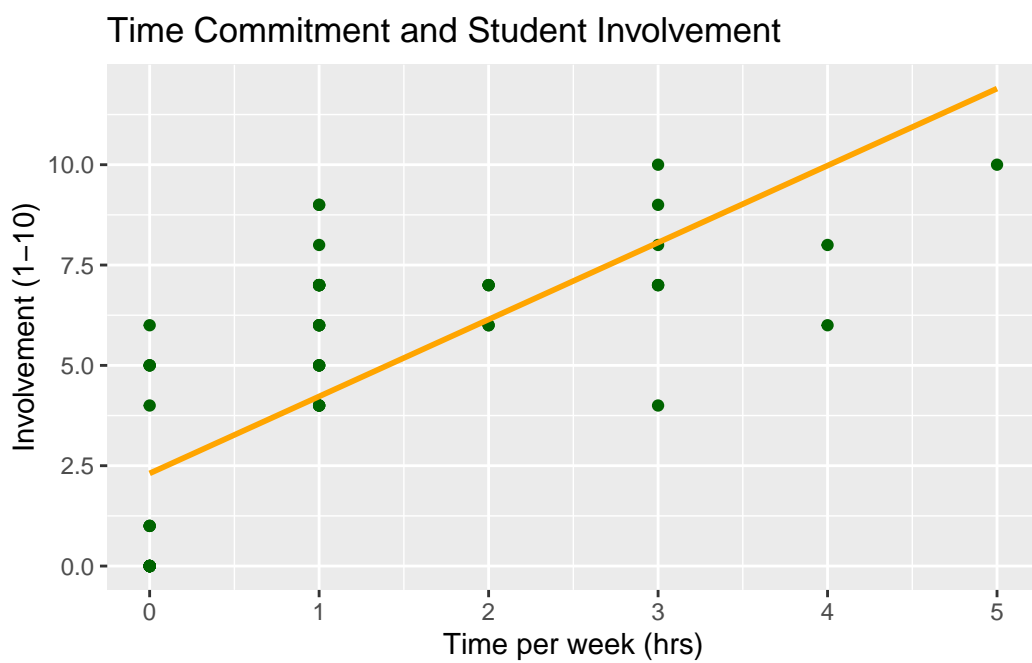


Figure 4: Checking Monotonicity

3.3.2 Time Commitment and Activity Count

We wanted to check whether the time commitment and participation measured by the activityCount variable were correlated.

Statistical Test:

As discussed before, both the timeCommitment and activityCount variables did not meet normality assumptions, so we proceed with the Spearman's rank correlation coefficient as our statistical test.

Normality of Variables:

To proceed with this test, we used QQ-plots to verify normality of timeCommitment (TC) and activityCount(AC). As seen in Figure 5 they are non-normal variables due to their deviation from the normal line.

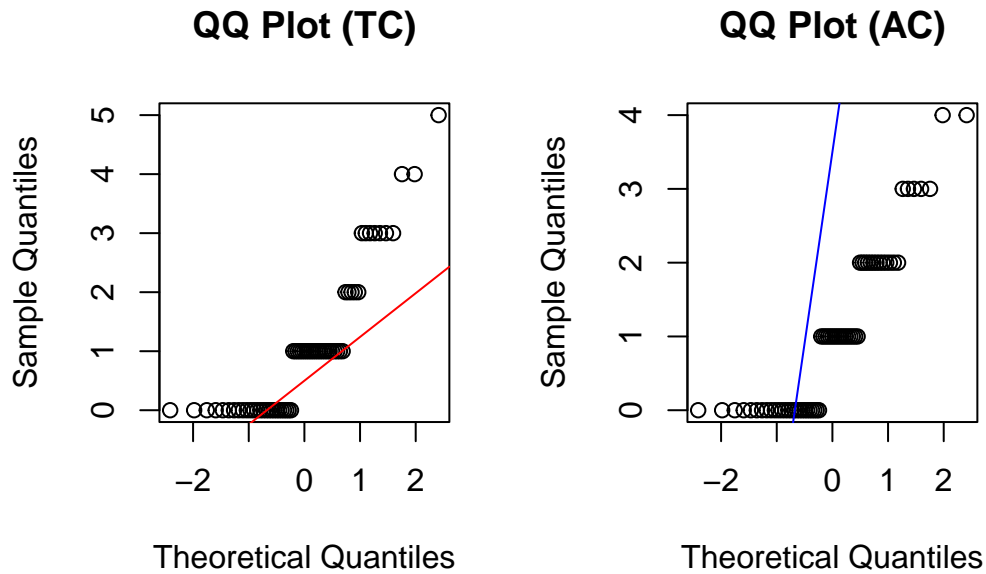


Figure 5: Normality of Variables

Monotonic Association: We confirmed that the relationship between these variables was monotonic, specifically increasing as seen in Figure 6

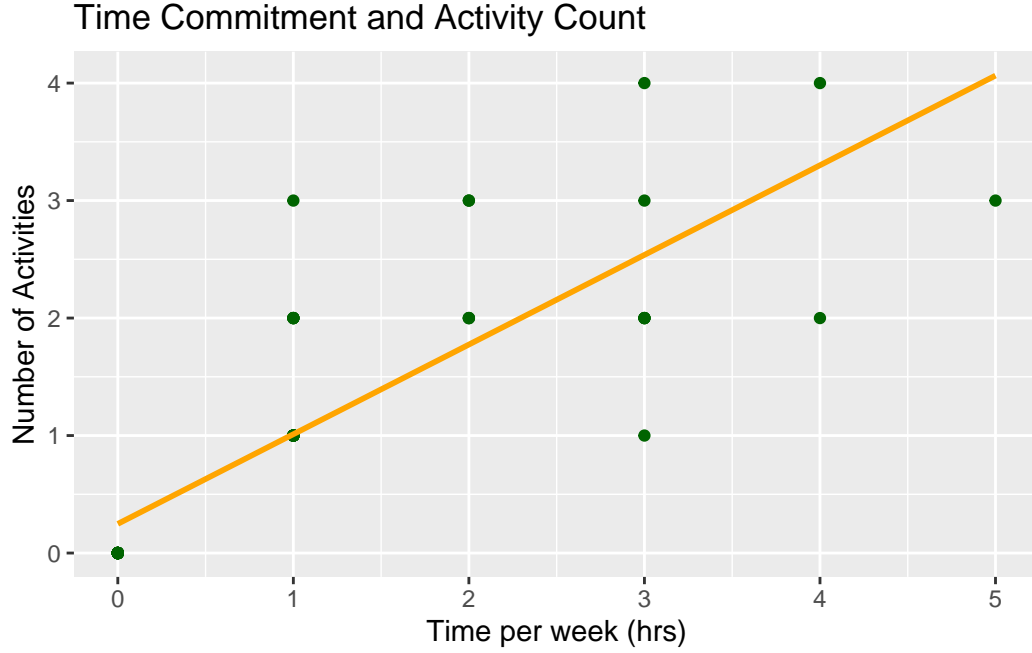


Figure 6: Checking Monotonicity

Table 4: Correlation Test Results

Statistic	Value
Spearman Coefficient (ρ)	0.937
p-value	1.346e-29

Results: The test results show that the coefficient value is 0.937, indicating a strong positive relationship. This would mean as time commitment increases, so does activity count. The p-value is < 0.05 and indicates strong evidence against the null hypothesis, hence the relationship is unlikely due to chance.

Hence, these results indicate a strong, statistically significant monotonic relationship between time commitment and activity count. As students increase their time commitment, so does their participation measured by involvement.

4 Discussion

In this study, we sought to identify the factors that significantly contribute to the Gross Domestic Product (GDP) and Inflation Rate. We applied a Bayesian generalized linear model

to a dataset encompassing a range of economic indicators over a period of years.

5 Limitations

This study aimed to understand the reasons why students select and participate in ECAs at UTM. Although our results did not yield significant findings, several limitations may have contributed to this outcome.

One such notable limitation is the way in which our data was collected. Although we proposed to collect data using a Stratified Random Sample, the data was analysed using a Simple Random Sample. Although randomness was ensured throughout data collection, the data was not analyzed with strata.

Another notable limitation is the possible selection bias as most participants were of the LEC0101 lecture section.

In the future studies, we would remove the multi-select option when choosing a program of study and days of the week students participate in activities as it can simplify data analysis by making it easier to categorize and aggregate data.

6 References