

Statistical Analysis

2024-10-21

Contribution

1. Introduction

This report analyzes a survey conducted among 3000 - 4000 students enrolled in statistics courses at UC Berkeley during the Fall 1994 semester. The survey aimed to investigate students' video game habits, preferences, and attitudes towards gaming. A total of 314 students from Statistics 2, Section 1 were eligible for the survey based on their participation in the second exam of the semester, with 95 students randomly selected to participate. Out of the selected group, 91 students submitted complete responses, with incomplete responses encoded as 99.

The survey covered questions including the number of hours spent playing video games, where and how often students play, their preferences for different types of games, etc. The second part of the survey covered students' attitudes toward video games, whether they liked playing games or not and why.

The primary goal of this analysis is to investigate students' video game behaviors and preferences, as well as potential correlations between these behaviors and demographic factors such as gender, work, and computer ownership. Specifically, we aim to investigate the proportion of students playing video games during the week prior to the survey, the relation between reported time spend playing video games, the frequency of playing, and timing of the survey, the main reasons students like or dislike playing video games, and any noticeable differences in video game preferences and habits based on demographics. We found that **insert results of analysis here**

The report is structured as follows: Section 1 addresses each of the research questions outlined above, Section 2 analyzes the survey data, including numerical and graphical comparisons of video game habits and preferences. Section 3 expands on the analysis to explore potential differences across subgroups. Section 4 summarizes the findings and discusses their implications, and Section 5 provides additional information and appendices related to the analysis.

2. Analysis

2.1 Estimate Proportions of Students Who Played a Video Game

2.1.1 Methods

The first step of our analysis was to estimate the proportion of students who played a video in the week prior to the survey. We found both a point estimate and an interval estimate for this proportion.

The point estimate represents a single estimated value of the proportion of students who played video games based on the sample. The sample proportion \bar{X} is calculated as the ratio of students who played video games (i.e., reported playing for more than 0 hours) to the total number of students surveyed.

The formula for the sample proportion is given by:

$$\bar{X} = \frac{k}{n}$$

where k is the number of students who played video games, and n is the total number of students surveyed.

The interval estimate provides a range that is likely to contain the true proportion of students who played video games. This is calculated using a 95% confidence interval.

The confidence interval is calculated as:

$$(\bar{X} - 2\frac{s}{\sqrt{n}}, \bar{X} + 2\frac{s}{\sqrt{n}})$$

where $\frac{s}{\sqrt{n}}$ is the standard error; s is the sample standard deviation.

2.1.2 Analysis

To find the point estimate, we found the sample proportion by counting the number of students whose playing time a week prior to the survey was greater than zero and the total number of responses to the survey. The proportion of students who played video games in the week prior to the survey is approximately 37.36%. While the point estimate gives a specific value, it does not account for the inherent variability that comes with sampling. To take this uncertainty into account, we calculated a 95% confidence interval. The 95% confidence interval for the proportion of students who played video games is between 27.22% and 47.51%.

2.1.3 Conclusion

A point estimator is a single value \bar{x} , useful for summarizing the central tendency. In this case, it estimates the proportion of students that played video games a week prior to

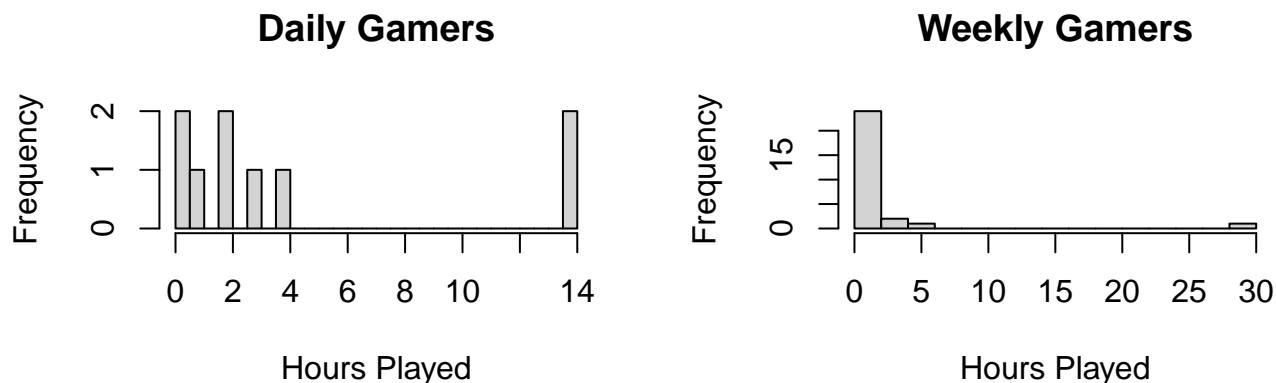
the survey given a sample. An interval estimate is a range of two values, where one can estimate that the population proportion is between those two values with a fixed confidence (probability), reflecting the variability in the sample data. The interval is based on the assumption that our sample is representative of the entire population of students who took the survey. The confidence interval ensures that if the survey were repeated multiple times, 95% of the intervals calculated in this manner would contain the true population proportion. In our case, the interval estimate is very wide, reducing its usefulness in estimating the population proportion, since the true proportion could plausibly fall anywhere between 27% and 47%.

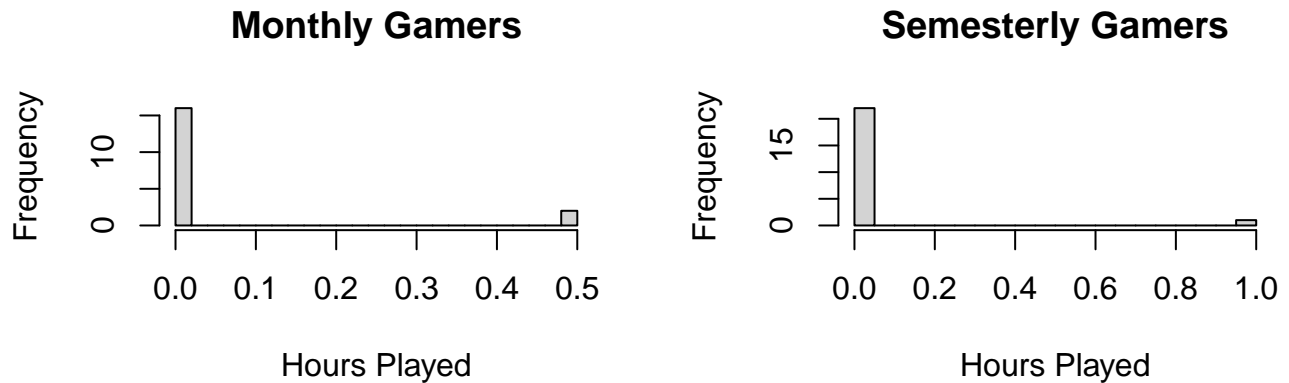
2.2 Comparison of Time Spent and Reported Frequency of Playing Video Games

2.2.1 Methods

Next, we wanted to look at the if the amount of time spent playing games in the week prior to the survey compares for different reported frequencies. In the survey, students were asked to report how often they play video games on a scale from 1-4, where 1 represents playing daily, 2 represents playing weekly, 3 represents playing monthly, and 4 represents playing semesterly. To compare the amount of time spent playing games between different reported frequencies, we graphed the distributions of time spent playing games and calculated the proportion of students who reported playing video games in each frequency category and the average time they spent playing. This comparison helped us assess whether students' reported gaming habits align with the actual hours they spent playing video games. We considered the potential impact of an exam held the week before the survey, which may have influenced the amount of time students spent playing video games that week.

2.2.2 Analysis





Based on these graphs, it becomes apparent that daily and weekly gamers spent a lot more time playing games a week prior to this survey compared to monthly and semesterly gamers.

Table 1: Sample Proportions Based On Reported Frequency

	Proportions
Daily	77.78%
Weekly	85.71%
Monthly	11.11%
Semesterly	4.35%

Similarly, the numerical summaries show that more daily and weekly gamers spent time playing games the week prior to the survey than monthly and semesterly gamers.

2.2.3 Conclusion

This survey was taken a week after a statistics exam, and the participants were students who participated in that exam. The fact that there was an exam in the week prior to the survey could possibly affect the previous estimates and this comparison. Students who reported playing daily or weekly generally followed through with significant gaming time during the week prior to the survey, seen by their higher point estimates and the shape of their time distributions. This suggests a strong alignment between reported gaming habits and actual behavior for these groups. On the otherhand, the majority of students who reported playing monthly or semesterly did not play video games during the week prior to the survey. Their self-reported frequency of playing (monthly or semesterly) may correctly reflect their long-term habits, however the timing of the survey (after the exam) may have influenced their decision not to play games during that specific week, which could explain the low percentages of gameplay in those groups. This could also explain why the overall proportion of students who played games in the past week was lower than expected. Other variables in the survey like “Hate Math”, “Grade Expected”, or “Play if Busy” may be confounding variables that

affect the results of our estimation. For example, those that respond “Yes” to the question “Hate Math” may dedicate less time to studying and more time to playing games. Similarly, those that already have high expectations for what grade they would earn on the upcoming exam might spend less time studying and more time playing games. Those that respond “Yes” to “Play if Busy” may still decide to spend time playing games despite being busy studying for the exam.

2.3 Point and Interval Estimate for Average Time Spent Playing Video Games

2.3.1 Methods

To estimate the average time spent playing video games during the week prior to the survey, we will calculate both a point estimate and a 95% confidence interval for the mean.

The point estimate for the average time is the sample mean of the time spent playing video games, excluding those who reported zero hours of gameplay. The formula for the sample mean \bar{X} is:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

where X_i is the number of hours played by the i th student and n is the total number of students in the sample.

We explored the shape of the data using a histogram to visualize the distribution of the reported time spent. Given that the data appeared heavily skewed, we chose to use bootstrapping, a resampling method, to estimate a 95% confidence interval for the mean. Since the sample distribution is not normal, bootstrapping is useful because it does not require the assumption of normality.

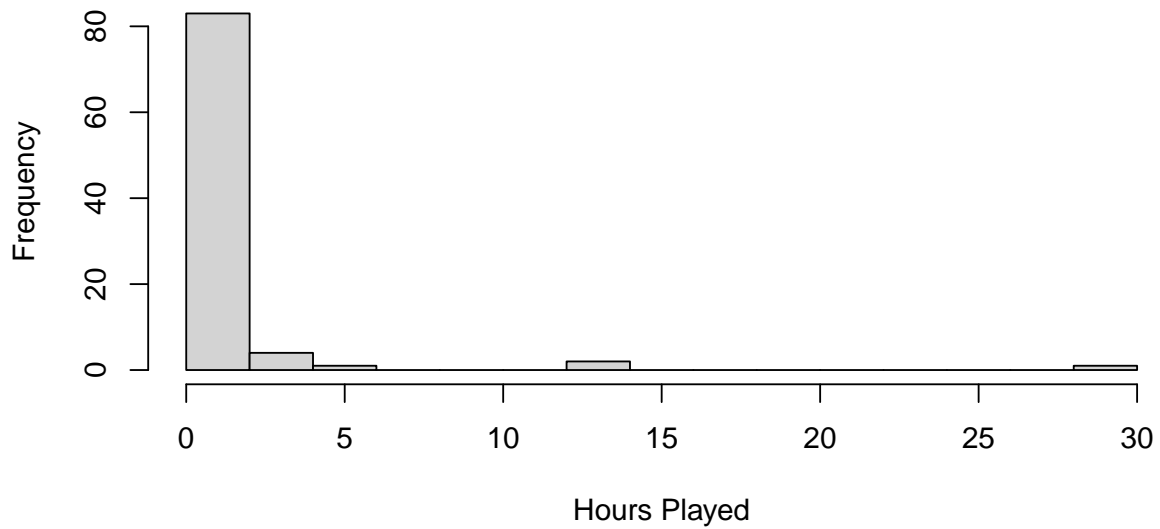
The procedure for bootstrapping involved resampling the data (with replacement) 10,000 times and calculating the mean for each resample. This allowed us to generate a distribution of means and compute the point estimate (the sample mean) and the 95% confidence interval using the 2.5th and 97.5th percentiles of the bootstrap distribution.

2.3.2 Analysis

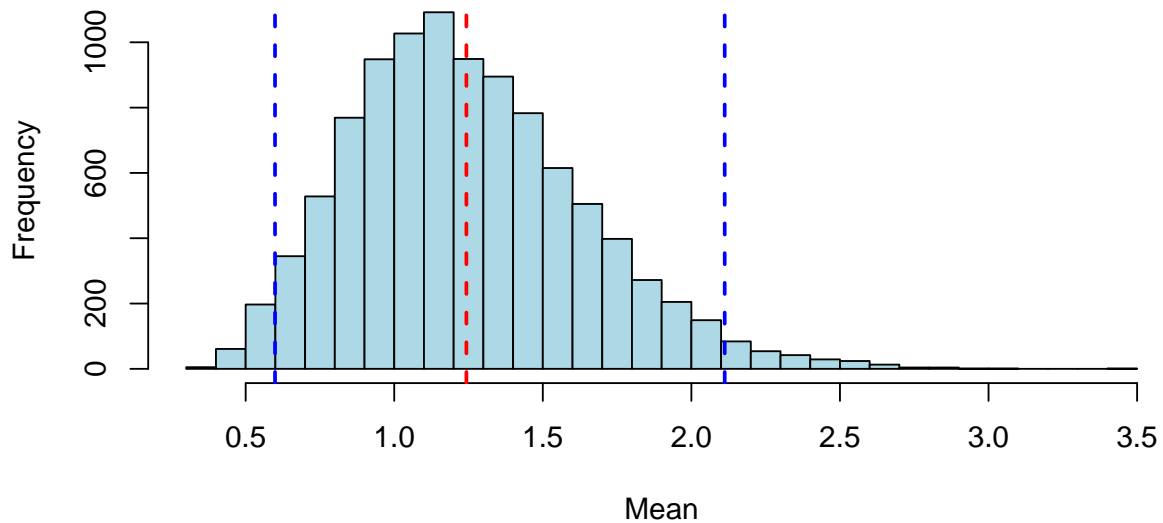
The point estimate for the average amount of time spent playing video games is 1.24 hours. To account for the right-skewed distribution, we generated a 95% confidence interval through bootstrapping. The resulting interval ranges from 0.60 hours to 2.11 hours.

The histogram of the bootstrap distribution, shown below, shows an approximately normal sampling distribution of the mean, despite the skewness of the original data due to the large sampling size (CLT), suggesting the bootstrap method was appropriate for generating the interval estimate. The red dashed line in the histogram represents the point estimate, while the blue dashed lines represent the boundaries of the 95% confidence interval.

Time Spent Playing Games a Week Prior to The Survey



Bootstrap Distribution of the Mean



2.3.3 Conclusions

The average amount of time students spent playing video games in the week prior to the survey is estimated to be 1.24 hours. The 95% confidence interval suggests that, on average, students spent between 0.60 and 2.11 hours playing video games during that week. This interval accounts for the skewed nature of the data and provides a more robust estimate

than taking two standard errors away from the mean, due to the non-normality of the distribution.

The simulation study supports the validity of bootstrapping to determine the interval estimate, as the bootstrap distribution of the mean follows a roughly normal shape, proving to be an effective way to estimate the average time spent on video games in this dataset.

To conclude, we estimate that students who played video games spent an average of about 1.24 hours playing games in the week prior to the survey, with the true mean likely falling between 1.48 and 2.46 hours.

2.4 Attitude Toward Video Games

2.4.1 Methods

To analyze the students' attitudes towards video games, we will focus on the responses to the questions assessing how much students like playing video games, the reasons for playing, and the reasons for disliking them.

The variable **play** in the dataset categorizes students' preferences for video games as follows: 1 = never played, 2 = very much, 3 = somewhat, 4 = not really, and 5 = not at all.

We will assess whether students generally enjoy playing video games by analyzing the distribution of responses. We will examine the reasons why students like or dislike video games by focusing on the most frequently mentioned reasons.

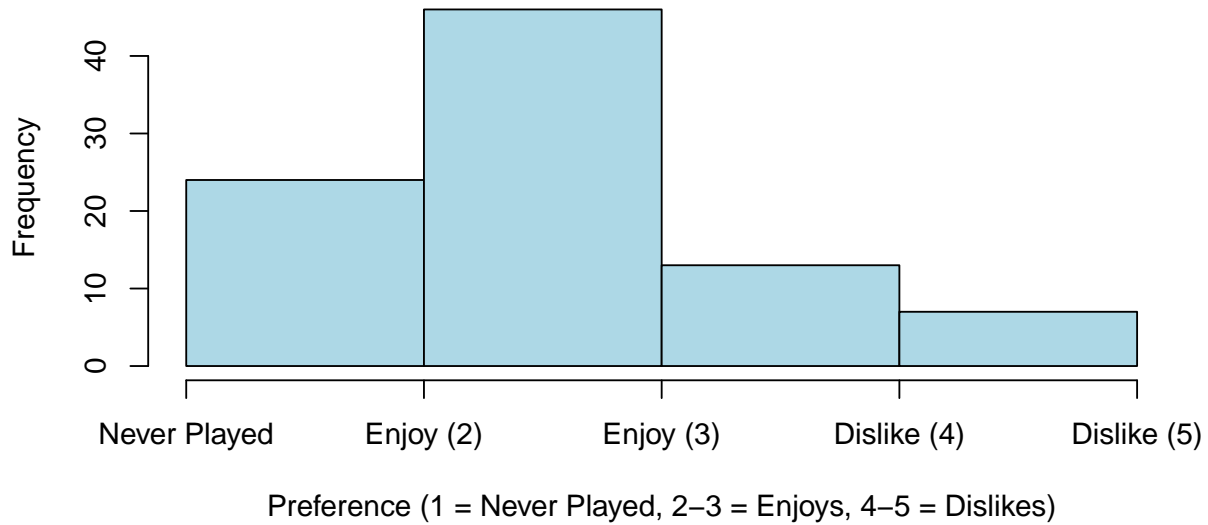
Since some students were asked to skip certain questions (e.g., those who never played video games), we will account for nonresponses when interpreting the results. This includes the proportion of nonrespondents for questions such as whether video games are educational and where they play.

We will create two lists based on the most frequently selected responses for reasons why students like or dislike video games: Reasons for Liking Video Games; Reasons for Disliking Video Games.

2.4.2 Analysis

The distribution of students' preferences for video games using the **play** variable is shown below.

Distribution of Video Game Preferences

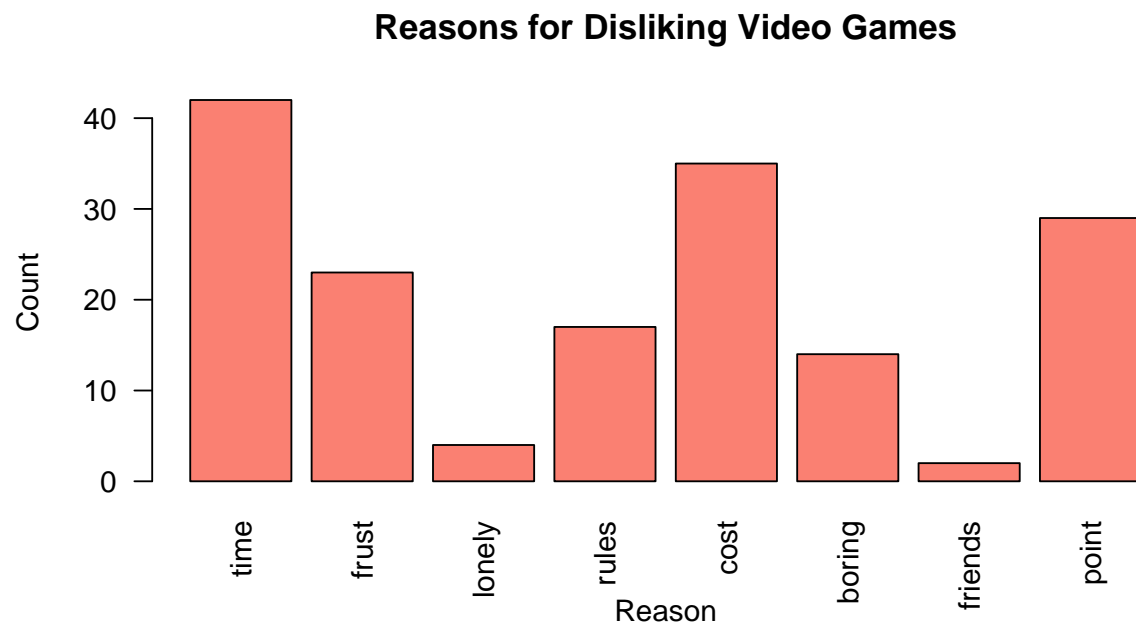
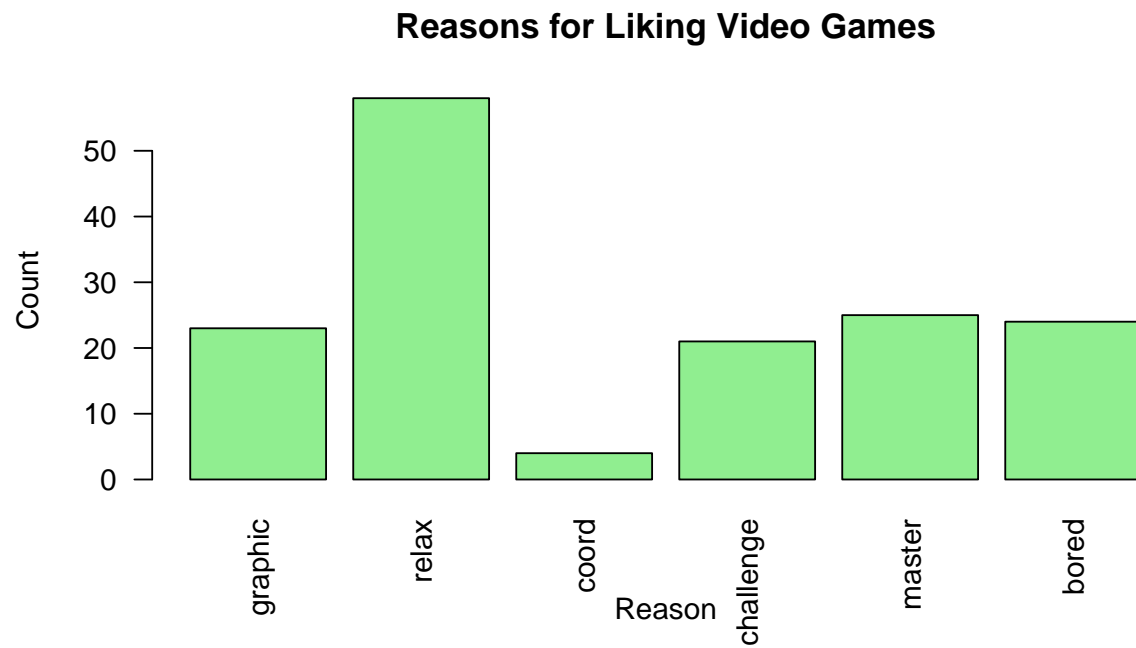


About 75.82% of students reported that they enjoy video games (combining responses for “like” values 2 and 3). 21.98% of students reported that they dislike video games (responses for “like” values 4 and 5). Only 1.10% of students reported that they have never played video games.

Table 2: Proportion of Students’ Preferences for Video Games

	Proportion
Enjoys Video Games	75.82%
Dislikes Video Games	21.98%
Never Played Video Games	1.10%

Next, we examined the most common reasons for liking or disliking video games from the second part of the survey.



Reasons for Liking Video Games (most frequently selected):

1. Relaxation
2. Feeling of Mastery
3. Bored
4. Graphics/Realism

Reasons for Disliking Video Games:

1. Too much time
2. Costs too much
3. It is pointless
4. Frustrating

2.4.3 Conclusion

Based on the analysis of the attitude questions, we find that a majority of students enjoy playing video games to some extent. Approximately 76% of students reported liking video games either very much or somewhat, while about 22% of students reported disliking them or not playing at all. This suggests that video games are generally popular among students, with a significant majority finding enjoyment in them.

The most common reasons students provided for liking video games include:

1. Relaxation - The highest count of students cited relaxation as a reason for liking video games.
2. Feeling of Mastery - Many students enjoy the sense of accomplishment and mastery over the challenges in video games.
3. Bored - Some students use video games as a way to relieve boredom.
4. Graphics/Realism - The quality of graphics and visual appeal was also a common reason for liking video games.

The most frequent responses for disliking video games are:

1. Too much time - The most common reason for disliking video games was the amount of time they consume.
2. Costs too much - The expense of gaming was a significant deterrent.
3. It is pointless - Some students view video games as having no real purpose or point.
4. Frustrating - Certain students found video games frustrating, which reduced their enjoyment.

It is noted that students who have never played video games or expressed strong dislike for them were instructed to skip certain questions, which likely explains the presence of nonrespondents for some of the attitude questions. Consequently, there may be a higher number of non-responses in questions related to other aspects of video games (like whether they are educational or where they play). The analysis of those specific questions would primarily reflect the views of students who already have a positive opinion of video games.

Overall, students generally enjoy playing video games, with relaxation being the most common reason for liking them and time commitment being the most common reason for disliking them.

2.5 Differences Between Students Who Like and Dislike Video Games

2.5.1 Methods

We will examine the differences between students who enjoy playing video games versus those who don't. We use survey data to examine differences based on three demographic variables:

1. Gender (Male vs. Female)
2. Work Status (Students who work for pay vs. those who don't)
3. Computer Ownership (Students who own a computer vs. those who don't)

To simplify these comparisons, we will collapse the responses to video game preference into two main categories:

- Enjoys Video Games: This includes students who selected “2” (very much) and “3” (somewhat).
- Dislikes Video Games: This includes students who selected “4” (not really) and “5” (not at all).

The students who selected “1” (never played) will not be included in this analysis. We will compute proportions for each demographic group (e.g. male/female, working/not working, owning a computer/not owning a computer) to observe differences in video game preferences. Graphical displays (i.e. bar plots and cross-tabulations) will be used to visualize the differences between the subgroups to help us understand whether demographic factors like gender, work status, or computer ownership are associated with a preference for video games.

2.5.2 Analysis

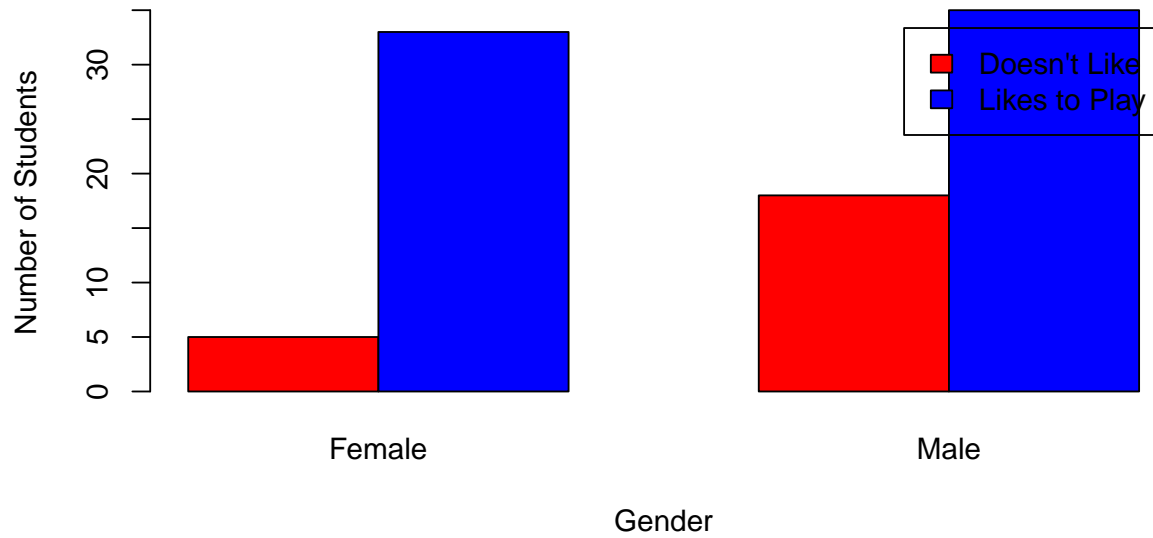
First, we categorized students based on whether they like to play video games or not:

Table 3: Sample Frequency of Video Game Preference

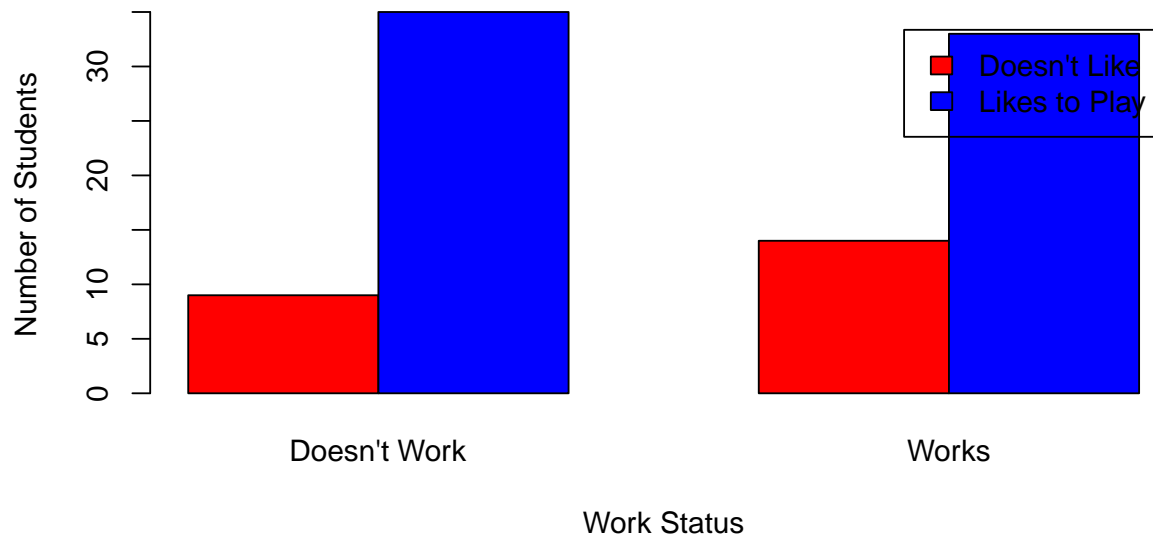
	Frequency	Proportion
Likes to Play Video Games	69	77.5%
Does not like to Play Video Games	20	22.5%

We compute the proportions of students in each category who enjoy or dislike video games and visualized the results using bar plots:

Video Game Preferences by Gender



Video Game Preferences by Work Status



Video Game Preferences by Computer Ownership

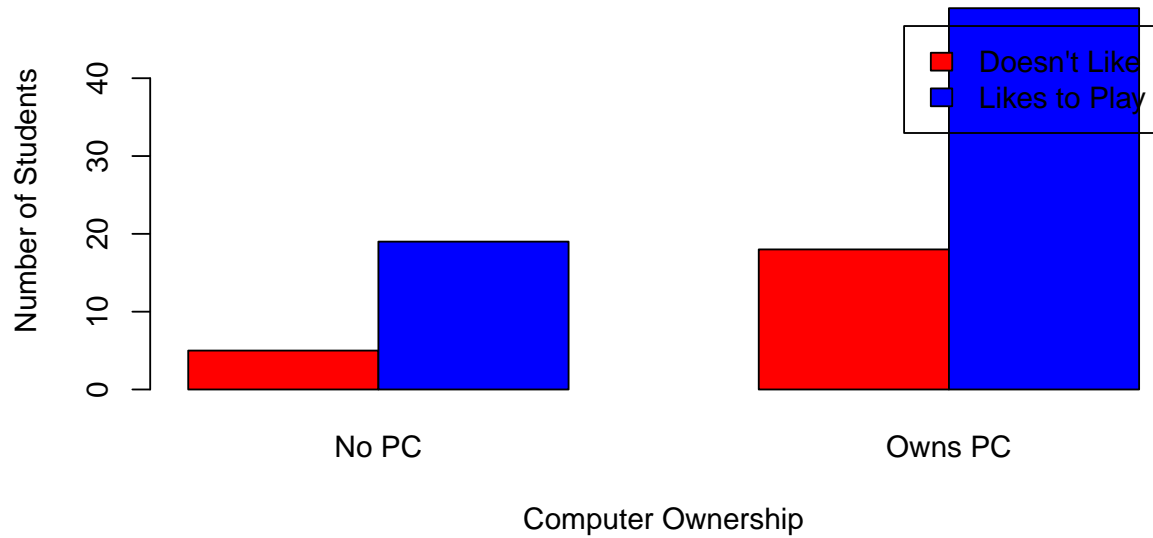


Table 4: Video Game Preferences by Gender

	Female	Male	Total
Likes to Play Video Games	0.363	0.385	0.747
Does not like to Play Video Games	0.055	0.198	0.253
Total	0.418	0.582	1.000

Table 5: Video Game Preferences by Work Status

	Works	Doesn.t.Work	Total
Likes to Play Video Games	0.362	0.385	0.747
Does not like to Play Video Games	0.154	0.099	0.253
Total	0.516	0.484	1.000

Table 6: Video Game Preferences by Computer Ownership

	Owns	Doesn.t.Own	Total
Likes to Play Video Games	0.538	0.209	0.747
Does not like to Play Video Games	0.198	0.055	0.253
Total	0.736	0.264	1.000

From the cross-tabulation and bar chart, we observe that male students are more likely to enjoy playing video games compared to female students. A larger proportion of male students report liking video games than female students. Students who work for pay tend to have a slightly lower preference for video games compared to those who don't work. Students who own a computer are more likely to enjoy playing video games than those who don't.

2.5.3 Conclusions

Our analysis reveals notable demographic differences between students who like to play video games and those who do not:

1. **Gender:** There is a no notable gender difference for students who like to play video games, with both male and female students being likely to enjoy video games. This suggests that video games may appeal equally to both genders. However, a greater proportion of males reported disliking video games, potentially due to the types of games commonly played by males.
2. **Work Status:** Similarly to gender, work status does not seem to have a significant effect on whether a student likes to play video games. Students who work for pay are relatively as likely to enjoy video games as those who do not work. However, there is a slightly greater proportion of students who work for pay and dislike video games; this could be due to the time constraints faced by working students, who may have less leisure time.
3. **Computer Ownership:** Computer ownership appears to have a major influence on whether a student likes to play video games or not. Owning a computer appears to be associated with a higher likelihood of enjoying video games. This is intuitive, as access to technology likely facilitates more frequent gaming, and in the reverse case, liking gaming may encourage the purchase for a computer. Out of the students who dislike gaming, there is, however, a larger proportion who own a computer.

This suggests that gender and work status may not be related to students' preferred for video games, while computer ownership appears closely related. Male students, those who do not work, and students who own computers are more likely to enjoy playing video games compared to their counterparts. However, in the population of those who are more likely to dislike playing video games, male students, those who do work, and those who own a laptop take precedent.

2.6 Grade Distributions

2.6.1 Methods

To analyze students' grade expectations in the course, we will: 1. Calculate the observed distribution of students' expected grades by determining the proportions of students expecting each grade with the letter assignments: 4 = A, 3 = B, 2 = C, 1 = D, 0 = F, 2. Compare

the observed distribution with the course's intended grade distribution (20% A's, 30% B's, 40% C's, and 10% D's or lower), 3. Simulate the effect of nonrespondents as potential failing students to evaluate if this assumption could significantly alter the observed distribution. Non-respondents could represent students who may expect to receive failing grades and, therefore, chose not to participate in the survey or attend discussion sections.

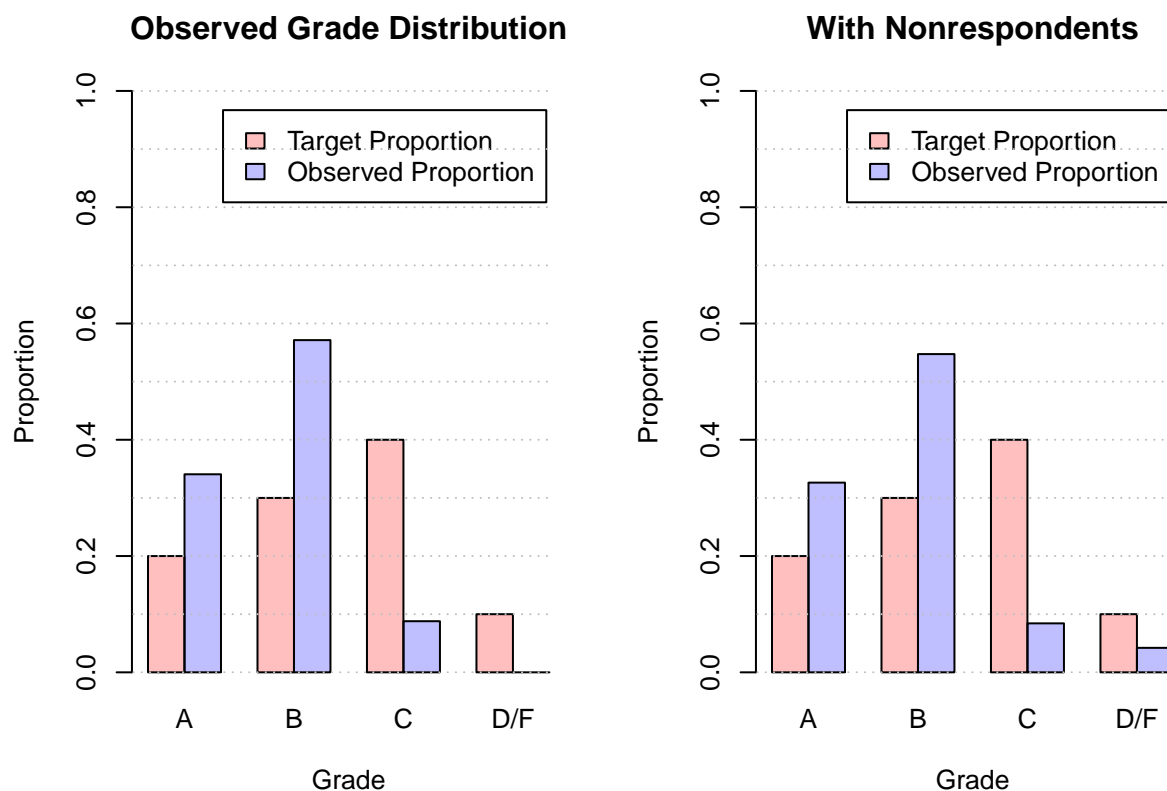
2.6.2 Analysis

The observed grade distribution from the survey data shows the following proportions:

Table 7: Comparison of Observed and Target Grade Proportions

Grade	Observed	With.Nonrespondents	Target
A	0.3406593	0.3263158	0.2
B	0.5714286	0.5473684	0.3
C	0.0879121	0.0842105	0.4
D/F	NA	0.0421053	0.1

By including the 4 non-respondents in the D/F category, we recalculated the proportions to evaluate how it would change the observed distribution's alignment with the target distribution. The bar plots below compares these observed proportions with the target distribution set by the course. Deviations from the target distribution suggests discrepancies between students' expectations and the course's grading goals.



Comparing the observed grade distribution with the target, the observed proportion of students expecting A's is 34.1%, which exceeds the target of 20%. The observed proportion for B's is 57.1% which is much higher than the target 30%. Only 8.7% of students expect a C, which is substantially lower than the target of 40%. There were no respondents categorized as D/F students in the initial observation, compared to the target of 10%.

When non-respondents are included as D/F students (assuming these students are no longer attending and likely failing), the proportion of students expecting A's and B's drops slightly to 32.6% and 54.7%, respectively. These values still exceed the course's target of 20% for A's and 30% for B's, though the gap narrows. The proportion of students expecting C's changes only slightly to 8.4%, which remains far below the 40% target. The D/F category now shows 4.2% of students, compared to the 10% target. Even with the non-respondents accounted for, the proportion of failing students remains lower than expected.

2.6.3 Conclusion

The observed grade distribution indicates that a larger proportion of students are expecting higher grades (A's and B's) than the target distribution set by the course. There is a significant deficit in the proportion of C's, which suggests that students are either expecting to over perform relative to expectations, lenient grading policies, or simple course material.

The inclusion of non-respondents, under the assumption that they are failing students, slightly adjusts the distribution toward the course's targets. However, it still does not bring

the observed distribution in alignment with the target proportions. Even with this adjustment, there are fewer failing students than expected, and higher grades are over-represented.

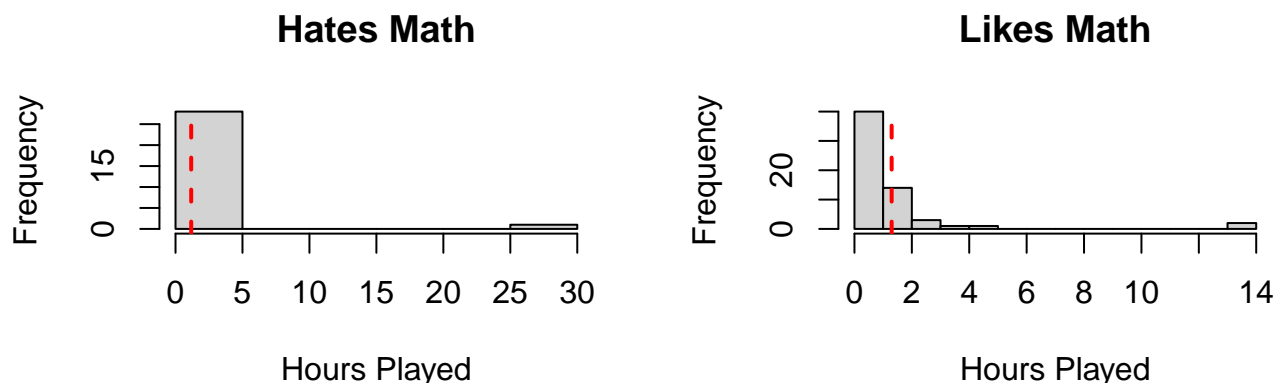
This discrepancy suggests that students' expected performance diverge from the course's grading goals. Course content, grading policies, or other factors (such as grade inflation or differing levels of student effort) could potentially cause this misalignment.

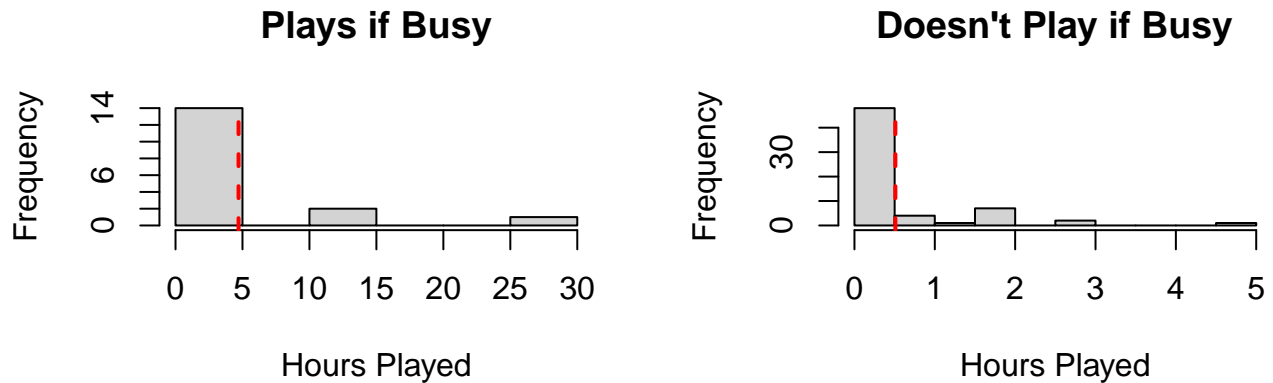
3. Advanced Analysis

3.1 Methods

We mentioned two variables “Hate Math” and “Play if Busy” that may affect the results of the estimation of the average amount of time spent playing video games in the week prior to the survey. Now, we want to further explore these variables by comparing the average amount of time spent playing video games in the week prior to the survey between students who reported hating math and those who reported that they did not hate math as well as the mean hours between students that reported playing video games even when they are busy and those who do not. To do this, we conducted two permutation tests to find out if there exists a statistically significant difference in the means between the groups.

First, we visualized the distributions of time spent playing video games of each group and calculated the average time spent playing games within each group (represented by the red dotted line).





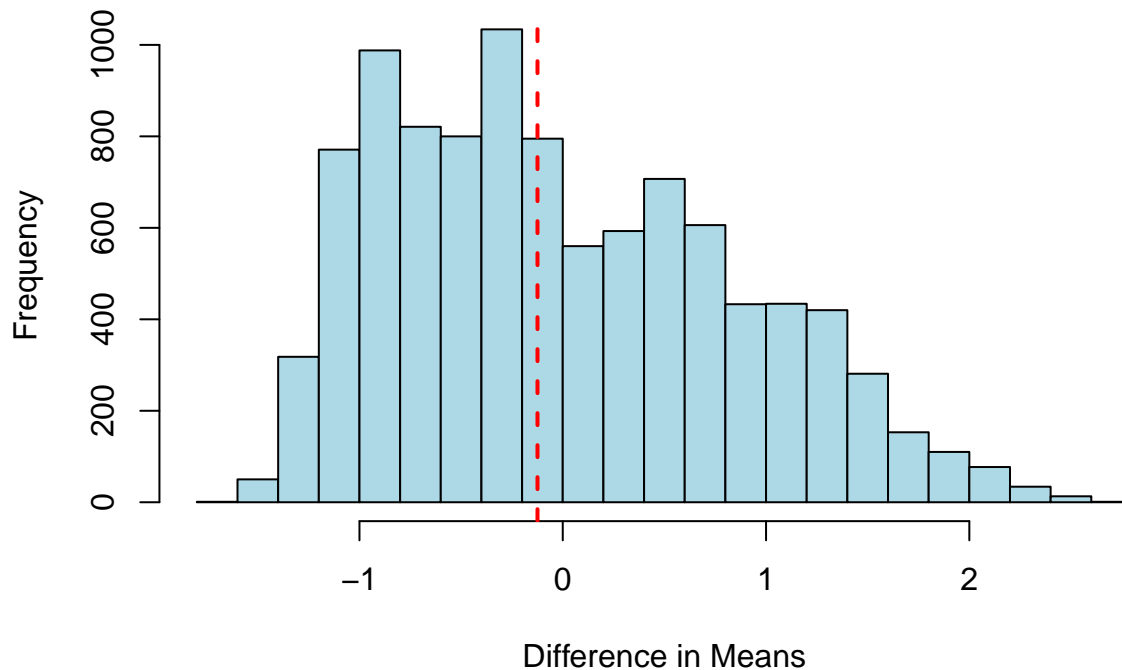
3.2 Analysis

Hypotheses

- **Null Hypothesis (H0):** There is no difference in the average number of hours played between students who hate math and those who do not. Any observed difference is due to random chance.
- **Alternative Hypothesis (H1):** There is a significant difference in the average number of hours played between students who hate math and those who do not.
- **Test Statistic:** Difference of means
- **Significance Level:** 0.0001
- **p-value:** The p-value for the t-test was **0.4905**, which is far above the 0.0001 significance level, which means we fail to reject the null hypothesis.

The results of the permutation test between students who reported hating math and students who reported not hating math is shown below. The dotted red line represents the observed difference between the average amount of time spent playing video games for students who reported hating math and students who reported not hating math.

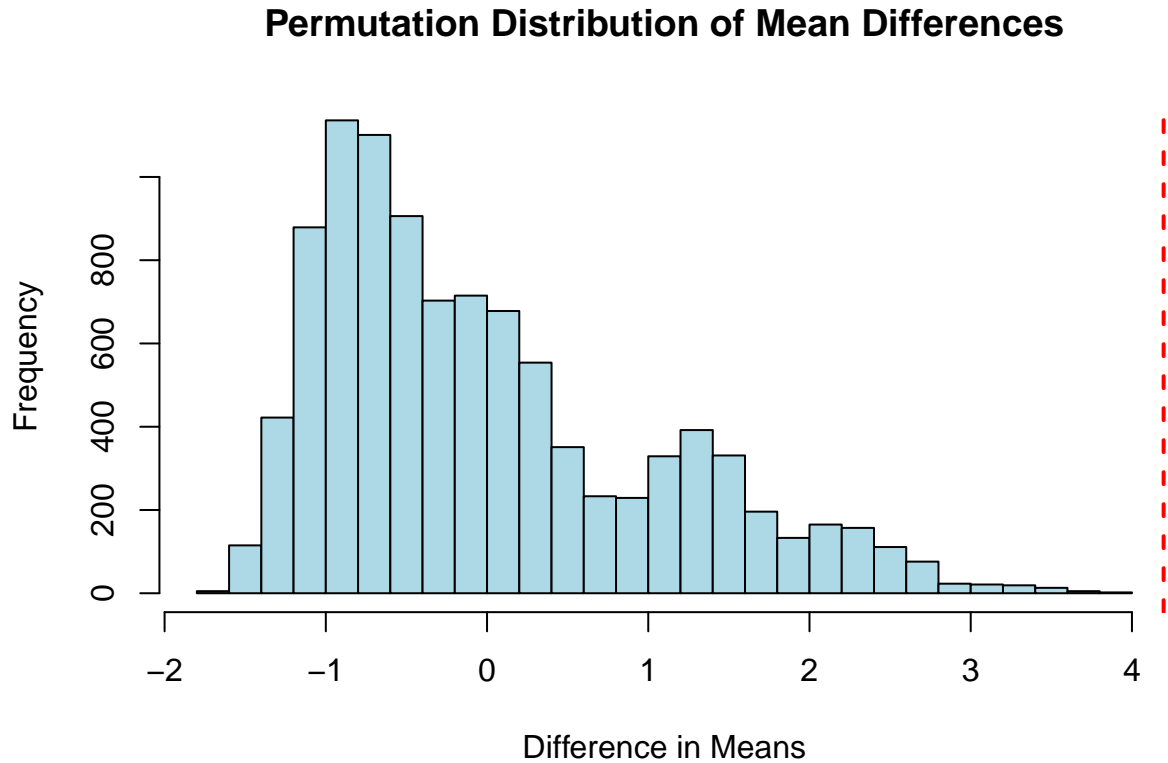
Permutation Distribution of Mean Differences



Hypotheses

- **Null Hypothesis (H0):** There is no difference in the average number of hours played between students who play video games when busy and those who do not. Any observed difference is due to random chance.
- **Alternative Hypothesis (H1):** There is a significant difference in the average number of hours played between students who play video games when busy and those who do not.
- **Test Statistic:** Difference of means
- **Significance Level:** 0.0001
- **p-value:** The p-value for the t-test was **0**, which is below the 0.0001 significance level, which means we reject the null hypothesis.

The results of the permutation test between students who reported hating math and students who reported not hating math is shown below. The dotted red line represents the observed difference between the average amount of time spent playing video games for students who play video games even when they are busy and students who do not.



3.3 Conclusion

In conclusion, the results of the first permutation test suggest that there is no statistically significant difference in the average hours played between students who hate math and those who do not. However, the results of the second permutation test indicate that students who play video games when they are busy tend to play significantly more hours in a week compared to students who do not play when busy, which aligns with our inference that “Play If Busy” could be a confounding variable that affects the proportion and mean estimations from Sections 2.1 and 2.2.

4. Discussion

5. Appendix