STA304 Technical report

1 Abstract

In campus life, students' satisfaction is determined by many factors, like academic stress, the time spent on study, extracurricular activities, etc. We made this survey to study the factors of students' satisfaction with campus and to improve and help students increase their campus happiness. In a third-year statistics survey and sampling course, the data was collected using the online Google form and the questionnaires through Piazza.

As shown in the survey, the factors of the lowest campus satisfaction are the academic stress pressure and food courts. 70% of students think their stress level is 6-8 out of 10, which is a relatively higher level. In conclusion, if the university wants to improve our campus satisfaction, we think the academic stress pressure should be alleviated, and the residences and food courts (food type and tastes) should be improved. It is worth mentioning that most students are highly satisfied with the campus facilities and security.

2 Introduction

Campus life is always an important factor in students' choice of university. There exist many different factors that might influence student satisfaction. The level of satisfaction with campus life plays a crucial role in students' mental health, academic performance, and willingness of students come to the university.

Understanding the areas that impact student satisfaction can help universities identify their strengths and areas in need of improvement. High campus satisfaction is associated with higher retention rates, higher application rates, and greater student development. In contrast, dissatisfaction leads to negative public perception and higher dropout rates. Our survey may lead universities to make improvements in making students happier. We aim to analyze the following research question:

- (RQ1) Is students' satisfaction affected by the number of days they come to campus in one week?
    - Null hypothesis: The number of days on campus per week is not associated with students' satisfaction.
    - Alternative hypothesis: The number of days on campus per week is associated with students' satisfaction.
- (RQ2) Does students' academic stress affect their satisfaction with campus life?
    - Null hypothesis: There is no association between students' academic stress and their satisfaction with campus life.
    - Alternative hypothesis: There is an association between students' academic stress and their satisfaction with campus life.
- (RQ3) Is students' satisfaction affected by the number of activities they attend per week?
    - Null hypothesis: There is no association between the number of activities attended per week and students' satisfaction.
    - Alternative hypothesis: There is an association between the number of activities attended per week and students' satisfaction.

The structure of the paper is as follows: Section 3 discusses our methods of data collection. Section 4 presents the data analysis. In section 5, we show the interpretation of our results. Section 6 covers the limitations of our study. Section 7 concludes our project, and Section 8 shows all the R code we use.

3 Methodology

From September to October 2024, in the STA304 course at the University of Toronto Mississauga, we investigated the factors that affect students' satisfaction with campus life through Piazza. We used simple random sampling, because the students in STA304 course are homogeneous, there is minimal need for stratification or clustering. Simple random sampling gives every student an equal probability of being selected, reducing bias and ensuring fairness. We ensured randomness by sending our questionnaire on Piazza to all students who decided to do projects in the STA304 course. (n = 91) students responded. There are 9 questions in total in our survey, which includes 7 multiple-choice questions and 2 short answer questions. The multiple choice questions asked about the number of days students go to campus each week, the study hours each day, their evaluation of some aspects of campus (academic pressure, campus facilities, residencies, food court, campus security, and Non-academic student support), their stress levels, the extracurricular activities they participated in, the time spent participating in extracurricular activities, and their overall satisfaction with campus life. The 2 short answer questions asked students' personal course IDs and the names of the groups they belong to.

4 Analysis

In our study, the population size is N = 200 and we plan to collect data through simple random sampling. We will do calculations to determine a sample size by using a formula based on the population mean parameter. For p and q, it is assumed that there is an equal proportion between those who are satisfied with the campus and those who are dissatisfied. Hence, given a bound of error of 0.0775, the sample size calculation will be like:

$$\frac{Npq}{(N-1)D+pq} = \frac{(200)(0.5)(0.5)}{(199)\cdot 0.0015 + (0.5)(0.5)} = 91$$

Based on the result, we sampled 91 students for our analysis.

For RQ2, we try to figure out the relationship between students' satisfaction and their academic stress by running a simple linear regression model. Before doing it, we check the assumptions, with assistance from ChatGPT.

Simple Linear Regression Assumptions

Breusch-Pagan Test Results	**Value**
BP Statistic	1.6375
Degrees of Freedom (df)	1
p-value	0.2007

Table 1: SLR assumption: Homoscedasticity check for Q2

In Table 1, we use Breusch-Pagan Test to do the homoscedasticity check. According to the table, the p-value is equal to 0.2007, which is greater than 0.05, significantly indicating homoscedasticity is satisfied.

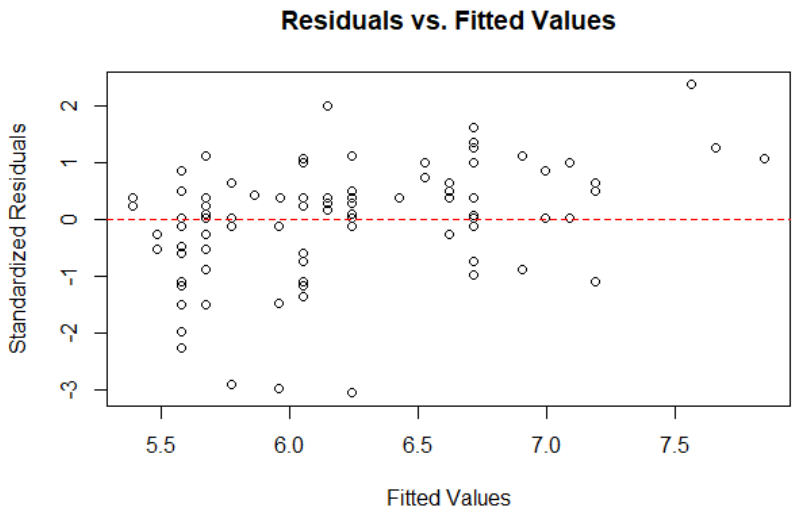

Figure 1: SLR Assumption: Linearity check for Q2

In Figure 1, there are no pattern observed in the plot of residuals versus fitted values. Therefore, the linearity assumption is satisfied.

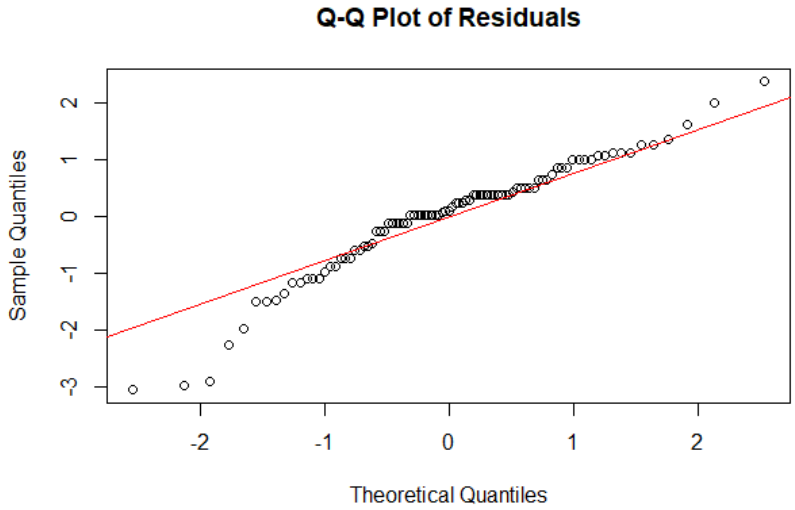

Figure 2: SLR Assumption: Normality check for Q2

In Figure 2, a Q-Q plot of residuals is used to check the normality assumption. Although there are some exceptions on the left side of the plot, most of the residuals align along the diagonal line, indicating normality assumption is met.
In addition, independence is satisfied since we use simple random sampling in this project, which ensure the independence of residuals since observations are chosen randomly.

Summary of Results for Question 2

Coefficients

Coefficients	Estimate	Std. Error	t value	Pr($>|t|$)
Intercept	4.65573	0.67296	6.918	6.74e-10
The level of stress of the participants	0.21844	0.09161	2.385	0.0192

Table 2: Table for coefficients of linear regression

Model Summary Statistics

Statistic	Value
F-statistic	5.686 on 1 and 89 DF
Degree of freedom	89
p-value	0.01922

Table 3: Model summary of linear regression

For RQ1 and RQ3, we try to find out the connection between students' satisfaction and the number of days they come to campus in one week, and the connection between students' satisfaction and the number of activities they attend weekly, respectively. Given the potential interaction effect between the number of days they come to campus per week and the number of activities they attend per week, we will use a multiple linear regression model. Similar to the previous analysis, we check the assumptions with assistance from ChatGPT.

Multiple Linear Regression Assumptions

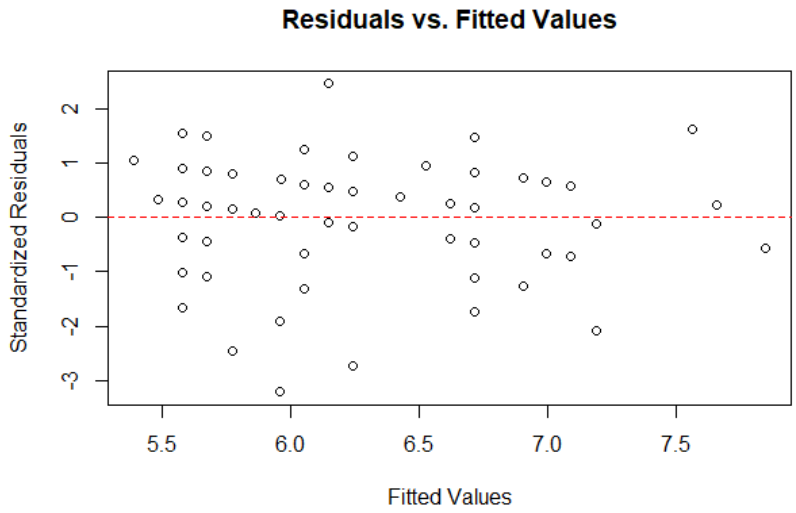

Figure 3: MLR Assumption: Linearity check for Q3 and Q1

Figure 3 presents the plot of residuals versus fitted values, showing no patterns. This indicates that the linearity assumption is met.

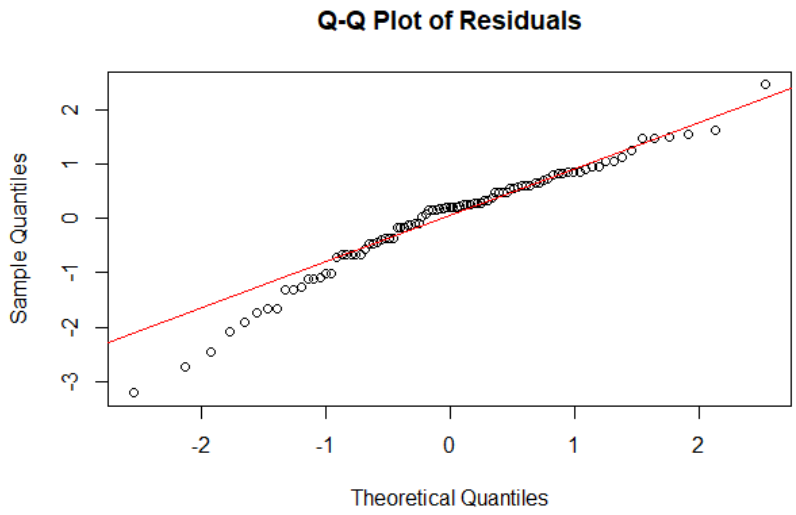

Figure 4: MLR Assumption: Normality check for Q3 and Q1

According to the Q-Q plot of residuals in figure 4, most of the residuals align along the diagonal line, indicating normality assumption is satisfied.

Breusch-Pagan Test Results	**Value**
BP Statistic	0.24416
Degrees of Freedom (df)	2
p-value	0.8851

Table 4: MLR assumption: Homoscedasticity check for Q3 and Q1

Table 4 shows the result of the Breusch-Pagan Test, where p-value equals 0.8851. Since 0.8851 is greater than 0.05, it significantly suggesting homoscedasticity assumption is satisfied.

Variance Inflation Factor (VIF) Results	**VIF**
Days in Campus/week	1.114706
Extracurricular activities/week	1.114706

Table 5: MLR Assumption: Multicollinearity check for Q3 and Q1

In Table 5, we want to check whether there are multicollinearity by calculating the Variance Inflation Factor(VIF) values for each predictor. Based on the table, both VIF are equal to 1.114706, less than 5. The results show there is no multicollinearity in the model.

Summary of Results for Question 1 and Question 3

Coefficients

Coefficients	**Estimate**	**Std. Error**	**t value**	**Pr**$(>\|t\|)$
Intercept	5.29538	0.54287	9.754	1.15e-15 ***
Days on campus/week	0.09546	0.13918	0.686	0.49461
Extracurricular activities/week	0.47156	0.15901	2.966	0.00389 **

Table 6: Table for coefficients of MLR

Model Summary Statistics

Statistic	Value
F-statistic	5.891 on 2
Degrees of Freedom (DF)	88
p-value	0.003971

Table 7: Table for the model summary of Q3 and Q1

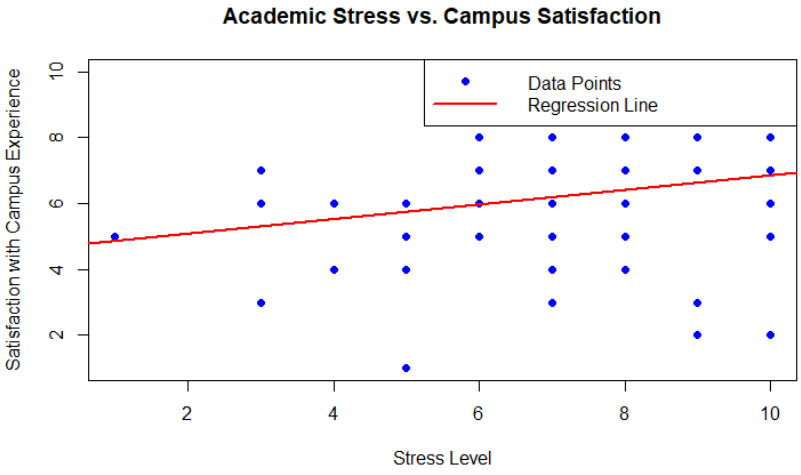

Figure 5: Graph representation of the Q2

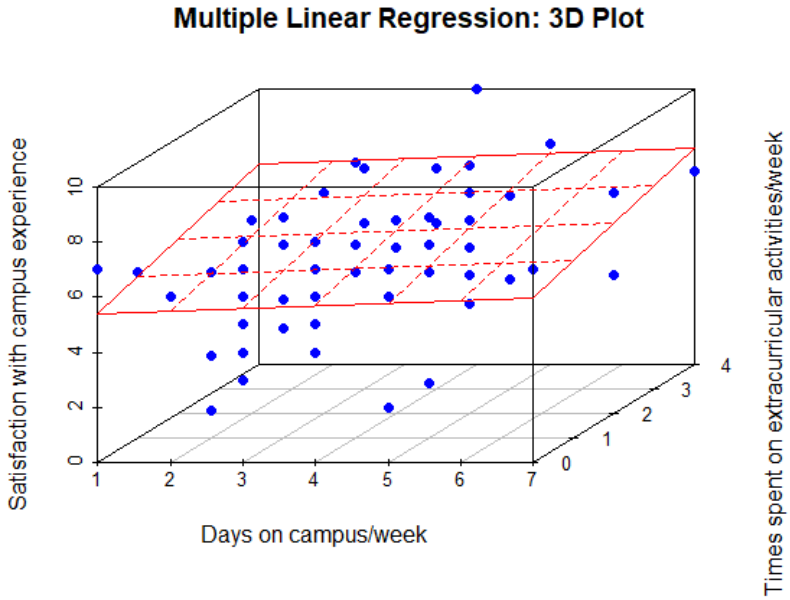

Figure 6: Graph representation of the Q3 and Q1

5 Discussion/Results

For RQ1, there is enough statistical evidence in our analysis to show that there is no relationship between the number of days students come to campus in one week and their campus satisfaction, which means the number of days students come to campus does not affect their satisfaction.

For RQ2, our analysis shows that there is strong evidence to conclude that there is an association between students' academic stress and their satisfaction with campus life. Interestingly, the graph above shows a potentially positive relationship between stress levels and student satisfaction, meaning that the higher the students' stress level, the higher their campus satisfaction.

The analysis and the graph above of RQ3 prove that there exists enough evidence that there is a positive relationship between the number of activities attended per week and students' satisfaction. This implies that the more extracurricular activities students participate in, the more satisfied they are on campus.

6 Limitations

In our study, we only surveyed students in the STA304 course, but statistics students can not represent the satisfaction of all students on campus, because students in different majors may have preferences and special needs for some campus factors due to the course and assignment characteristics of each major. In the next survey, we will need to send the questionnaire to a wider range of people who have different majors to ensure a diverse sample.

Another limitation is the authenticity of students' answers to questions. Some students may fill out our questionnaire casually just for project marks, or some students may not fill it in seriously because of their busy academics, which leads to counterintuitive results. For example, the result of our survey question 2 is that the higher the students' stress level, the higher the campus satisfaction. In the future, we can add a question that gives students answers in the title to test whether they answered the survey carefully, such as "Rate your living environment (choose 0 to prove you filled in this survey carefully)", and remove the responses of people who choose other options in this question. We can also set clearer questions and options, and add an "unsure" option to each question to prevent students from answering casually.

7 Conclusion

This project analyses the factors influencing students' satisfaction with campus life, like academic stress, extracurricular activities, and the days on campus, etc. We collect data through questionnaires and analyze data by simple linear regression and multiple linear regression. We interestingly found that there is a potential for the stress level of STA304 students to be positively correlated with students' satisfaction. There may be some limitations here, some students did not read the questionnaire carefully, or some misclicks, which caused some results that do not make sense. However, the time spent on extracurricular activities positively correlates with students' satisfaction, which seems more reasonable. We realized that we could add some incentives to encourage students to fill out the questionnaire carefully and reduce restrictions. It is worth noticing that we have strong evidence that the days on campus do not correlate with students' satisfaction. In other words, it is so important for students to have enough satisfaction on campus. By improving these factors we mentioned before, students can have a more attractive and complete campus experience. Our project also offers some advice for the university administrators.

8 Appendix

```
setwd("C:\\Users\\astry\\OneDrive\\Documents\\sta304")
library(scatterplot3d)
library(lmtest)
data  <- read.csv("sta304␣data1.csv")
names(data)

#####################################################################
```

```
# Sample Size Computations (SRS)                                    #
#####################################################################

N = 200
p = 0.5
q = 1-p
B = 0.0776

D = B^2 / 4
(N * p * q) / ((N-1) * D + p * q)

# Q2. Does students academic stress affect their satisfaction with campus
    life? Q2

model <- lm(Satisfaction_with_campus_experience ~ Stress_level, data = data)

# Model Assumptions check
# Residuals vs. Fitted Values plot q2 for Linearity
plot(model_mlr$fitted.values, rstandard(model),
     xlab = "Fitted␣Values", ylab = "Standardized␣Residuals",
     main = "Residuals␣vs.␣Fitted␣Values")
abline(h = 0, col = "red", lty = 2)

# Breusch-Pagan Test for heteroscedasticity q2
bptest(model)

# Q-Q Plot q2 for Normality
qqnorm(rstandard(model), main = "Q-Q␣Plot␣of␣Residuals")
qqline(rstandard(model), col = "red")

# Show the summary of the regression analysis
summary(model)

#show the visualization
plot(data$Stress_level, data$Satisfaction_with_campus_experience,
     main = "Academic␣Stress␣vs.␣Campus␣Satisfaction",
     xlab = "Stress␣Level",
     ylab = "Satisfaction␣with␣Campus␣Experience",
     pch = 19, col = "blue")  # ‘pch = 19‘ adds solid circle points
# Add regression line
abline(model, col = "red", lwd = 2)  # Adds a red regression line with a
    thicker widt

legend("topright", legend = c("Data␣Points", "Regression␣Line"),
       col = c("blue", "red"), pch = c(16, NA), lty = c(NA, 1), lwd = c(NA, 2))

# Q1.s students satisfaction affected by the number of days they come to
    campus in one week? and Q3. Is students satisfaction affected by the
    number of activities they attend per week?

model_mlr <- lm(data$Satisfaction_with_campus_experience ~ data$Days_on_campus_
    per_week +
                  data$Times_to_participate_in_extracurricular_activities_per_
                      week, data=data)

# Show the summary of the multiple regression analysis
summary(model_mlr)

#Multiple linear regression assumption check
# Residuals vs. Fitted Values plot for linearity
plot(model_mlr$fitted.values, rstandard(model_mlr),
     xlab = "Fitted␣Values", ylab = "Standardized␣Residuals",
     main = "Residuals␣vs.␣Fitted␣Values")
```

```
abline(h = 0, col = "red", lty = 2)


# Breusch-Pagan Test for heteroscedasticity
bptest(model_mlr)

# Q-Q Plot for normality
qqnorm(rstandard(model_mlr), main = "Q-Q␣Plot␣of␣Residuals")
qqline(rstandard(model_mlr), col = "red")


# Variance Inflation Factor (VIF) for multicollinearity
library(car)
vif(model_mlr)

# Visualization
s3d <- scatterplot3d(data$Days_on_campus_per_week,
              data$Times_to_participate_in_extracurricular_activities_per_week,
              data$Satisfaction_with_campus_experience,
              pch = 19, color = "blue",
              main = "Multiple␣Linear␣Regression:␣3D␣Plot",
              xlab = "Days␣on␣campus/week",
              ylab = "Times␣spent␣on␣extracurricular␣activities/week",
              zlab = "Satisfaction␣with␣campus␣experience")

s3d$plane3d(model_mlr, lty.box = "solid", col = "red")

# AI link
https://chatgpt.com/share/674d2b33-a210-8010-97b2-bad84aade87a
```