Technical Report: Academic Advising Role in Student Success

Group 7 LEC0102 | November 20th, 2023

Meghna Ghosh, Luvleen Rehal, Sarada Sai Turaga, Mohammed Yusuf Shaikh, Lucas Lyu, Gurvir Sasan, Shahan Uppal

Introduction

This study explores the relationship between academic advising and its use to support student success. Academic advisors offer guidance on understanding degree requirements, course selections, career paths, and personal development, aiming to assist students in navigating their academic journey. We hope to answer the following research questions in our study:

1. How does frequent utilization of academic advising services relate to academic performance?
2. How does students' rate of effective communication with academic advisors affect the student's improvement in their academic performance?
3. How does the utilization of academic advising increase the probability of students graduating on time?

Note: We use the term "Higher Seeking" in our report to label students who chose an option greater than or equal to three, and we use the term "Lower Seeking" in our report to label students who chose an option less than three for our survey question: "On a scale of 1 – 5, how likely are you to seek academic advising"?

Hypothesis 1 - Two-Sample T-Test

The following hypothesis will model the relationship between the utilization of academic services and academic performance:

Let $\mu 1$ be the population mean CGPA of students categorized as "Higher Seeking" for academic advising.

Let $\mu 2$ be the population mean CGPA of students categorized as "Lower Seeking" for academic advising.

(Null hypothesis): H_0: $\mu 1 - \mu 2 = 0$

(Alternative Hypothesis): H_A: $\mu 1 - \mu 2 \neq 0$

Hypothesis 2 – Linear Regression Test

The relationship between students' rate of effective communication with academic advisors and how students noticed improvement in their academic performance, it will be modelled by Y = β_0

$+ \beta_1 X_1 + \epsilon$, where Y represents the communication effectiveness and X_1 represents performance effectiveness.

(Null hypothesis): H_0: $\beta_1 = 0$

(Alternative Hypothesis): H_A: $\beta_1 \neq 0$

Hypothesis 3 - Linear Regression Test

The relationship between the use of academic advising services and on-time graduation is modelled by $Y = \beta_0 + \beta_2 X_2 + \epsilon$, where Y represents the probability of on-time graduation and X_2 represents the utilization of academic services.

(Null hypothesis): H_0: $\beta_2 = 0$

(Alternative Hypothesis): H_A: $\beta_2 \neq 0$

Methodology

From the beginning of October 2023 to early November 2023, we conducted a survey to gather data to understand the relationship between academic advising and its effects on STA304H5 students. The sampling method of choice was Simple Random Sampling (SRS), and to ensure randomness, the survey was presented on the class forum piazza. This provided equal accessibility to every student, omitting lecture timings and availability as possible bias factors. Furthermore, we contacted students randomly from both lectures, encouraging them to fill out our survey to account for the students not being active on piazza. The sample size that we picked was 50, and we received a total of 48 responses. We presented ten questions regarding students' academic performance, their rating and usage of academic advisors, and other miscellaneous demographic-related questions (a program of study, CGPA, and student performance).

Analysis

Sample Size

The total number of students enrolled is 200 (N = 200) with a bound of error of estimation B = 0.127. Since p and q are unknown, it can be assumed that $p = q = \frac{1}{2}$. The sample size computation follows:

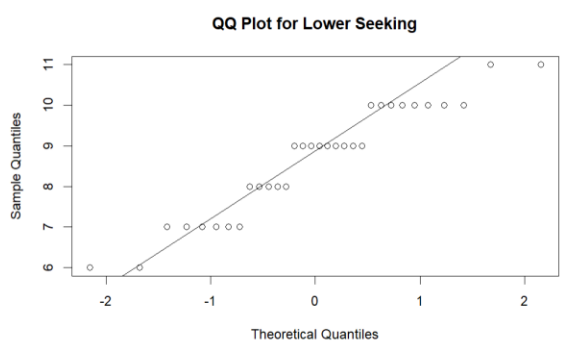
$$n = \frac{Npq}{(N-1)D + pq} = \frac{200 \times 0.5 \times 0.5}{(200-1)\frac{(0.127)^2}{4} + 0.5 \times 0.5} = 47.50965099 \approx 48, \text{ where } D = \frac{B^2}{4} = 0.0025$$

Therefore, we need to sample 48 students, which we have completed.

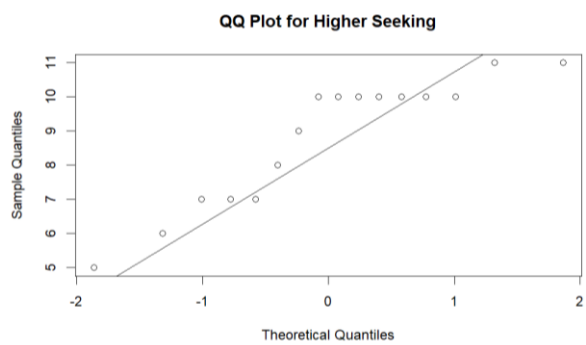
Assumptions

1. The samples are independent and obtained randomly. *{Hypothesis 1, 2, 3}*
2. The variances of the two independent groups are equal. *{Hypothesis 1, 2, 3}*
3. The samples are normally distributed, which is crucial when the sample sizes are small. *{Hypothesis 1, 2, 3}*

For **Hypothesis 1**, the first and second assumptions are met because we use simple random sampling without replacement. Also, CGPA values are not dependent on the amount of academic advising someone receives, proving independence. The third assumption is proved in Figure 1.1 and Figure 1.2. Since the points demonstrate a straight line, it follows that the data is normally distributed.

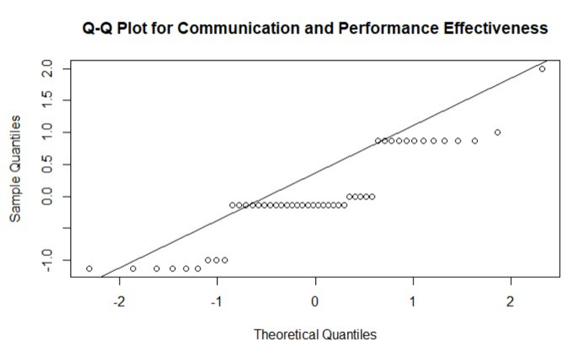


[Figure 1.1]



[Figure 1.2]

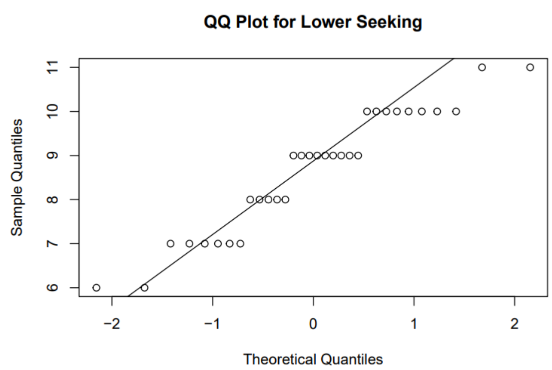
For **Hypothesis 2**, the first and third assumptions are met because we use simple random sampling without replacement. Also, communication effectiveness with academic advisors is not dependent on the range of the student improvement in performance, i.e., proves independence. The third assumption is proved in Figure 1.3. Since the points demonstrate a straight line, it follows that the data is normally distributed.



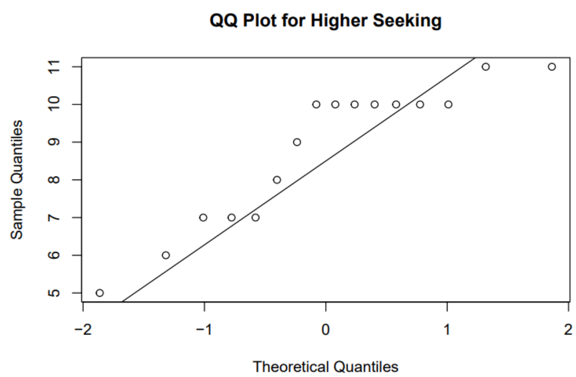
[Figure 1.3]

For **Hypothesis 3**, the first and second assumptions are met because we use simple random sampling without replacement. Also, the degree completion time longer than the standard duration is not dependent on academic advising services. The third assumption is proved in Figure 1.4 and

Figure 1.5 since the points demonstrate a straight line, it follows that the data is normally distributed.



[Figure 1.4]



[Figure 1.5]

Factor analysis is the linear correlation between student performance and academic advising factors. There are certain assumptions that must be met:

1. Data can be factored.
2. No significant outliers.
3. Appropriate sample size.
4. Variable errors are uncorrelated.

Relevant Outputs

Proportion for Seeking Advising	Proportion for Meeting Frequency	Proportion for Rate of Graduation	Mean CGPA with higher seeking academic advising	Mean CGPA with lower seeking academic advising	Mean time for graduation with higher seeking advising	Mean time for graduation with lower seeking advising
0.6875	0.4792	0.3542	8.6061	8.9333	1.1515	0.8667

Hypothesis 1:

Two-Sample t-test

```
	Welch Two Sample t-test

data:  dataset$Cgpa_range by dataset$Advising_Group
t = 0.29625, df = 23.469, p-value = 0.7696
alternative hypothesis: true difference in means between group Higher Seeking and group Lower Seeking is not equal to 0
95 percent confidence interval:
 -0.9336244  1.2461244
sample estimates:
mean in group Higher Seeking  mean in group Lower Seeking 
                     8.81250                      8.65625 
```

Hypothesis 2:

Linear Regression Model

```
lm(formula = Project_Dataset$Academic_advisor_communication_effectiveness ~ 
    Project_Dataset$Performance_Effectiveness, data = Project_Dataset)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.1316 -0.1316 -0.1316  0.8684  2.0000 

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                 3.1316     0.1180  26.537  < 2e-16 ***
Project_Dataset$Performance_Effectiveness  -1.1316     0.2585  -4.377 6.87e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7274 on 46 degrees of freedom
Multiple R-squared:  0.294,	Adjusted R-squared:  0.2787 
F-statistic: 19.16 on 1 and 46 DF,  p-value: 6.869e-05
```

ANOVA Test

```
                         Df Sum Sq Mean Sq F value   Pr(>F)    
Project_Dataset$PE_values  1  10.14  10.137   19.16 6.87e-05 ***
Residuals                 46  24.34   0.529                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hypothesis 3:

Simple Linear Regression Test
“Higher seeking” academic advising

```
Call:
lm(formula = Degree_Completion_Extension ~ Seeking_Academic_Advising, 
    data = subset(dataset, Seeking_Academic_Advising >= 3))

Residuals:
     Min       1Q   Median       3Q      Max 
-1.30534 -1.05153 -0.05153  0.94847  2.20229 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)  
(Intercept)                 2.0668     1.1130   1.857   0.0728 .
Seeking_Academic_Advising  -0.2538     0.3031  -0.837   0.4088  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.208 on 31 degrees of freedom
Multiple R-squared:  0.02212,	Adjusted R-squared:  -0.009423 
F-statistic: 0.7013 on 1 and 31 DF,  p-value: 0.4088
```

“Lower seeking” academic advising

```
Call:
lm(formula = Degree_Completion_Extension ~ Seeking_Academic_Advising,
    data = subset(dataset, Seeking_Academic_Advising < 3))

Residuals:
    Min      1Q  Median      3Q     Max
-1.5714 -0.4107 -0.2500  0.7500  1.4286

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                 2.8929     0.8414   3.438  0.00441 **
Seeking_Academic_Advising  -1.3214     0.5218  -2.532  0.02501 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

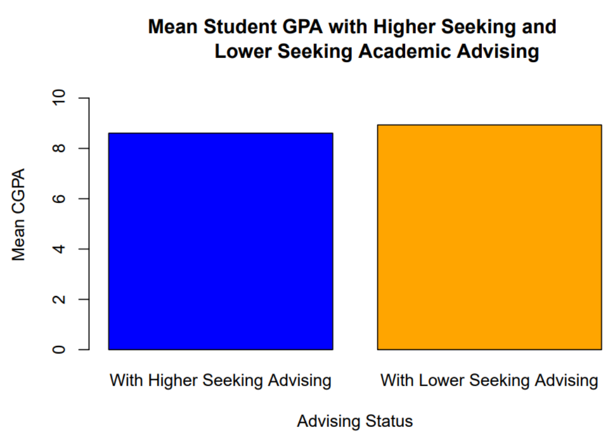
Residual standard error: 1.008 on 13 degrees of freedom
Multiple R-squared:  0.3304,    Adjusted R-squared:  0.2788
F-statistic: 6.413 on 1 and 13 DF,  p-value: 0.02501
```

Advanced Methodology:

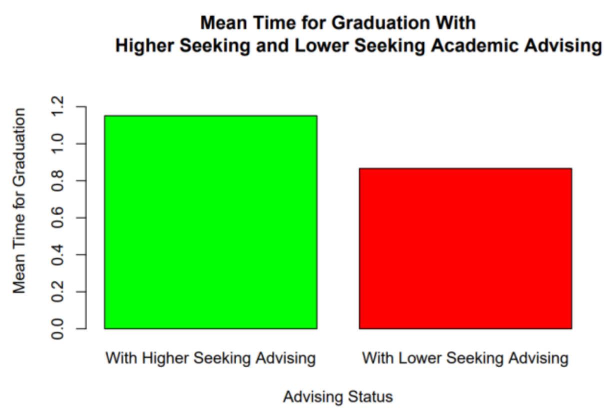
Factor Analysis

```
##
## Loadings:
##                                            RC1    RC2    RC3
## Cgpa_range                                 0.176  0.818
## Program_Satisfaction                       0.547  0.285
## Degree_Completion_Extension                0.159 -0.832
## Seeking_Academic_Advising                  0.132  0.135  0.802
## Advisor_availability_rating               -0.860
## Meeting_Frequency                                -0.202  0.798
## Performance_Effectiveness                  0.641 -0.146  0.444
## Academic_advisor_communication_effectiveness -0.846  0.137 -0.236
##
##                  RC1   RC2   RC3
## SS loadings    2.245 1.542 1.537
## Proportion Var 0.281 0.193 0.192
## Cumulative Var 0.281 0.473 0.666
```

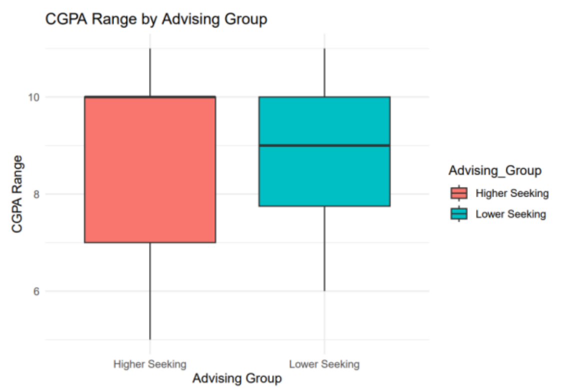
Graphs



[Figure 1.6]: Barplot to measure the GPA among students who are more likely to seek academic advising versus students who are less likely to seek academic advising.

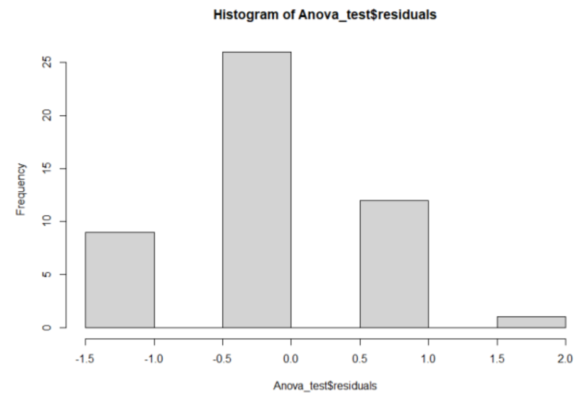


[Figure 1.7]: Barplot to measure the estimated time of degree completion longer than the standard duration (4 years) among students who are more likely to seek academic advising versus students who are less likely to seek academic advising.



[Figure 1.8]: Boxplot to measure the CGPA of students who are more likely to seek academic advising versus students who are less likely to receive academic advising.

[Figure 2.0] The line best fit for regression model from Hypothesis 2.



[Figure 1.9]: Histogram is constructed to show the residuals from ANOVA test. It shows a very slight skewness towards the right between -0.5 & 0. It can cause a small bias.

Discussions/Results

****Disclaimer: For brevity purposes, we will provide the interpretation of our first hypothesis in detail and provide the remaining tests in a summarized table.**

In Figure 1.6, we observe that the average CGPA for the higher seeking group doesn't significantly differ from the lower seeking group. Both groups exhibit an average CGPA within the 2.7 – 3.0 GPA range, 8.6060 for the higher seeking group and 8.9333 for the lower seeking group.

Moving to Figure 1.7, we note that the higher seeking group take, on average, 1.15 additional years beyond the standard 4-year duration to complete their degree. In contrast, the less advising-seeking group averages 0.87 additional years.

<u>Results from Hypothesis #1:</u>

The correlation coefficient of -0.03 tells us that there is a weak and negative relationship between the students who are seeking academic advising and their CGPA range.

We note a small t-stat of 0.29625 and p-value of 0.7696. Considering that the p-value is greater than 0.05, this means that the difference between the means (the average CGPA range between the higher seeking and lower seeking groups) is not statistically significant and we don't have enough evidence to reject the null hypothesis. Figure 1.8 illustrates this point as we see very little difference in the medians between the higher seeking group and the lower seeking group.

Our 95% confidence interval of (-0.9336244, 1.2461244) also contains 0, which further suggest that there exists no significance for difference in the average CGPA range between the higher seeking and lower seeking groups.

Figure 1.8 highlights a *higher variance* for 'Higher Seeking Group' than 'Lower Seeking Group'.

Summary for Hypothesis #2:

Measure	Value	Specified Conclusion
C.C (Correlation Coefficient)	-0.5422	This shows the two variables have moderate negative correlation between Academic Advisor Communication Effectiveness and Performance effectiveness.
Approximate ANOVA F – Stat	19.16	Strong relationship between Academic Advising Communication Effectiveness and Performance Effectiveness.
Approximate ANOVA P-Value	0.0000687	Strong relationship between Academic Advisor Communication Effectiveness and Performance Effectiveness.

Linear Regression P – Value	0.00006869	Since < 0.05; the model is statistically significant. The model is a good fit.
Linear Regression Coefficient	3.1316	A 3.1316 increase in Performance Effectiveness per unit change in Academic Advising Communication Effectiveness.
Approximate Adjusted R – Squared	0.2787	Tells us Adjusted R-squared = 0.2787 show the model is a good fit for simple linear regression

ANOVA (Analysis of Variance) test is performed because of the two categorical values that have been created under the new variable 'PE_value' (Yes and no). The test is done under linear regression hypothesis test. The aim of the hypothesis is to check if there is a difference in the average rate of effectiveness in the communication between student and academic advisor across the group of response (Yes/No) provided by students who noticed improvement in their academic performance and those who did not.

We found MSE to be 0.5291762 which corroborates our assumption for ANOVA test that the residuals come from a normal distribution, as MSE is close to 0. The Smaller the MSE, the regression model is a better fit. According to the Histogram, there is a slight skewness, between -0.5 and 0, which can be interpreted that if the sample size increases, the error (MSE) can be reduced, and our regression model would be a better fit.

<u>Summary for Hypothesis #3:</u>

	Approximate Adjusted R – Squared	**F - Stat**	**P - Value**	**Interpretation of Model**	**Specified Conclusion**

Higher Seeking Group	-0.009	0.7013	0.4088	Model is likely not significant in predicting degree completion extension.	No clear relationship between students who seek academic advising and degree completion extension.
Lower Seeking Group	0.2800	6.4130	0.0250	Model is likely significant in predicting degree completion extension.	Clear relationship between students who do not seek academic advising and degree completion extension.

<u>Insight for the Factor Analysis:</u>

By looking at communalities output, we notice that those three factors together explained around 66.6% of the total variance, with factor 1 explaining around 28.1%, factor 2 explaining around 19.3% and factor 3 explaining around 19.2%. Factors 1, 2 and 3 have respective Eigenvalues of 2.245, 1.542 and 1.537. According to Kaiser's rule, as all three factors have Eigenvalues > 1, we deem those three factors as valid.

As for the Rotated factor matrix output, we use the threshold of +/- 0.5 to deem a loading valid. Here we see that "Program Satisfaction", "Advisor availability rating", "Performance Effectiveness" and "Academic advisor communication effectiveness" are variables that load onto factor 1. Factor 1 has a strong negative correlation with "Advisor availability rating" and "Academic advisor communication effectiveness", indicating a common factor relating to being unable to find the advisor and finding the conversation with the advisor useless. On the contrary, the positive correlation of factor 1 with "Program Satisfaction" and "Performance Effectiveness" indicates the other extreme where students are both satisfied with their current program and deeming advisory effective aligns with this factor. Overall, factor 1 seems to represent a spectrum of students' attitudes toward academic advising and past experiences with academic advising.

"Cgpa_range" loaded onto factor 2 with a high positive correlation while "Degree Completion Extension" loaded onto factor 2 with a high negative correlation. This indicates that factor 2 aligns with the group of students who have strong academic performance and are expecting to graduate within the standard 4-year study.

"Seeking Academic Advising" and "Meeting Frequency" both loaded onto factor 3 with high positive correlation. This makes sense as these two factors are highly correlated. This indicates that factor 3 may represent the group of students that very actively seek academic advising and continue to consistently interact with academic advisors.

Conclusions and Limitations

Although insightful connections were made throughout the study, it is important to note the constraints to accurately represent the interpretations that were made. There are three primary constraints: (1) small sample size, (2) self-reported bias, and (3) causality vs. correlation.

Regarding the first limitation, our sample size was relatively small. This may mean the results do not fully capture the behaviour of the student population. Hypothesis 2 results illustrates this, where the MSE is quite low, and the histogram is slightly right skewed. Ideally with a larger sample size, biased results are less likely to occur. Additionally, the data collected for the survey focused on self-reported results, which assumes response bias and inaccuracies, delineating the discrepancies that exist amongst the variances in hypothesis 1. Lastly, it is important to note the difference between correlation and causality, and have other factors that can influence student success, such as individual motivation, socioeconomic background, and external life events.

In conclusion, the purpose of this study was to investigate the connection between academic advising and student success. Throughout this process, we came across many valuable insights that highlighted the interconnectivity between students' engagement with advising services, their contentment with the program, and their timely degree completion. It is important to remember the limiting factors of this study, and the widespread generalizations that it incurs.

Appendix

```
---
output:
  pdf_document: default
  html_document: default
```

```
---
---
title: "Analysis"
author: "Group 7: Unusual Deviations"
date: '2023-11-20'
output: pdf_document
—

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

```{r Load Dataset}
library(readxl)
library(psych)
# File path
excel_file <-
"C:/Users/luvle/OneDrive/Documents/STA304/Dataset_Sta304.xlsx"
# Read data
dataset <- read_excel(excel_file)
```

----- PROPORTIONS -----
Seeking Academic Advising
```{r Proportion of Students Seeking Academic Advising}
seeking_advising <- sum(dataset$Seeking_Academic_Advising >= 3) /
nrow(dataset)
cat("seeking advising proportion:", seeking_advising, "\n")
# Used B as 0.1272
n <- ((200)*(seeking_advising)*(1-seeking_advising))/
  ((200-1)*(((0.1272)^2)/4)+ (seeking_advising) * (1-
seeking_advising))
cat ("Sample size:", ceiling(n))
```

Meeting Frequency
```{r Proportion of Students Meeting with their Academic Advisor}
meeting_frequency <- sum(dataset$Meeting_Frequency >= 2) /
nrow(dataset)
cat("meeting frequency proportion:", meeting_frequency, "\n")
# Used B as 0.1272
n <- ((200)*(meeting_frequency)*(1-meeting_frequency))/
  ((200-1)*((0.1272)^2/4)+ (meeting_frequency) * (1-
meeting_frequency))
cat ("Sample size:", ceiling(n))
```
```

````
```

# Rate of Graduation
```{r Proportion of Students based on Rate of Graduation}
rate_of_graduation <- sum(dataset$Meeting_Frequency >= 2 &
                            dataset$Degree_Completion_Extension <= 2)
/ nrow(dataset)
cat("rate of graduation proportion:", rate_of_graduation, "\n")
# Used B as 0.1272
n <- ((200)*(rate_of_graduation)*(1-rate_of_graduation))/
  ((200-1)*((0.1272)^2/4)+ (rate_of_graduation) * (1-
rate_of_graduation))
cat ("Sample size:", ceiling(n))
```

# --- MEAN ----
# Mean Student GPA and Academic Advising
```{r Mean Student GPA and Academic Advising}
# Create Data Frame
data_cgpa_advising <- dataset[, c("Cgpa_range",
"Seeking_Academic_Advising")]
# Mean CGPA and seeking academic advising
mean_cgpa_with_advising <-
mean(dataset$Cgpa_range[dataset$Seeking_Academic_Advising >= 3])
mean_cgpa_without_advising <-
mean(dataset$Cgpa_range[dataset$Seeking_Academic_Advising < 3])
# Comparing the two mean CGPAs
# higher seeking is bigger than or equal to three
# lower seeking is less than three
cat("Mean CGPA with higher seeking academic advising:",
mean_cgpa_with_advising, "\n")
cat("Mean CGPA with lower seeking academic advising:",
mean_cgpa_without_advising, "\n")
```

# Plot for Mean Student GPA and Academic Advising
```{r Plot for Mean Student GPA and Academic Advising}
mean_cgpa_data <- data.frame(
  Advising_Status = c("With Higher Seeking Advising", "With Lower
Seeking Advising"),
  Mean_CGPA = c(mean_cgpa_with_advising, mean_cgpa_without_advising)
)
# Create a bar plot
barplot(mean_cgpa_data$Mean_CGPA,
        names.arg = mean_cgpa_data$Advising_Status,
````

````
        main = "Mean Student GPA with Higher Seeking and
        Lower Seeking Academic Advising",
        xlab = "Advising Status",
        ylab = "Mean CGPA",
        col = c("blue", "orange"),
        ylim = c(0, max(mean_cgpa_data$Mean_CGPA) * 1.2),
        beside = TRUE
)
```

# Mean Rate of Graduation and Academic Advising
```{r Mean Rate of Graduation and Academic Advising}
# Creating a data frame
data_degree_completion_advising <-
  data.frame(Degree_Completion_Extension =
dataset$Degree_Completion_Extension,
             Seeking_Academic_Advising =
dataset$Seeking_Academic_Advising)
# Calculating the mean time for students to graduate with academic
advising support
mean_time_with_advising <-
  mean(data_degree_completion_advising$Degree_Completion_Extension
       [data_degree_completion_advising$Seeking_Academic_Advising >=
3])
# Calculating the mean time for students to graduate without academic
advising support
mean_time_without_advising <- mean(data_degree_completion_advising$

Degree_Completion_Extension[data_degree_completion_advising$

Seeking_Academic_Advising < 3])
# Comparing the two mean times for graduation
cat("Mean time for graduation with higher seeking advising :",
    mean_time_with_advising, "\n")
cat("Mean time for graduation with lower seeking advising:",
    mean_time_without_advising, "\n")
```

# Plot for Mean Rate of Graduation and Academic Advising
```{r Plot for Mean Rate of Graduation and Academic Advising}
# Create data frame for mean time for graduation
mean_time_data <- data.frame(
  Advising_Status = c("With Higher Seeking Advising", "With Lower
Seeking Advising"),
  Mean_Time = c(mean_time_with_advising, mean_time_without_advising)
````

~~~
)
# Create a grouped bar plot
barplot(mean_time_data$Mean_Time,
        names.arg = mean_time_data$Advising_Status,
        main = "Mean Time for Graduation With
        Higher Seeking and Lower Seeking Academic Advising",
        xlab = "Advising Status",
        ylab = "Mean Time for Graduation",
        col = c("green", "red"),
        ylim = c(0, max(mean_time_data$Mean_Time) * 1.2),
        beside = TRUE
)
```
# --- Hypothesis ----

# Hypothesis 1: Two-Sample Test Between Seeking Academic Advising and
CGPA
```{r Hypothesis 1 Two-sample test}
cor(dataset$Seeking_Academic_Advising, dataset$Cgpa_range)
dataset$Advising_Group <- ifelse(dataset$Seeking_Academic_Advising <=
3, "Lower Seeking", "Higher Seeking")

# Perform the t-test
t_test_result <- t.test(dataset$Cgpa_range ~ dataset$Advising_Group,
data = dataset)
# Display the results
print(t_test_result)

# Q-Q plot for Lower Seeking group
qqnorm(dataset$Cgpa_range[dataset$Advising_Group == "Lower Seeking"],
main = "QQ Plot for Lower Seeking")
qqline(dataset$Cgpa_range[dataset$Advising_Group == "Lower Seeking"])

# Q-Q plot for Higher Seeking group
qqnorm(dataset$Cgpa_range[dataset$Advising_Group == "Higher Seeking"],
main = "QQ Plot for Higher Seeking" )
qqline(dataset$Cgpa_range[dataset$Advising_Group == "Higher Seeking"])

#Boxplot for Higher seeking and Lower Seeking (It shows Higher seeking
group shows higher seeking group variance compared to lower seeking
group)
boxplot(dataset$Cgpa_range ~ dataset$Advising_Group, color = 'group')
```

# Plot for hypothesis 1
~~~

````
```{r Plot for hypothesis 1}
library(ggplot2)
# Create a box plot
ggplot(dataset, aes(x = Advising_Group, y = Cgpa_range, fill =
Advising_Group)) +
  geom_boxplot() +
  labs(
    title = "CGPA Range by Advising Group",
    x = "Advising Group",
    y = "CGPA Range"
  ) +
  theme_minimal()
```

# Hypothesis 2: Linear Regression Test between Communication
Effectiveness and Performance Effectiveness
```{r Hypothesis 2 Linear Regression}
reg_model <- lm(dataset$Academic_advisor_communication_effectiveness ~
dataset$Performance_Effectiveness, data = dataset)
summary(reg_model)
# Plotting the function of Regression Model
plot(dataset$Academic_advisor_communication_effectiveness,
dataset$Performance_Effectiveness)
abline(reg_model)

# Getting MSE with n=48
sum(residuals(reg_model)^2)/(48-2)

# For ANOVA Test the value are converted from 0 and 1 to 'No' and
'Yes' respectively
# So we have 2 categorical groups for ANOVA test (Group of 'Yes' and
'No')
dataset$PE_values <- ifelse(dataset$Performance_Effectiveness == 1,
"Yes", "No")

# Correlation of 2 Variables
cor(dataset$Academic_advisor_communication_effectiveness,
dataset$Performance_Effectiveness) # It has negative variable;
Moderate

# Conducting the ANOVA test
Anova_test <- aov(dataset$Academic_advisor_communication_effectiveness
~ dataset$PE_values, data = dataset)
summary(Anova_test) # p-value = 0.0000687
````
````

~~~~
# Getting Histogram for residuals
hist(Anova_test$residuals)
```

```{r Q-Q Plot for Communication and Performance}
residuals_anova <- residuals(Anova_test)
qqnorm(residuals_anova, main = "Q-Q Plot for Communication and
Performance Effectiveness")
qqline(residuals_anova)
```

# Hypothesis 3: Simple Linear Regression Between Graduation and
Academic Advising
```{r Hypothesis 3 Simple Linear Regression}
#HO: Academic advising has no effect on the rate of graduation
#HA: Academic advising does have an effect on the rate of graduation
# Fit a linear regression model for degree completion with academic
advising >= 3 as the predictor
model_with_advising_geq_3 <- lm(Degree_Completion_Extension ~
Seeking_Academic_Advising,
                                 data = subset(dataset,
Seeking_Academic_Advising >= 3))

# Fit a linear regression model for degree completion with academic
advising < 3 as the predictor
model_with_advising_lt_3 <- lm(Degree_Completion_Extension ~
Seeking_Academic_Advising,
                                data = subset(dataset,
Seeking_Academic_Advising < 3))

# Print the summary of the models
summary(model_with_advising_geq_3)
summary(model_with_advising_lt_3)

# Plot the linear regression models
plot(dataset$Seeking_Academic_Advising,
dataset$Degree_Completion_Extension,
     main = "Linear Regression Model", xlab = "Seeking Academic
Advising",
     ylab = "Degree Completion Extension")
abline(model_with_advising_geq_3, col = "red")
abline(model_with_advising_lt_3, col = "blue")

residuals <- residuals(model_with_advising_geq_3)
~~~~

````
# Create QQ plot
qqnorm(residuals, main = "QQ Plot for Higher Seeking")
qqline(residuals)

residuals1 <- residuals(model_with_advising_lt_3)

# Create QQ plot
qqnorm(residuals1, main = "QQ Plot for Lower Seeking")
qqline(residuals1)
```

# --- Factor Analysis ----
```{r Factor Analysis}
#Data Preparation, need to remove non-numeric variables
numeric_data <- dataset[, sapply(dataset, is.numeric)]
factor_result <- principal(numeric_data, nfactors = 3, rotate =
"varimax")
print(factor_result$loadings)
print(factor_result$communalities)
```
````
````