PSTAT 122 Final

Isaiah Singer

June 5, 2025

```
#Libraries
library(readxl)
library(knitr)
library(ggplot2)
library(dplyr)
library(broom)
library(pwr)
```

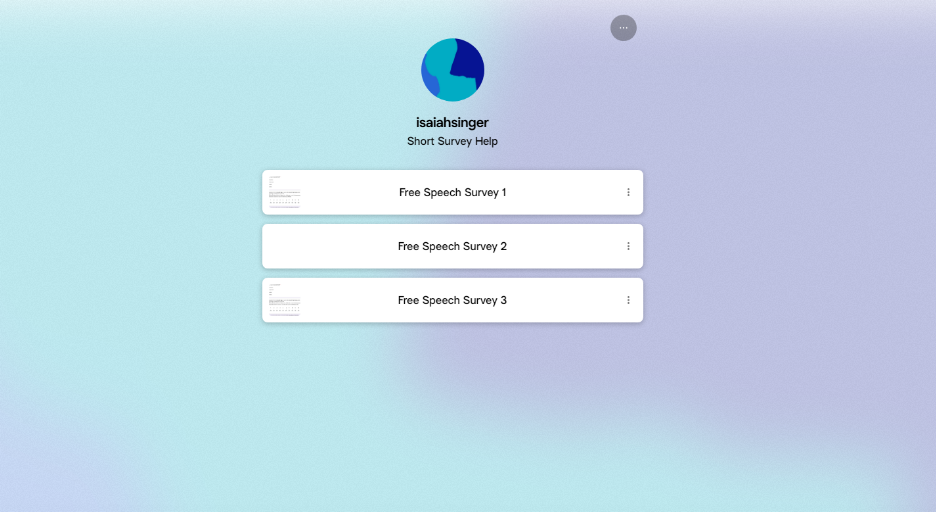
Final Project: Free Speech At UCSB

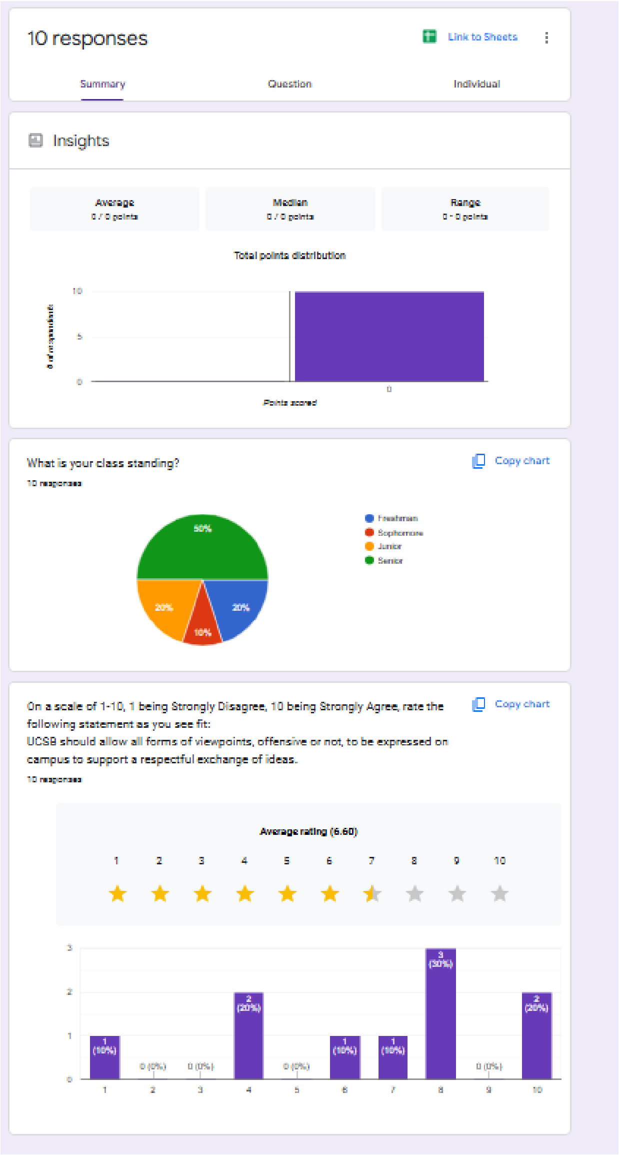
Introduction:

What constitutes as free speech has been as issue that has plagued not only universities in the recent years, but across society as well. The line between free speech, hate speech, and things of that nature is generally drawn very thin. This incites fear in students due to the possibility of being ostracized for their opinions on certain topics. While the boundaries of free speech are currently unknown in all of society as stated above, this study will take a focus on UCSB. Around the world, there have been similar studies done to determine how the student bodies at their respective university felt about the current restrictiveness of free speech. A study done at Goethe University Frankfurt in 2020 showed that a majority of students were “tolerant of different, even controversial viewpoints”(Revers, 2020). Despite the fact that the majority holds that belief, it is a common trend for people to silence themselves, whether that be out of fear or oppression. An important point to note is that it is commonly believed that political orientation has a large impact on an individual’s stance on the situation. Though that is not something that can be entirely disputed, nor is that the point of this study, a psychologist by the name of Mauricio Alvarez did a multilevel analysis across 37 states as well as a study that asked liberals and conservative their stance on free speech. The finding suggested that both liberals and conservatives were in support of free speech, although for different reasons. Conservatives “appear to be motivated by a focus on collectively held values” while liberals focused more on self-expression(Alvarez, 2018). Since this issue can be seen as divisive, it’s important to know that both sides of the political spectrum feel similarly. Since UCSB is generally associated with a predominantly liberal student body and staff, there doesn’t seem to be a political argument to be made considering the study by Alvarez .Free speech is an integral part of academic growth since growing adults need to be able to debate ideas that they oppose and other ideas they’d like to defend in order to be educated and to educate. Conversely, allowing all forms of expression opens gateways for potentially hurtful or offensive language, which would otherwise have been restricted.The consequence of the duality of the situation forces students to weigh the importance of each side and determine what is most beneficial to campus life. The purpose of this experiment is to determine whether students on campus feel similarly, or if they prefer a restrictive atmosphere that would lead to less growth, but less people taking offense as well.

Methods:

The study chosen for this experiment was based on the first option offered, a Generalized Randomized Block Design.I created an online survey for UCSB students with the goal of determining how students felt about free speech on campus, and whether the tone of the statement changed their overall opinion. In order to do so, I created 3 separate surveys on Google Forms(https://linktr.ee/isaiahsinger). the surveys are phrased in three ways: Treatment 1 phrases it in a way to allow respectful exchanges of ideas between students, Treatment 2 takes the stance of a threat to free speech, and Treatment 3 takes a scholarly approach where free speech can promote critical thinking and individual growth. In all the surveys, the block factor was the class standing of the student, with the choices being Freshman, Sophomore, Junior, and Senior. The factor of interest would be the rating from 1-10, 1 being strongly disagree, and 10 being strongly agree. With a sample size of n = 21, it may be difficult to make concrete conclusions about how students feel about free speech given our small n. The randomization process was less than optimal and was done by providing the linktree link above and having participants click the survey of their choosing. It was distributed through the UCSB Snapchat story which is only accessible by UCSB students and is available to all students regardless of class standing. Technical issues were not much of a problem, but the major issue was actually getting students to participate in the surveys. Without an incentive, it can be difficult for university students to participate in activities given the busyness of many students. The ANOVA assumptions required in my study would be random assignment, no noticeable trend to the structure of the data, and equal variances. There doesn't currently seem to be reason to believe that my data would make any departures from normality, but the inefficiency of the random assignment may impact those results, and that will be something that will be thoroughly tested in a separate section of the study. Below you will find images of my survey(s) and the randomization method, being the linktree link mentioned above:





10 responses

Link to Sheets

Summary Question Individual

Insights

Average
0 / 0 points

Median
0 / 0 points

Range
0 - 0 points

Total points distribution

What is your class standing?

10 responses

Copy chart

On a scale of 1-10, 1 being Strongly Disagree, 10 being Strongly Agree, rate the following statement as you see fit:
UCSB should allow all forms of viewpoints, offensive or not, to be expressed on campus to support a respectful exchange of ideas.

10 responses

Copy chart

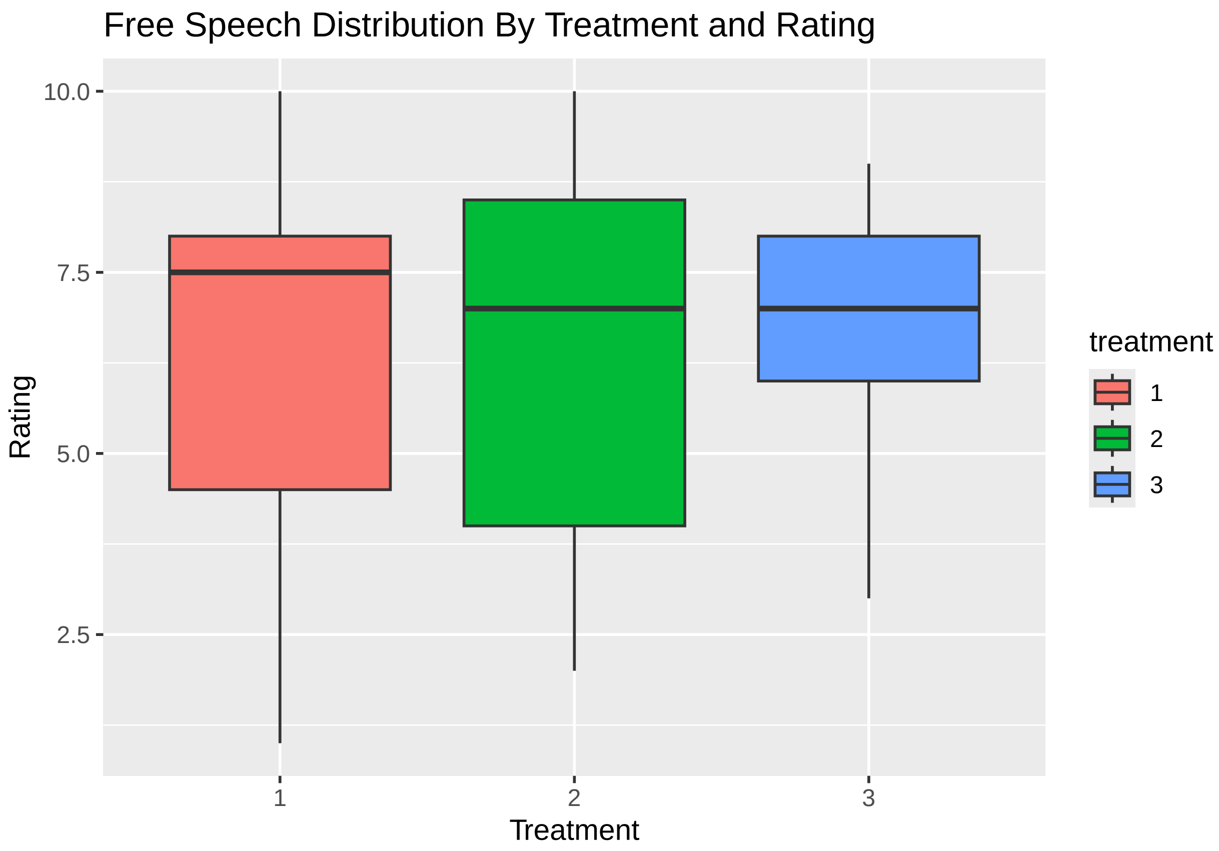
Average rating (6.60)

Results:

```r
#Reading Data
speech <- read_excel("final_data.xlsx")
speech$treatment <- factor(speech$treatment)
```

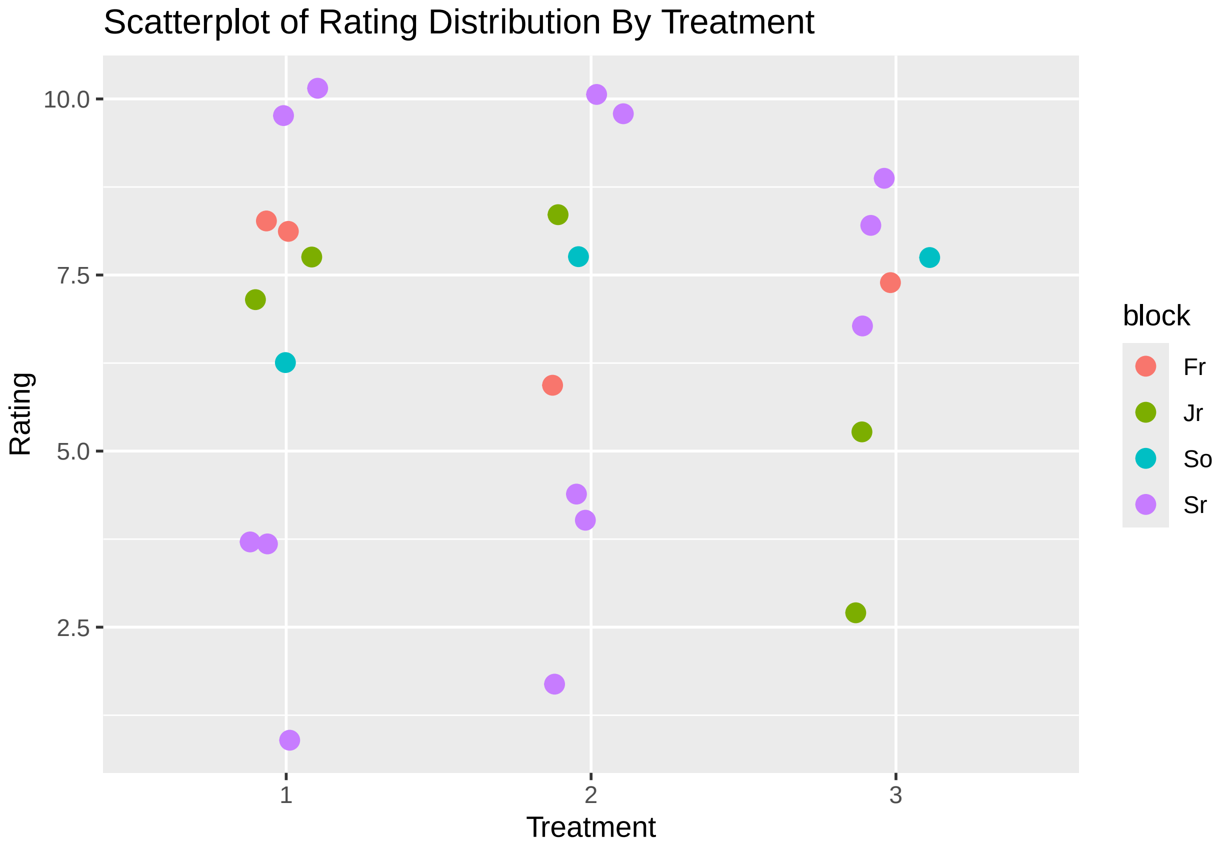
To begin with, let's start with a boxplot of our data which you may find below. The boxplot will be a representation of the distribution of our variables treatment and rating, and seeing if the treatment had a noticeable effect on the given ratings or not. Of course, we will need to do further investigation to determine whether that really is the case or not, but the boxplot will help us have a rough understanding of how our data may look.

```r
#Free Speech Boxplot
ggplot(speech, aes(x = treatment, y = rating, fill = treatment)) +
  geom_boxplot() +
  labs(
    title = "Free Speech Distribution By Treatment and Rating",
    x = "Treatment",
    y = "Rating"
  )
```



At face value, there doesn't actually seem to be too much difference in rating based on the treatment groups. Due to the fact that there weren't enough participants in every single block for all 3 surveys, creating a graph to represent ratings based on our block factor is not exactly ideal. From this graph, we can see that our second treatment had the largest variance out of the three groups, and the median value between treatment 2 and 3 is very similar. Our first treatment method had the highest median with the second lowest variance, while our third treatment method had the lowest variance among the groups. It's hard to make any concrete assumptions based on this graph aside from what I have previously said, and this is still just meant to give us a general understanding of how our data may behave for the next tests that we will do. If I were to make an assumption, the data may not be statistically significant and if anything comes back as statistically significant, I will have to verify my p-values through the normality test to ensure that it really is the case. In order to get a better grasp of our block factor as well, we will include another graph below that may help shed some light on how the participants were voting.

```
#Scatterplot of same data with block as the color
ggplot(speech, aes(x = treatment, y = rating, color = block)) +
  geom_jitter(width = 0.15, size = 3) +
  labs(
    title = "Scatterplot of Rating Distribution By Treatment",
    x = "Treatment",
    y = "Rating"
  )
```



For the most part, the data seems to come off as random, even when the block factor is included. The highest scores of rating were given by Seniors, yet the lowest scores by rating were also given by Seniors. It is important to note that Seniors were more present in the sample size calculations than participants of other class standings. Freshman and Sophomores gave similar ratings around the range of approximately 5 to 8. This could be due to the fact that they haven't solidified their opinions on matters as much as Juniors and Seniors may have. The reason for this being there is clearly more variance in the responses from Juniors and Seniors than there are from Freshman and Sophomores.

Now that we have given some graphical representations of our data to try to get an idea of what we're working with, the only way to truly know is with our p-values that we receive from our ANOVA test. The ANOVA will help give numerical values to the significance that we aren't able to achieve with the graphs. Once again, due to the uneven sample sizes, there may be some errors with the ANOVA and that will be followed up later in the study. With that being said, let us perform an ANOVA test and see the results for ourselves.

```
#ANOVA test to determine significance
speech_anova <- aov(rating ~ treatment + block, data = speech)
anova_res <- anova(speech_anova)
kable(anova_res)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	2	0.1714286	0.0857143	0.0103653	0.9896938

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
block	3	4.7113979	1.5704660	0.1899147	0.9019517
Residuals	19	157.1171735	8.2693249	NA	NA

Let's start with forming our null and alternative hypothesis. Our null hypothesis, denoted as H_0, will be defined to be there was no mean difference in rating score based on our factors. Our alternative hypothesis, denoted as H_a, will be defined to be that at least one of the group means differed from the rest.

From our ANOVA, we can see that neither our treatment method nor our block factor makes a statistically significant difference on the rating that the participant provided. With our set α level of .05, we received a p-value of ≈.9897 for our treatment and ≈.9012 for our block factor. Both of these p-values are considerably higher than our set alpha level, and we would then conclude that they are not statistically significant and fail to reject H_0.

Since we are unsure if our p-values can be trusted or not, we are also going to run a permutation test of our data and see how the p-values compare to one another. If there is a lot of variance in the p-values, we would then know that our ANOVA p-values can't be trusted. The benefit of the permutation test is that it works better for data with small sample sizes since no required assumptions need to be met such as normality or equal variances.

```r
#Permutation test with assistance of ChatGPT
set.seed(17)
obsF <- anova(speech_anova)["treatment", "F value"]
reps <- 5000
permF <- replicate(reps, {
  speech_perm <- speech
  speech_perm <- speech_perm %>%
    group_by(block) %>%
    mutate(treatment = sample(treatment)) %>%
    ungroup()
  anova(aov(rating ~ treatment + block, data = speech_perm))["treatment", "F value"]
})
p_perm <- mean(permF >= obsF)
p_permdf <- data.frame(
  `p-value` = round(p_perm, 3)
)
kable(p_permdf, caption = "Permutation Test p-value")
```

Table 2: Permutation Test p-value

p.value
0.992

Surprisingly, the p-value we received from our permutation test gave us a p-value of .992. The p-value we received from our ANOVA earlier was ≈.989. These values are relatively similar and there doesn't actually seem to be reason to not trust our p-values to my surprise.

Moving on to our pairwise comparisons, we will see if any of the blocking groups are statistically significant and fail to reject H_0. Considering the sample size, there is a very high chance that we will receive a Type I error. But regardless, we will proceed with the test and make conclusions from the p-values.

```
#Pairwise comparisons
speech_tukey <- TukeyHSD(speech_anova, "treatment")
tukey_res <- tidy(speech_tukey)
kable(tukey_res)
```

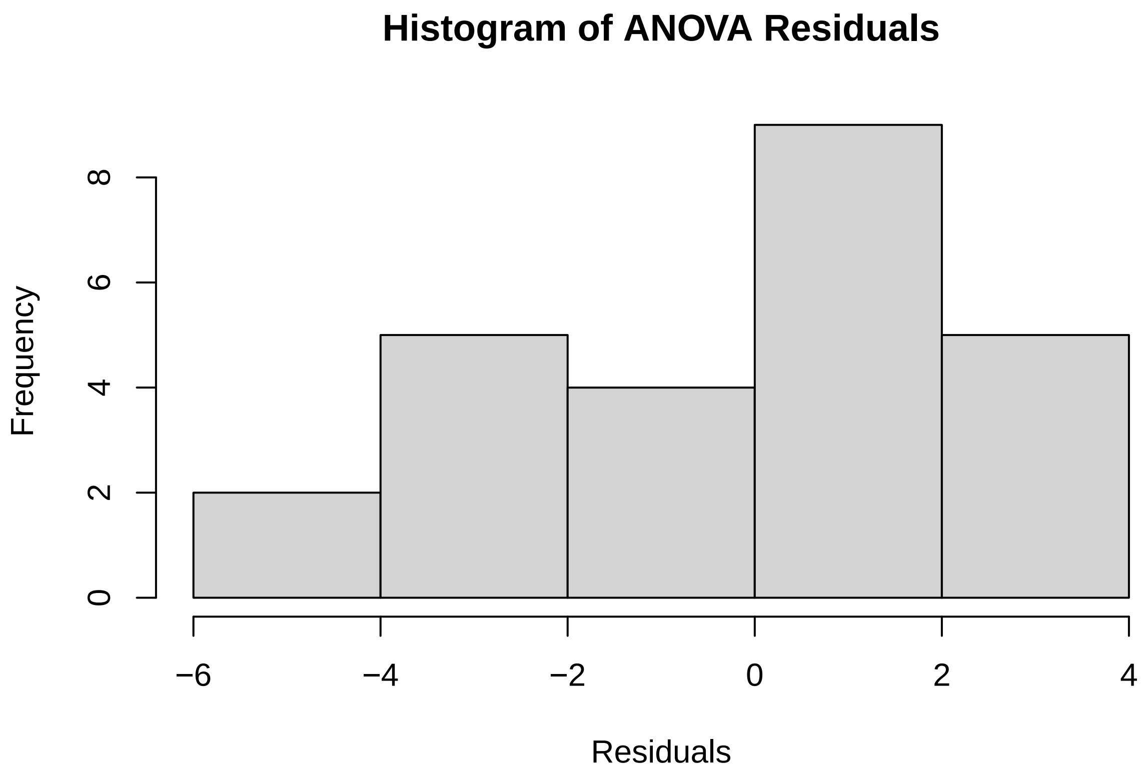
term	contrast	null.value	estimate	conf.low	conf.high	adj.p.value
treatment	2-1	0	-0.1000000	-3.565269	3.365269	0.9970417
treatment	3-1	0	0.1142857	-3.485869	3.714440	0.9964215
treatment	3-2	0	0.2142857	-3.566633	3.995204	0.9886430

The pairwise comparisons showed similar results as our ANOVA tests. With that being said, that would mean the wording of my different statements had no statistically significant effect on the ratings that the participants gave the surveys. There is an obvious lack of power with my small sample size, and if a test with a larger sample size were done on the entire student body of UCSB, these results may differ greatly from what we are seeing currently.

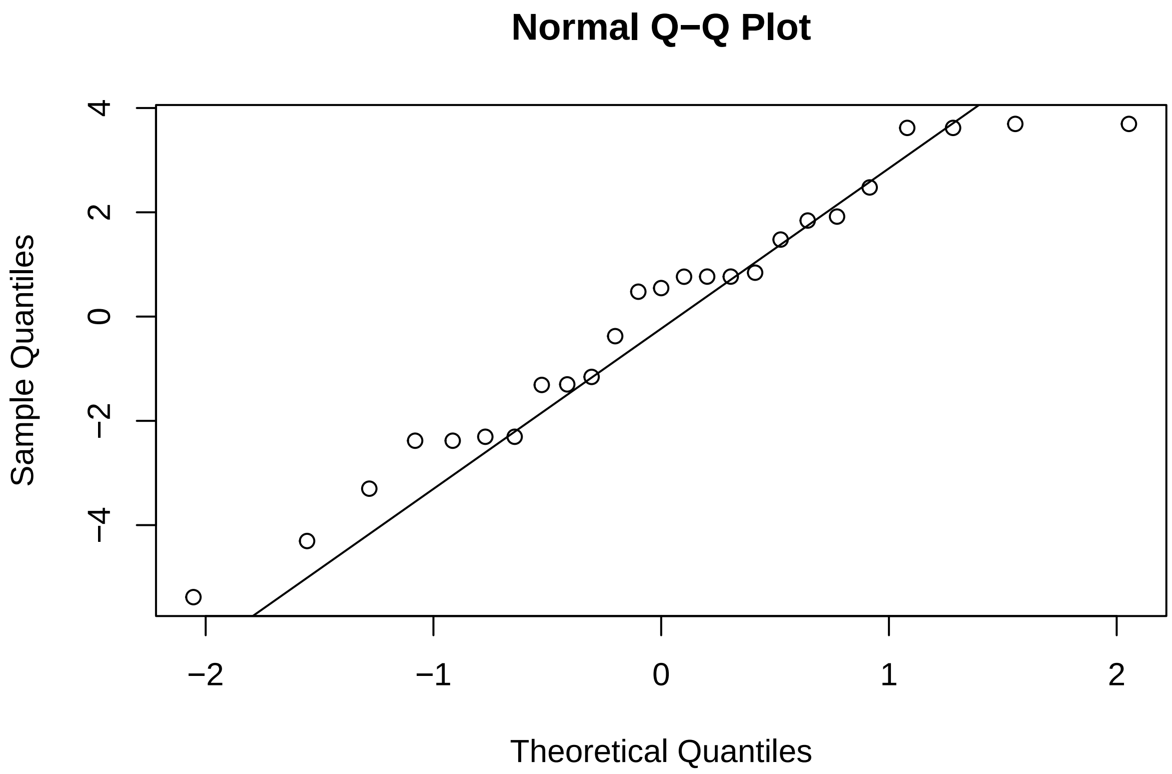
Continuing on, we can now test the normality of our data to see if is making any departures from normality or not. In order to do so, I will show a variety of tests and graphs below that will help us determine the normality. And below the graphs, I will draw conclusions on what the case may be.

```
#Normality Checking

#Histogram of plotting the residuals of our ANOVA
hist(speech_anova$residuals, main = "Histogram of ANOVA Residuals",
     xlab = "Residuals")
```



```
#QQ-Plot of ANOVA Residuals
qqnorm(speech_anova$residuals)
qqline(speech_anova$residuals)
```

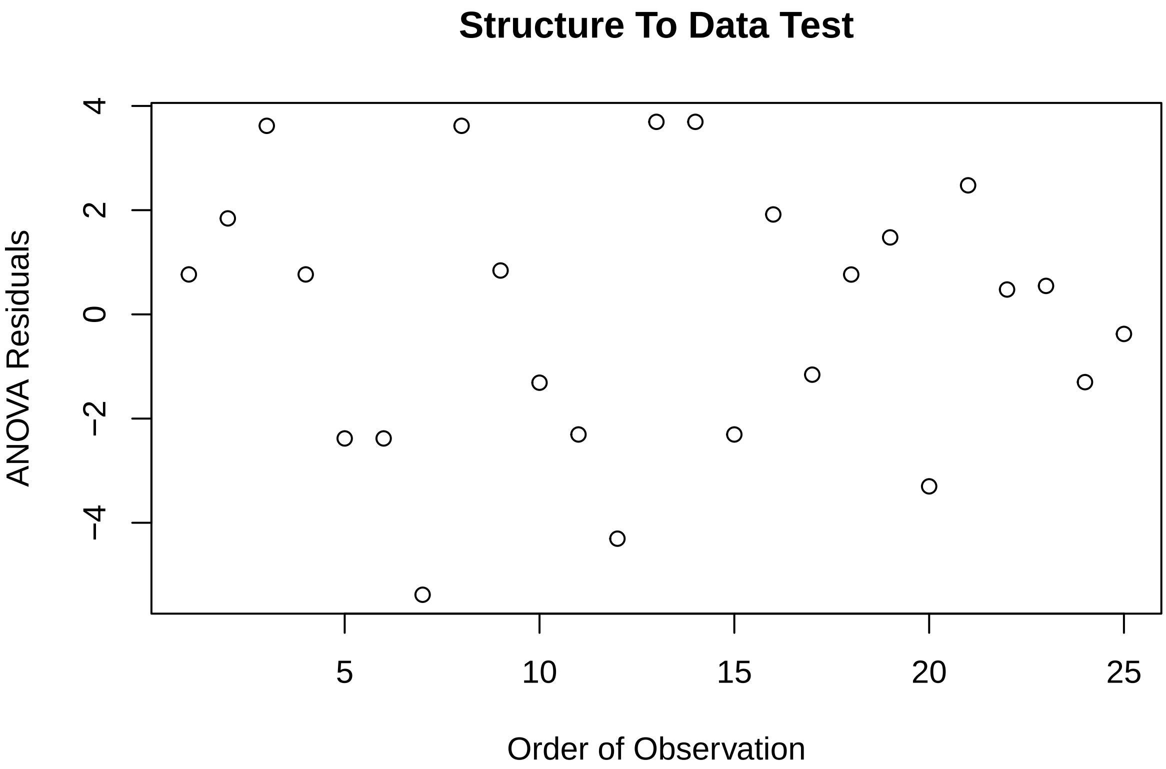


```
#Shapiro-Wilk Test
shap_test <- tidy(shapiro.test(speech_anova$residuals))
kable(shap_test, caption = "Shapiro-Wilk ANOVA Normality Test", col.names = c("W", "p-value", "Method")
```

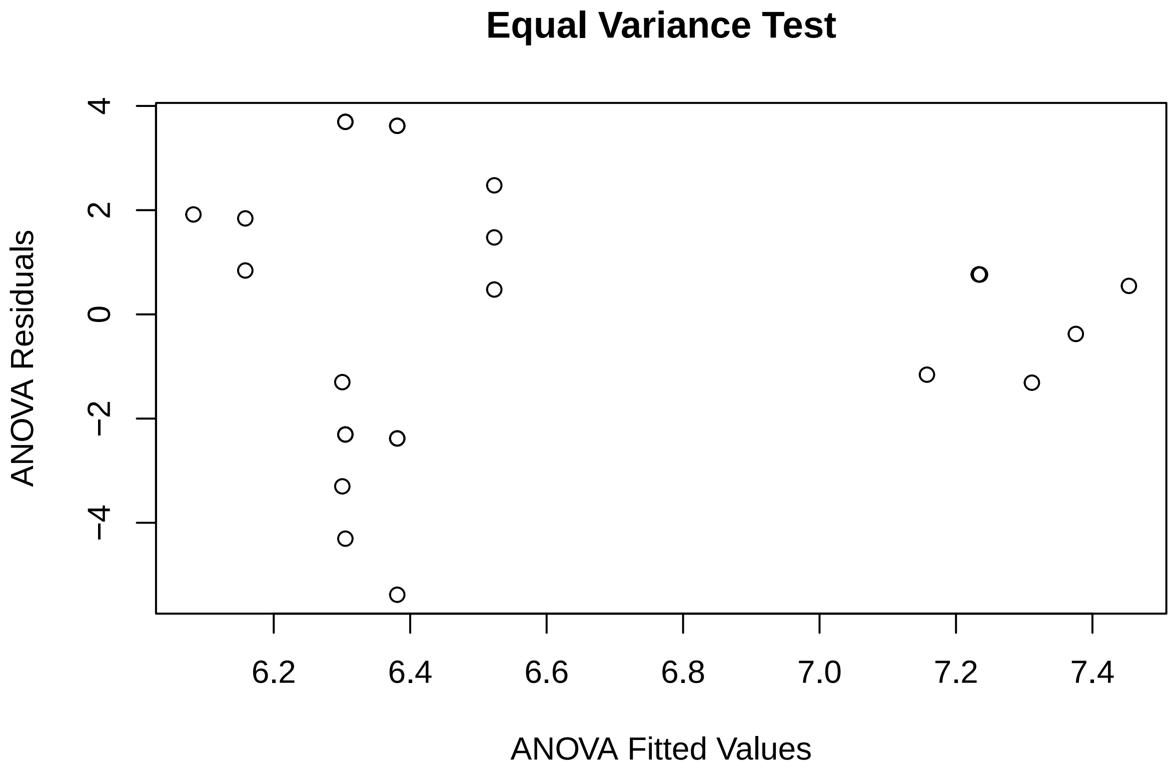
Table 4: Shapiro-Wilk ANOVA Normality Test

W	p-value	Method
0.9561999	0.3439973	Shapiro-Wilk normality test

```
#Structure to the Data
x <- 1:length(speech_anova$residuals)
plot(speech_anova$residuals ~ x, main = "Structure To Data Test", xlab = "Order of Observation",
     ylab = "ANOVA Residuals")
```



```
#Equal Variance
plot(speech_anova$residuals ~ speech_anova$fitted.values, main = "Equal Variance Test", xlab = "ANOVA F
     ylab = "ANOVA Residuals")
```



We can see a noticeable left-skew on our histogram which is not a good sign at all. There may be some resemblance of a bell curve but it is hard to determine considering the skew. In our QQ-Plot , we see a lot of departures from our line which is also a cause for concern. The S-shaped curved may be a sign of some departures from normality as well.From our Shapiro-Wilk test, we received a p-value of ≈ .344. Given that information alone, we would claim that we would fail to reject our null and that there is no reason to assume our data is non-normal, but that may not be the case here. Given our small sample size, it is very possible that we are receiving a Type II error. The structure to our data does seem random and there does not seem to be any noticeable trends within the data, so that much is good. But, when it comes to our equal variances, there does appear to be some signs of nonlinearity. The reason that is the case is because the variance does not seem to be constant among the residuals. In conclusion, our Shapiro-Wilk test gave us a p-value of .344 which would normally indicate that our data is normal. The caviat with this data set is that nearly all the rest of our tests is not supporting the p-value that our Shapiro-Wilk test has outputted. That is definitely a cause for concern and it's uncertain whether our p-values for this experiment can be trusted.

For the sample size calculations, I will be doing post-hoc power calculations. The reason for that being because the other options were a literature review or pilot study. While I mentioned that the studies referenced above can relate to UCSB, there is no officially released study on free speech at UCSB. For that reason, I would find the literature review to be somewhat arbitrary for this specific study and opted against it. Since I did a real data option, the surveys that I created are my pilot study. With that being the case, it would be easy to get the effect size from my real data. Therefore, post-hoc seemed most optimal for my real data study, and determining the effect size versus the sample size over a large range of values will give me a better understanding of the present data. Below you will find the code for the post-hoc calculations and graphical representation of the effect size and sample size:

```
#Post-hoc power calculations with assistance of ChatGPT
```

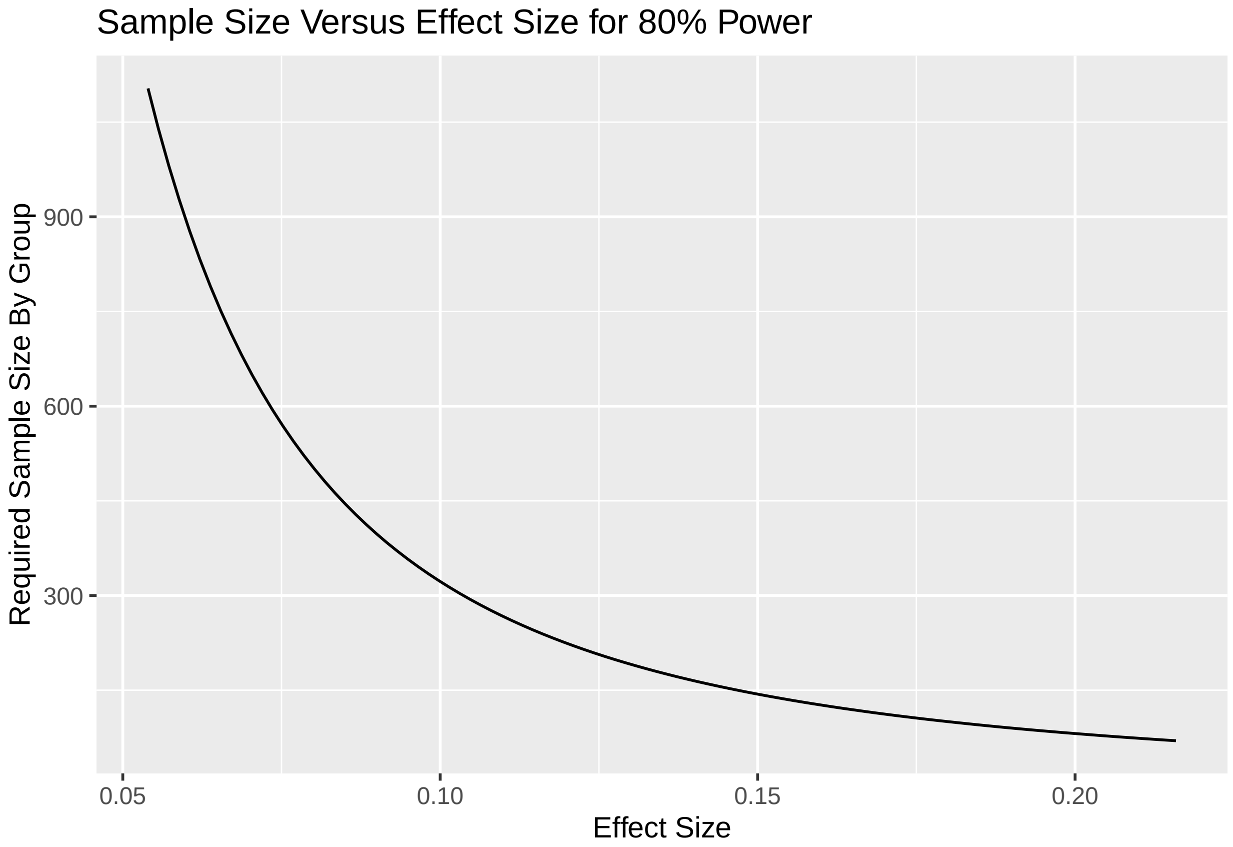
```r
#Computing Mean Squares Between and Within denoted as MSB and MSW
by_treatment <- speech %>%
  group_by(treatment) %>%
  summarize(n = n(), mean = mean(rating), var = var(rating))
grand_mean <- mean(speech$rating)
#Sum of Squares and Degrees of Freedom
SSB <- sum(by_treatment$n * (by_treatment$mean - grand_mean)^2)
SSW <- sum((by_treatment$n - 1) * by_treatment$var)
dfB <- n_distinct(speech$treatment) - 1
dfW <- nrow(speech) - n_distinct(speech$treatment)
#Mean Squares
MSB <- SSB/dfB
MSW <- SSW/dfW

#Post-hoc f
f_vals <- sqrt(MSB/MSW)

#Range of Effect Sizes
f_range <- seq(from = f_vals/2, to = f_vals*2, length.out = 100)

#Required n Per Group
ss <- sapply(f_range, function(f){
  pwr.anova.test(k = 3, f = f, sig.level = .05, power = .8)$n
})

#Tibble
ss_df <- tibble(f = f_range, n_per_group = ss)
ggplot(ss_df, aes(x = f, y = n_per_group)) +
  geom_line() +
  labs(
    x = "Effect Size",
    y = "Required Sample Size By Group",
    title = "Sample Size Versus Effect Size for 80% Power"
  )
```



The f value we received from our ANOVA was calculated by taking the square root of the mean square of our treatment divided by the mean square of our residuals. This value was ≈.1018. With that being said, we can now analyze the graph of the sample size vs effect size. Considering our approximation .10, we would then look at the x-axis for effect size for .10 and see where that corresponds to our y-axis for the required sample size. As you can see, at .10 effect size, we would require a sample size of around 300 for our power to reach 80%. This is likely due to the fact that it is difficult to find statistically significant evidence solely from a slight change in tone or emphasis in a sentence without an extremely large sample size backing it. Because of that, the results were relatively underwhelming and I would've needed to potentially outsource to different universities in order to get my sample size to that range. Another option could've been increasing my effect size by changing the structure of the sentences more so that our mean squares would give us a larger f value for the effect size. If our effect size is larger, we require a smaller sample size since it is diminishing.

Discussion:

Overall, the results of the lab were pretty underwhelming. All of the data I collected came back as statistically insignificant. That goes as far as the ANOVA p-values, the pairwise comparisons p-values, and the permutation test p-values. This was due to the small sample size that I was unable to achieve since many students are unwilling to participate in a survey for nothing in return. From the post-hoc test, we saw that we would've needed a sample size of at least 300 to hit 80% power with an effect size of .10. Regardless of the small sample size, we were not able to find any conclusive evidence that the different approaches of my treatment methods had an effect on whether students wanted free speech to be less restricted or not. If we were to extrapolate from that point, regardless of the goal being scholarly, the sake of free speech, or whatever other reason you may come up for loosening the restrictiveness of free speech, it doesn't seem to change the opinion for students at UCSB. This may mean that students are content with the restrictiveness of speech at UCSB, or it may mean that students don't feel restricted in the way they speak on campus.

In order to draw conclusions such as those, it would be imperative to get a larger sample size and test the data again though. For the time being, the only conclusions that can be drawn from my study are that my different treatments did not have an effect on the rating that participants gave to my surveys. The normality assumptions of our data seem to have been met given the p-value of .3434 from our Shapiro-Wilk test, but there were strong possibilities given the histogram, QQ-Plot, and Equal Variances that there may be some departures to normality as well as nonlinearity in our variances. At the end of the day, free speech is not a topic of discussion that can be polarized so easily. A book written by Robert O'Neil states the same and says that the "nature and status of campus speech offers an intriguing paradox"(O'Neil, 1997). It even goes to show that debates among free speech has dated back decades ago, and it's only become even more difficult to differentiate between free speech and hate speech in contemporary society. There are many reasons to let students speak more freely, but for every reason that there is, there's another reason on the other end as to why they should not. For now, I will have to resign myself with the current data that I have collected and draw the conclusion that students at UCSB may not feel that free speech is a pressing issue.

References:

Revers, M., & Traunmüller, R. (2020, October 26). Is free speech in danger on university campus? some preliminary evidence from a most likely case - KZFSS Kölner Zeitschrift für Soziologie und Sozialpsychologie. SpringerLink. https://link.springer.com/article/10.1007/s11577-020-00713-z#author-information

Alvarez, M. J., & Kemmelmeier, M. (n.d.). Free speech as a cultural value in the United States. Journal of Social and Political Psychology. https://journals.psychopen.eu/index.php/jspp/article/view/5031

O'Neil, R. M. (n.d.). Free speech in the College Community. Google Books. https://books.google.com/books?hl=en&lr=&id=MuhT8XIXZSMC&oi=fnd&pg=PR7&dq=free%2Bspeech%2Bat%2Buniversities%2Bstudy&ots=_uXhab9Jr4&sig=TBS9mDwQQD1UOvFBRY4cwT--kqg#v=onepage&q=free%20speech%20at%20universities%20study&f=false