

Introduction

Motivation: An Analysis of The Relationship Between Distance From Campus And GPA Of Commuter Students. Since the academic performance is related to the commuting distance, I am also assuming commuting distance will affect relationships between students. And the frequency of university-related activities is a good measurement. In my assumption, the students with long commuting time are facing some degree of difficulty of coming to campus. Under that situation, they tend to decrease the frequency of coming to campus.

Research Question: Is there a relationship between commuting time to UT and university-related activity participating frequency of students in UT

Questions Asked In Survey:

What is your usual commuting time to UT Austin campus in minutes? (input "0" if living on campus)

What is your usual UT-related activity participation frequency per week? (Sports events, dinner with friends, Film /discussion, Public talk, etc.)

Data collection: I will collected data in front of Student Activity Center, Main Tower and Jester. I asked almost each person passing by whether he/she had a minute to help complete a survey instead of choosing some certain ones. This helps increase the possibility of random sampling. I removed one outlier from data. There were 99 records.

Method

Method:

First step: I create a histogram to check the value range of output variable “Frequency”, the distribution tends to be right skewed. So the median, quartile 1, quartile 3 and IQR values are calculated.

Second step: I create a scatterplot and get the regression function in Excel since I need to find the potential linear relationship between the two variables. I removed one outlier (8, 23) from the data.

Assumptions/Conditions:

To check whether the assumptions were violated, I generated a residual plots graph in Excel. After checking the residual plots. There were no clear trend that violates the linearity assumptions. And were no obvious outliers violates then no outlier assumption.

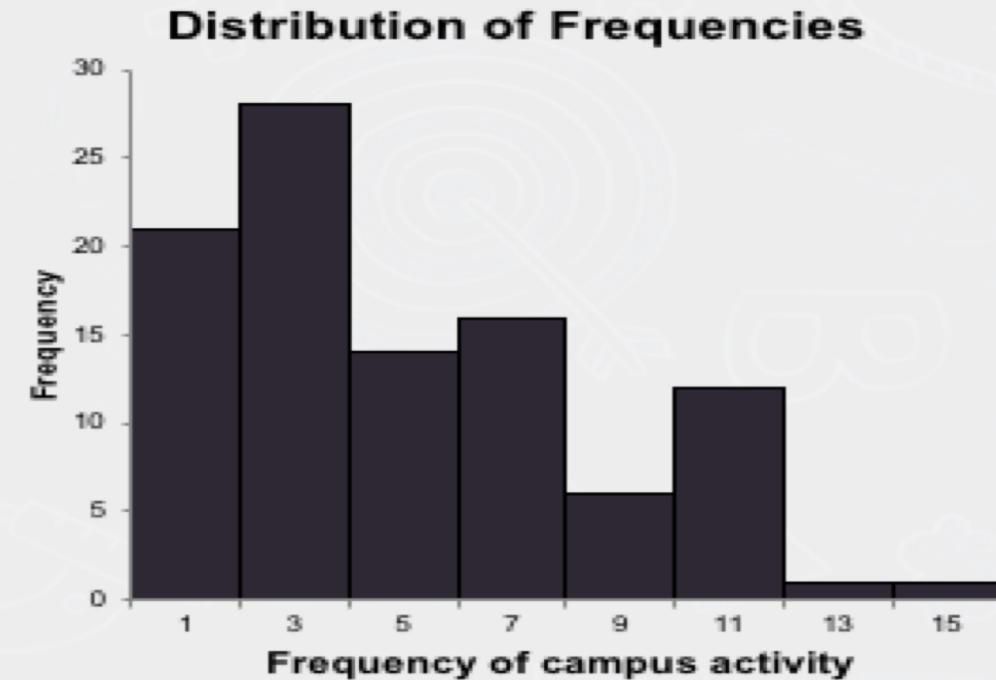
Dependent Variable

Dependent Variable: Frequency of campus activity

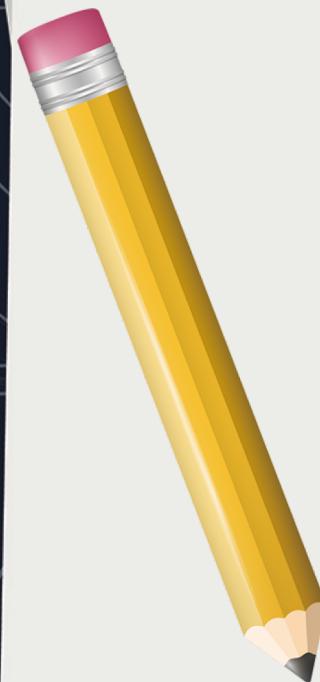
Since the values to be right skewed and unimodal,
I calculated descriptive statistics below:

| | |
|---------------|----|
| Min Frequency | 0 |
| Max Frequency | 15 |
| Count | 99 |
| Median | 4 |
| Quartile 1 | 2 |
| Quartile 3 | 7 |
| IQR | 5 |

Histogram



SDS302 Final Poster Project



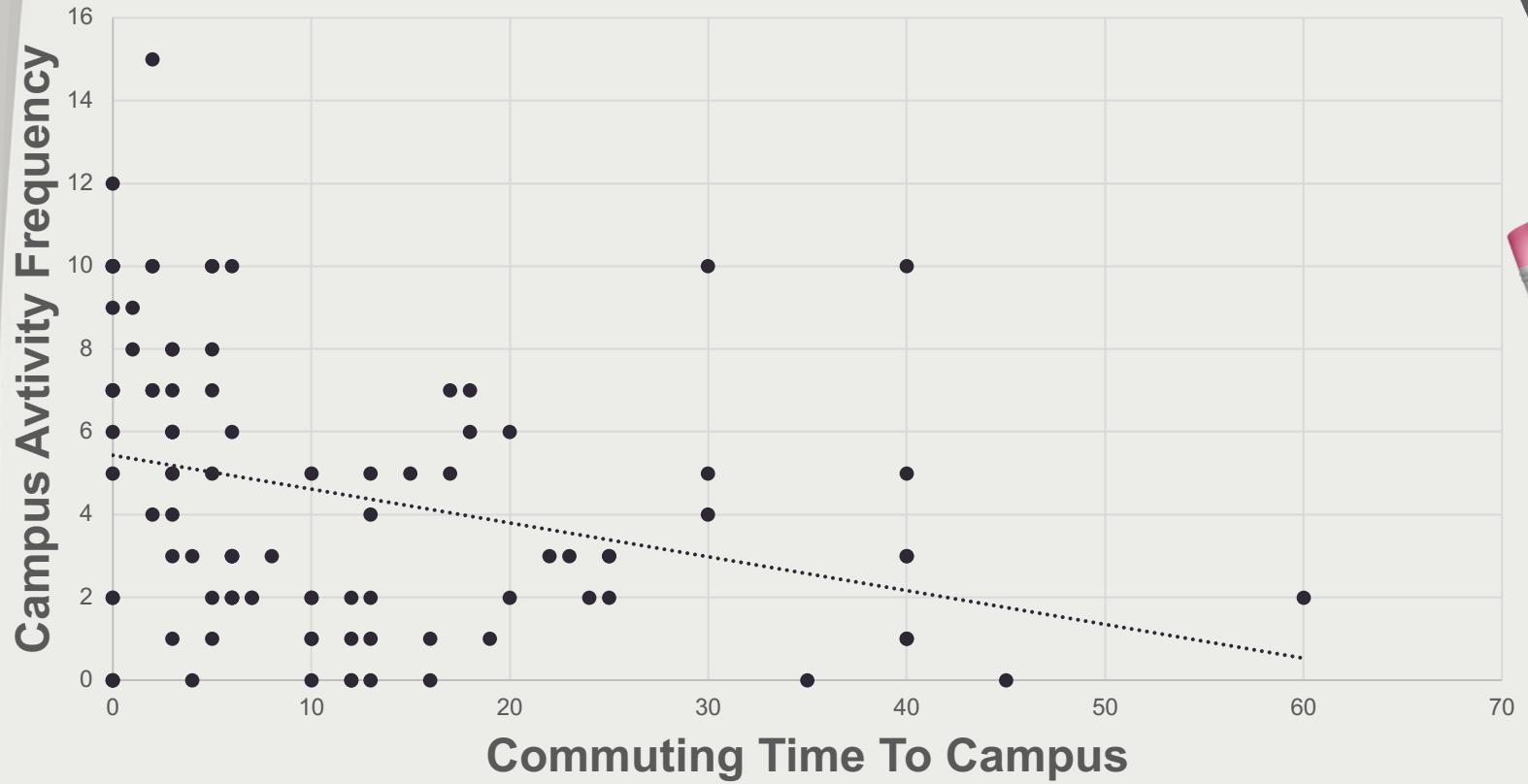
Date of Presentation: 5/8/2018

Student Name: Bulbasaur

Lab meeting time: Thursday 5pm

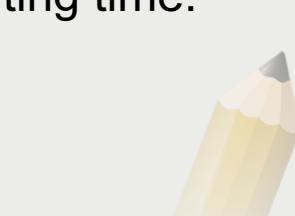
Results

Frequency VS Cummuting Time



$R^2: 0.09$

r (correlation): -0.30



Residual Plot

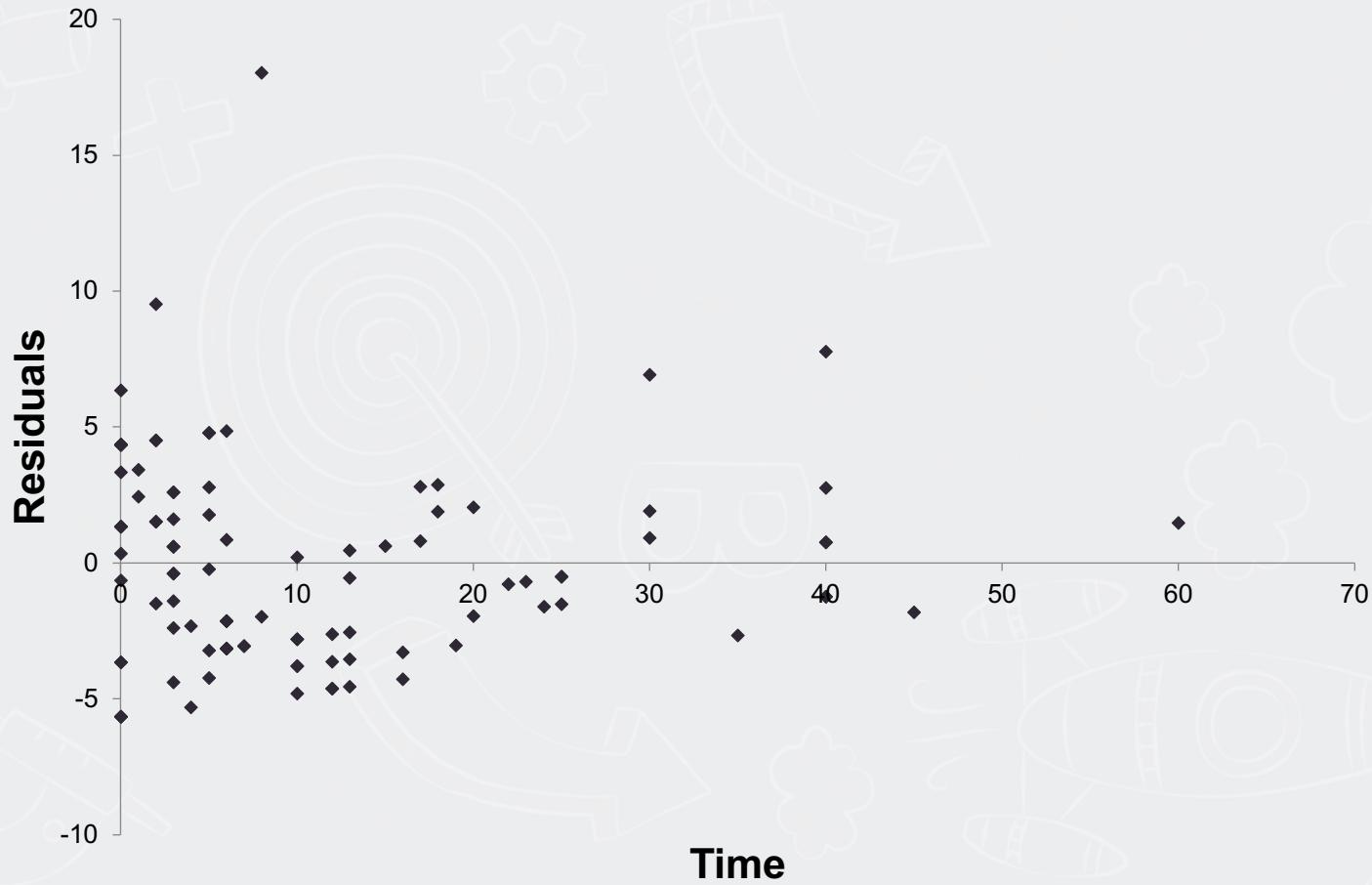
Issue 01

As I mentioned before, the data point at top left is an obvious outlier. I removed this record before linear function calculating.

Issue 02

There are more points on the side of shorter commuting time. But considering this is the case in real that more students live in the area closer to campus, this shouldn't be taken as an data issue.

Time Residual Plot



Conclusion

Equation: $y = -0.08x + 5.43$

r (correlation): -0.30

R²: 0.09

This linear equation indicates that as the commuting time increases by 1 minute, the predicted Campus Activity Frequency decreases by -.08.

-0.08

Additionally, a person living on campus(0 minute commuting time) is predicted to have a Campus Activity Frequency of 5.43

5.43

The R² value of .0904 indicates that 9.04% of the variation in Campus Activity Frequency is accounted for by the commuting time

0.09

The correlation value of -0.30 indicates that there is negative relation between the two variables. While the absolute value of 0.30 indicates that the relation is weak .

-0.30

Limitations

Potential bias: most of my survey results were collected during afternoon. There might be some students show up on campus only in morning or at night.

Seems that the correlation between the two variables is not as large as I expected.

Sample size: I collected 100 data, but this is still a relatively small sample.

My second question is pretty limited, since students might have potential campus activities than I included in the question.

Implications



- 1. What is an implication of your study? What recommendations do you have for people (doctors, policy makers, students, nutritionists, etc) based on your study.**

One implication of my study is that students have shorter commuting time to campus tend to attend on campus activities more often. One suggestion to on campus activity organizers: advertise more around dorms and communities close to campus! You will attract more students attend your activity!

- 2. What is a follow-up study that you would conduct to try and better understand the results of your current study. Would you include new variables? Would you ask a different population? Would you redo your same study with changes?**

Seems commuting time is not the only factor affecting the on campus activity frequency, I can redo study by including more variables like credit hours taking this semester to generate a better prediction model with higher R^2 value. The same population will be fine.