# 1.Introduction:

SUMMARY

In the dynamic landscape of contemporary education, the role of peer evaluation has garnered significant attention as a potential catalyst for academic growth and development. Peer evaluation, a practice where students assess and provide feedback on the work of their peers, is increasingly being incorporated into educational settings as a means to foster collaboration, critical thinking, and a deeper understanding of subject matter. The project at hand aims to explore the intricate relationship between peer evaluation and students' academic performance. Specifically, our primary objective is to estimate the proportion of students who believe that peer evaluation has a positive effect on their academic performance in STA304H5 fall 2023. This investigation is not only timely but also holds the potential to inform educational practices, providing valuable insights for educators, administrators, and policymakers as they navigate the ever-evolving landscape of pedagogy.

Variables

| Variables | Definitions |
|---|---|
| Major of Study | S=Statistics<br>O=Other majors |
| Year of Study | 3=Year3<br>4=Year4 |
| Participated in Peer Evaluation | Y=Yes<br>N=No |
| Peer Evaluation Improved understanding | Y=Yes<br>N=No |
| Peer Evaluation Help Identify Improvement | Y=Yes<br>N=No |
| Peer Evaluation Positively Impact Academic Performance | Y=Yes<br>N=No |
| After Peer Evaluation Feeling More Confidence | Y=Yes<br>N=No |

## 2. Methodology

From October 25, 2023, to November 2, 2023, we surveyed the relationship between peer assessment and student study performance at the University of Toronto. (We used a computer random number generator to apply simple random sampling; 50-person questionnaires were collected from all students enrolled in the STA304H5 course (N=200).) The survey mainly asked students about their educational backgrounds and understanding of peer assessment. It included two multiple questions and four yes-or-no questions related to their major, year of study, understanding of course material, academic performance, confidence, impact, and one open-ended question.

To address the objective of this study, we will establish two point estimators, $\widehat{p_1}$, and $\widehat{p_2}$ ; under this definition:

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} y_i, \text{ where}$$

$$y_i = \begin{cases} 0 & \textit{with prob } 1-p \\ 1 & \textit{with prob } p \end{cases}$$

If $\widehat{p_1}$ and $\widehat{p_2}$ are both determined to be greater than 0.5, we will conclude that peer evaluation has a positive impact on students' academic performance. To do so, we shall do two things:

1. conduct a t test under the null hypothesis that the aforementioned point estimators are not greater than 0.5 at a significant level of 0.05
2. Compute a 95% confidence interval for both of the point estimators. The confidence intervals are to be computed in two distinct ways, formulaically and via bootstrapping.

We will treat a rejection of the null hypothesis and the confidence interval being uniformly above 0.5 as evidence in favor of drawing a positive conclusion (That is, peer evaluations benefit a student's academic achievements).

In order to evaluate the strength of our conclusion, we will compute the Kuder Richarson constant(KR20) of the observations.

## 3. Analysis

### 3.1 Sample Selection

Our study focused on the total number of students N = 200 enrolled in the course STA304H5. We used the statistical method of Simple Random Sampling method to extract 50 samples from it, which n = 50. Since the proportion is unknown, we assume p=q=0.5; and we select the error bound B = 0.1228, We use the following computation to evaluate the sample size as below:

$$n = \frac{Npq}{(N-1)D + pq} = \frac{200*0.5*0.5}{(200-1)*0.0038 + 0.5*0.5} = 50, \text{ where } D = \frac{B^2}{4} = 0.0038.$$

## 3.2 Assumptions

1. Independence of Observations: Each respondent's answers are assumed to be independent of one another. In other words, the responses of one student do not influence the responses of another student.
2. Random Sampling: The sample was obtained using a computer random number generator, applying simple random sampling. This assumes that every member of the population (students enrolled in the STA304H5 course) had an equal chance of being selected.
3. Law of Large Numbers: We computed $\widehat{p}$ = 0.58 for the proportion of students who believe peer assessment helps their academic performance. Using n = 50, np = 29 >10, n(1-p) = 21 > 10. Similarly for the proportion of students who believe peer assessment has a positive impact, where = 0.68. Np = 34>10, and n(1-p) = 16 > 10. The assumption is applicable.

## 3.3 Point Estimators

To quantify the results for calculation and visualization, for all yes-or-no questions, we set the answer “Yes” to a score of “1” and the answer “No” to a score of “0”. As stated earlier,

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} y_i, \text{ where}$$
$$y_i = \begin{cases} 0 & \text{with prob } 1-p \\ 1 & \text{with prob } p \end{cases}$$

i) The point estimate $\widehat{p_1}$ is denoted as the proportion of students who believe peer assessment helps them identify areas of academic performance that need improvement. In this situation, we got 50 observations, and 29 responses were “Yes”, and 21 responses were “No”. Therefore, we can have $\widehat{p_1}$ calculated as:

$$\widehat{p_1} = \frac{\text{Total number of answered "Yes"(1)}}{\text{Total number of response}} = \frac{29*1+21*0}{50} = 0.58$$

which can be comprehended as 58% of students who believe peer assessment helps them identify areas of academic performance that need improvement.

ii) The point estimate $\widehat{p_2}$ is denoted as the proportion of students who believe peer assessment has a positive impact on their academic performance. In this situation, we got 50 observations, and 34 responses were “Yes”, and 16 responses were “No”. Therefore, we can have $\widehat{p_2}$ calculated as:

$$\widehat{p_2} = \frac{Total\ number\ of\ answered\ "Yes"(1)}{Total\ number\ of\ response} = \frac{34*1+16*0}{50} = 0.68$$

which can be comprehended as 68% students who believe peer assessment has a positive impact on their academic performance.

## 3.4 Hypothesis Test & Test Statistics

The Main Hypothesis:

Let p’ be the true proportion of Fall 2023 STA304H5 students who believe peer assessment has a positive impact on their academic performance.
The null and alternative hypothesis are written as:

$$H_0: p' = 0.5 \quad v.s \quad H_a: p' > 0.5$$

The Second Hypothesis:

Let p be the true proportion of Fall 2023 STA304H5 students who believe peer assessment helps them identify areas of academic performance that need improvement.
The null and alternative hypothesis are written as:

$$H_0: p = 0.5 \quad v.s \quad H_a: p > 0.5$$

<u>Test Statistics:</u>

For the main hypothesis, the hypothesized population proportion(p) is 0.5, the sample proportion($\hat{p}$) is 0. 68, the sample size(n) is 50.

Hence, the the test statistic is calculated as $Z^* = \frac{\hat{p}'-p}{\sqrt{\frac{p*(1-p)}{n}}} = \frac{0.68 - 0.5}{\sqrt{\frac{0.5*0.5}{50}}} = 2.5456.$

The p value of this test is 0.1292.

For the second hypothesis, the hypothesized population proportion(p) is 0.5, the sample proportion($\hat{p}$) is 0. 58, the sample size(n) is 50.

Hence, the test statistic is calculated as $Z^* = \frac{\hat{p}-p}{\sqrt{\frac{p*(1-p)}{n}}} = \frac{0.58 - 0.5}{\sqrt{\frac{0.5*0.5}{50}}} = 1.1314.$

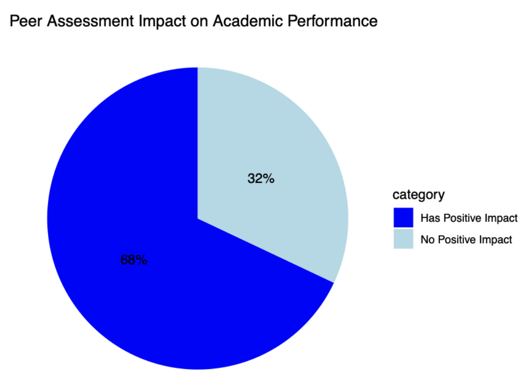
The p value of this test is 0.0055

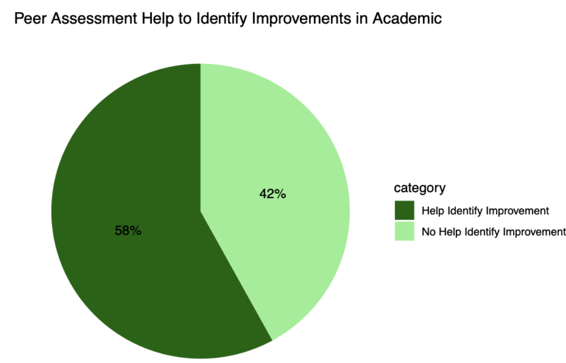
<u>Constructing 95% confidence interval for population proportion:</u>

For the main hypothesis, 95% confidence interval for the true proportion of Fall 2023 STA304H5 students who believe peer assessment has a positive impact on their academic performance is $(\widehat{p'} - B, \widehat{p'} + B)$, which is equal to (0.5572, 0.8028).

For the second hypothesis, 95% confidence interval for the true proportion of Fall 2023 STA304H5 students who believe peer assessment helps them identify areas of academic performance that need improvement is $(\widehat{p} - B, \widehat{p} + B)$, which is equal to (0.4572, 0.7028).

## 3.5 Data Visualization




## 3.6 Kuder-Richardson

Due to the reason that most questions in our questionnaire are binary questions, we want to choose the advanced method of Kuder-Richardson to test the reliability of our questionnaire, thus to see if the result we concluded is reliable or not.
As given, the formula of KR-20 is:

$$r = \frac{K}{K-1}\left[1 - \frac{\sum_{J=1}^{K} p_i q_j}{\sigma_X^{2}}\right]$$

In our survey, K is the number of binary test questions in the questionnaire; $p_j$ is the proportion that students answered "Yes"(denoted as "1") in $j^{th}$ question, and $q_j$ is the proportion that students answered "No"(denoted as "0") in $j^{th}$ question; Last but not least, for the $\sigma_X^2$, we have a formula: $\sigma_X^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$, where n is the total sample size. We take Yes as 1, and No as 0.

As in the questionnaire, we have 5 binary questions in total, so that K = 5;

$p_1 = 0.72;\ q_1 = 0.28$

$p_2 = 0.54;\ q_2 = 0.46$

$p_3 = 0.58;\ q_3 = 0.42$

$p_4 = 0.68;\ q_4 = 0.32$

$p_5 = 0.66;\ q_5 = 0.34$

$\sigma_X^{\ 2} = 3.58$

Thus, r = $\frac{K}{K-1}\left[1 - \frac{\sum_{J=1}^{K} p_i q_j}{\sigma_X^{\ 2}}\right] = \frac{5}{5-1}\left[1 - \frac{0.72*0.28+0.54*0.46+0.58*0.42+0.68*0.32+0.66*0.34}{3.58}\right] =$

$\frac{5}{4} \cdot (1 - \frac{1.1356}{3.58}) = 0.8535$

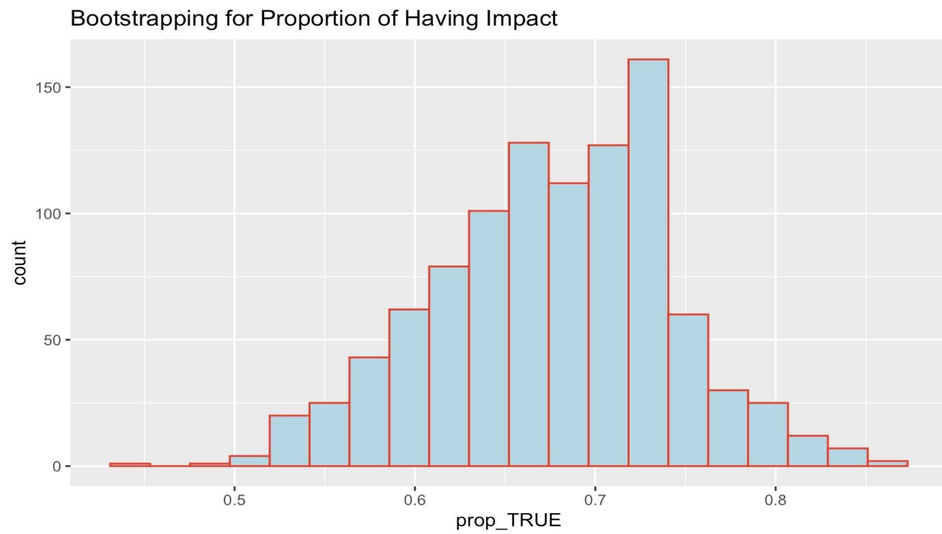
The Kuder-Richardson formula shows that our questionnaire's KR-20 score 0.8535 is at a moderate to high level, which means it can be generally considered to represent a reasonable level of internal consistency reliability.

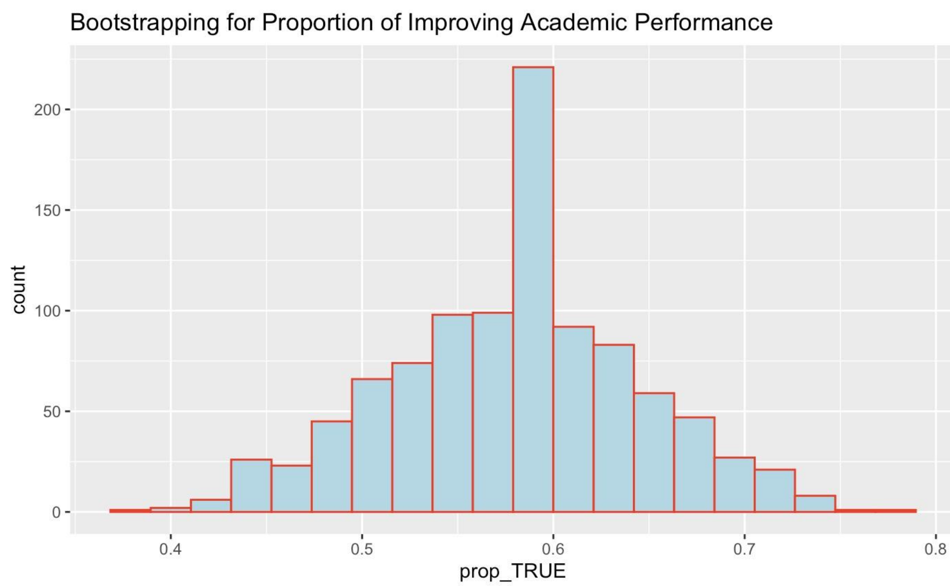
Therefore, we can say that our questionnaire has good reliability to conclude that at least 50% of the students believe that peer evaluation is beneficial for their academic performance by our hypothesis.

## 3.7 Bootstrapping

The following histograms are constructed by bootstrapping. Based on the original data set, we conducted 1000 bootstrap tests to obtain confidence intervals for two events. We calculate the 95% Confidence interval for the true proportion of STA304 student who feel peer evaluation has positive impact on their academic performance is (0.54, 0.80) by using R. In this case, we can conclude that we can reject the null hypothesis since $H_0$: $p' = 0.5$ and 0.5 does not belong to (0.54, 0.80)

Moreover, we also calculate the 95% Confidence interval for the true proportion of STA304 students who believe peer assessment helps them identify areas of academic performance that need improvement (0.44, 0.72) by using R. In this case, we can conclude that we can not reject the null hypothesis since $H_0$: $p = 0.5$ and 0.5 belongs to (0.44, 0.72).




# 4. Discussion/Results

Based on all the calculation and analysis mentioned above, we wish to conclude, as stated in the methodology, whether the true estimators are above 0.5.

We shall do so by commenting upon the t test results and the confidence intervals of the true point estimators for the two questions. For the specific calculating process and the formula, please check the appendix and refer back to the analysis section.

The point estimate $\hat{p} = 0.58$, the test statistics is 1.1314 with p-value = 0.1292, since $0.1292 > 0.05$, we fail to reject the null hypothesis. We have sufficient evidence and confidence to prove the true proportion of Fall 2023 STA304H5 students who believe peer assessment helps them identify areas of academic performance that need improvement is equal to 50%.

4.2 Having a positive impact on academic performance

The point estimator $\hat{p}$= 0.68, the test statistics is 2.5456 with p-value = 0.0055, since 0.0055 < 0.05, we will reject the null hypothesis. We have sufficient evidence and confidence to prove the true proportion of Fall 2023 STA304H5 students who believe peer assessment has a positive impact on their academic performance is greater than 50%.

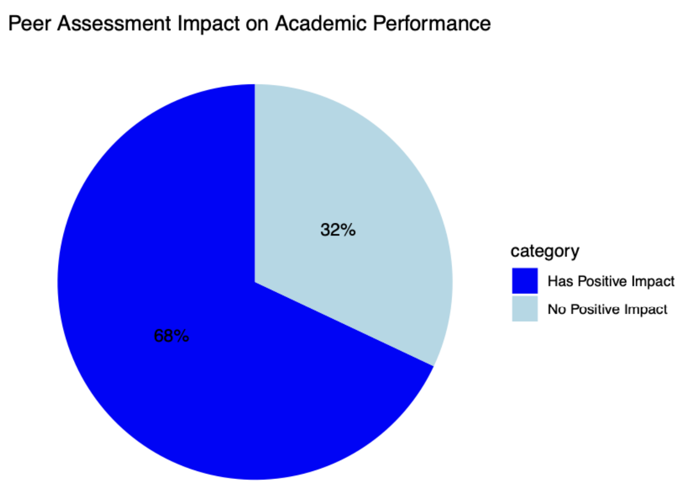
## 5.Conclusion/Limitation

-In conclusion, our study aimed to investigate the relationship between peer evaluation and student performance in the STA304H5 course during the fall of 2023. The analysis involved surveying 200 students using a Simple Random Sampling method, employing hypothesis testing, point estimators, data visualization, Kuder-Richardson reliability testing, and bootstrapping. The results indicate that a significant proportion of students believe that peer evaluation has a positive impact on their academic performance and helps them identify areas that need improvement. The Kuder-Richardson score of 0.8535 suggests good reliability in our questionnaire.

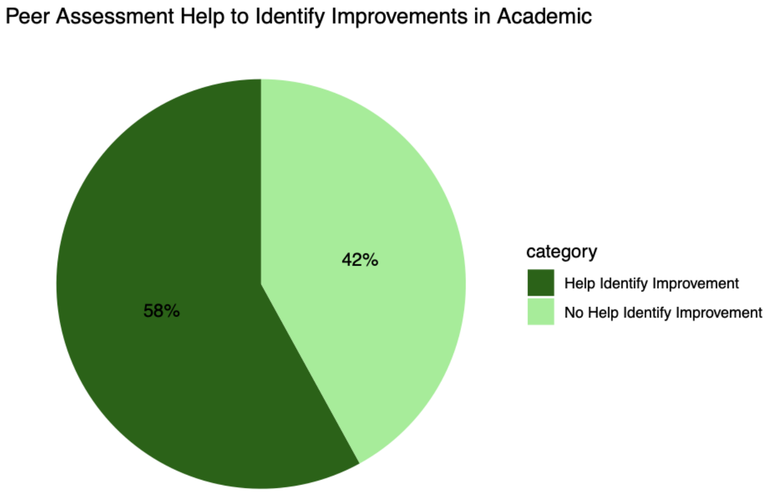
-The study faces several limitations that deserve attention. First, there is a risk of sampling bias. Despite efforts to implement random sampling, the sample may not fully represent the entire student population, potentially introducing bias. Second, our reliance on self-reported data in the study might be influenced by participant bias or misinterpretation of questions, impacting the accuracy of our findings. Third, the results may be too specific to the STA304H5 course and may not apply to other courses or academic disciplines. Fourth, the study's limited time frame might not capture the full range of experiences and opinions. Fifth, our questionnaire, designed in a binary question format, may oversimplify the complex nature of students' attitudes towards peer evaluation, limiting the depth of understanding. Lastly, assumptions of independence might be present, as student responses could be influenced by external factors, potentially violating the independence assumption.

## 6. Appendix

```r
positive_impact <- 34
no_positive_impact <- 16

data <- data.frame(
  category = c("Has Positive Impact", "No Positive Impact"),
  count = c(positive_impact, no_positive_impact)
)

data$percentage <- (data$count / sum(data$count)) * 100

library(ggplot2)

ggplot(data, aes(x = "", y = percentage, fill = category)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y") +
  theme_void() +
  geom_text(aes(label = paste0(round(percentage), "%")), position = position_stack(vjust = 0.5)) +
  scale_fill_manual(values = c("Has Positive Impact" = "blue", "No Positive Impact" = "lightblue")) +
  ggtitle("Peer Assessment Impact on Academic Performance")
```

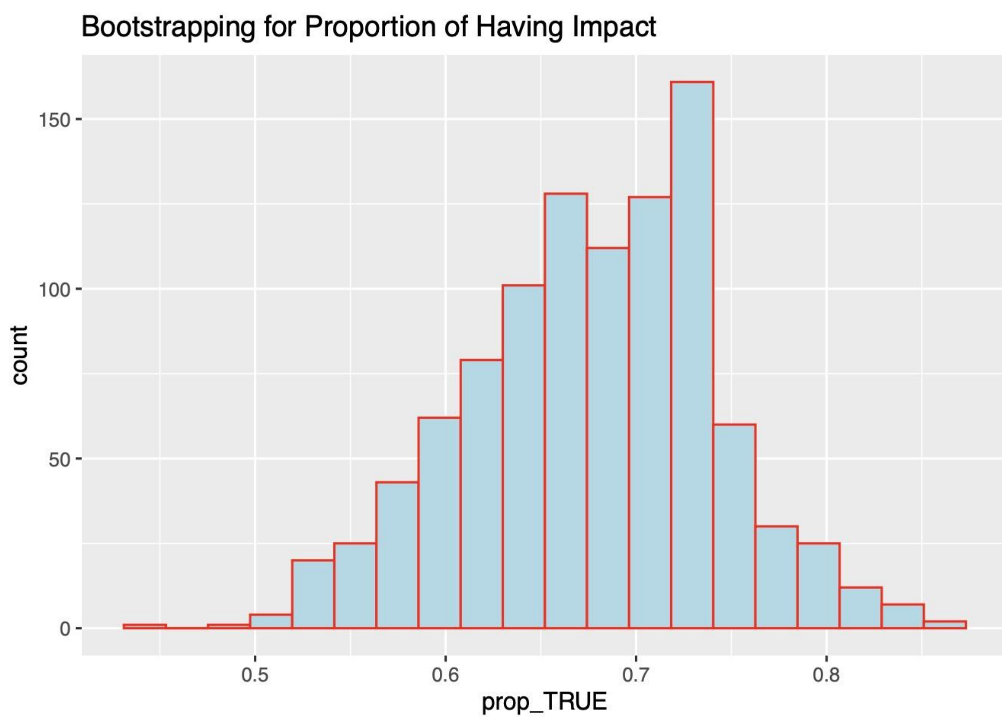



```
help_identify_improv <- 29
no_help_identify_improv <- 21

data <- data.frame(
  category = c("Help Identify Improvement", "No Help Identify Improvement"),
  count = c(help_identify_improv, no_help_identify_improv)
)

data$percentage <- (data$count / sum(data$count)) * 100

library(ggplot2)

ggplot(data, aes(x = "", y = percentage, fill = category)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y") +
  theme_void() +
  geom_text(aes(label = paste0(round(percentage), "%")), position = position_stack(vjust = 0.5)) +
  scale_fill_manual(values = c("Help Identify Improvement" = "darkgreen", "No Help Identify Improvement
  ggtitle("Peer Assessment Help to Identify Improvements in Academic")
```




```r
#Cleaning data
sta304 <- read_excel("Downloads/sta304.xlsx")
sta304 <- sta304[, -1]
```

```r
#Bootstrapping for impact
set.seed(0)
boot_impact = do(1000)*prop(~Impact == "Y", data = resample(sta304))
ggplot(boot_impact) +
  geom_histogram(aes(x = prop_TRUE), bins = 20, fill = "lightblue", color = "red") +
  ggtitle("Bootstrapping for Proportion of Having Impact")
```



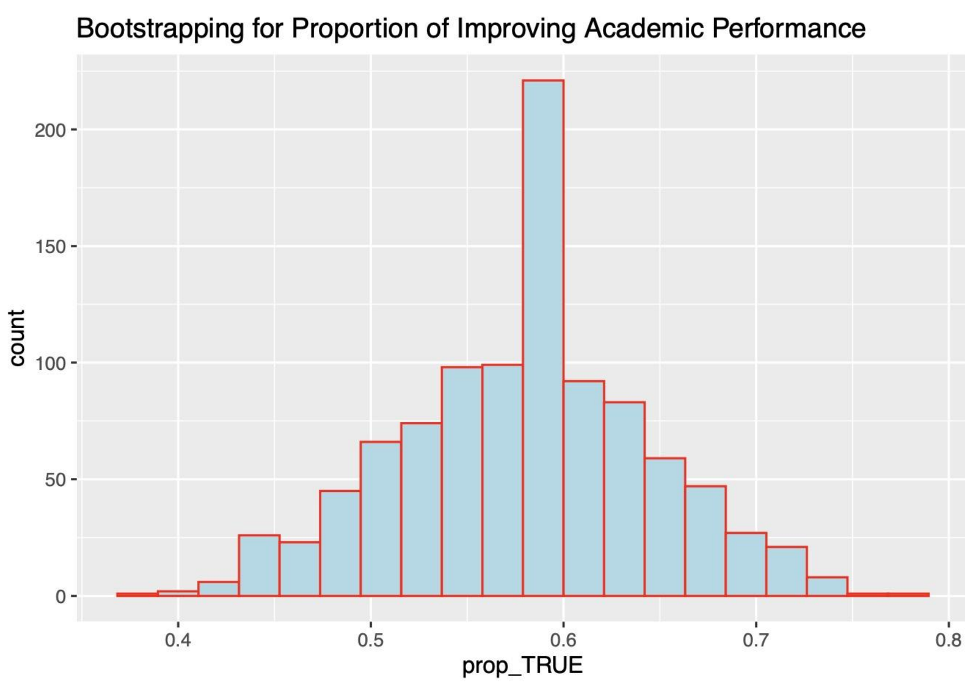
```r
#Standard Error
summarize(boot_impact, std_err_prop = sd(prop_TRUE))
```

```
##   std_err_prop
## 1   0.06554746
```

```r
#Bootstrapped Confidence Interval
confint(boot_impact, level=0.95)
```

```
##        name lower upper level     method estimate
## 1 prop_TRUE  0.54   0.8  0.95 percentile     0.68
```

```r
#Bootstrapping for Academic
set.seed(1)
boot_academic = do(1000)*prop(~Academic == "Y", data = resample(sta304))
ggplot(boot_academic) +
  geom_histogram(aes(x = prop_TRUE), bins = 20, fill = "lightblue", color = "red")+
  ggtitle("Bootstrapping for Proportion of Improving Academic Performance")
```




```r
#Standard Error
summarize(boot_academic, std_err_prop = sd(prop_TRUE))
```

```
##   std_err_prop
## 1   0.06969087
```

```r
#Bootstrapped Confidence Interval
confint(boot_academic, level=0.95)
```

```
##        name lower upper level     method estimate
## 1 prop_TRUE  0.44  0.72  0.95 percentile     0.58
```