Rick Boelen

Hwang Min Yu

Wen Zeng

MATH 189

04/25/2020

**Homework 2 Data Analysis Report**

# 

# Introduction

This study is primarily focused on gathering and interpreting data considering what populations play video games. The surveyed population is the undergraduates in the Introductory Probability and Statistics class, Section I, of Fall 1994.

The students that were selected were among those that had taken the second exam of that term. The students that were sampled from those that had taken the second exam of that term. The students were assigned a number from 1 to 314, and a pseudo-random number generator selected 95 students to participate in the study. Student anonymity was preserved to encourage honest responses. Some of the individuals from the sample did not answer part of the survey, and so we decided to omit those observations for the corresponding analysis. In order to minimize nonrespondents, data collectors first visited the discussion sections for lecture I in the week the survey was collected. In order to encourage attendance, exam papers were returned in discussion that week. Students that had not attended discussion were located in lecture to complete the survey. Data collectors were asked to inform students of the purpose of the survey and to guarantee student anonymity in order to encourage accuracy in reporting.

Our analysis will help provide insight on the preferences students have with video games and the supposed association this has with the labs the college was deciding to design. With random sampling, this sample can be validated to some extent to be a representative sample of the population, which would be future student bodies from the statistics class. One thing to consider is the change in preferences with students, as time moves on and how long this data will hold reliable for. Moreover, the survey was taken a week before their exams as explained above. This could potentially show large changes in the results due to the students’ shift in priorities.

Framing issues with the data itself is that this could be regarded as a small sample considering the diversity of topics video games bring. Therefore, we incorporated a bootstrapping method to ensure further reliability of our data. Moreover, it is hard to find viable links to video games and labs because of them being two different entities, so our assumption for now is that findings for video games have some possibility of being regarded as findings for the design of labs. Finally, the students were the ones filling out this form and with the flexibility of the choices, there were some restrictions that were violated, this made the data more difficult to work with and we had to omit missing values or make assumptions about the data.

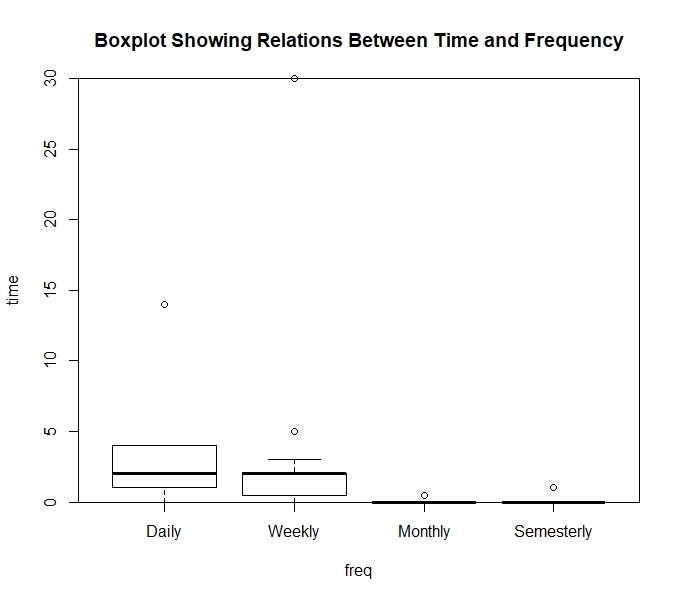
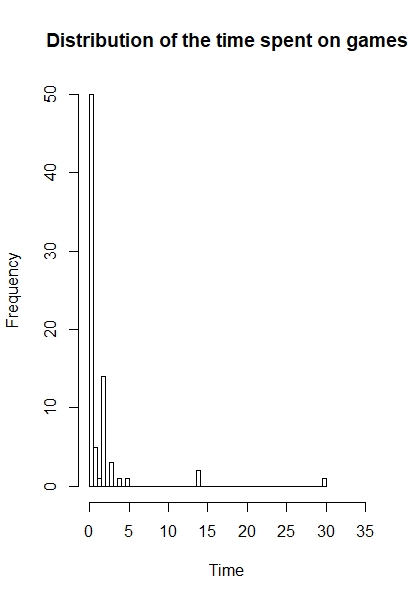
# Methods/Analysis

*Scenario 1:* Estimation of fraction of students who played a video game in the week prior to the survey.

*Method*: We first isolated the respondents who had reported time played as more than 0 hours in the previous week into a new table, “gamers”. By dividing the number of gamers by the number of respondents, we get a point estimate of 0.3736264. Then, to get an interval estimate we employ bootstrapping to sample 400 times from our sample. This gives us an interval of (0.2885014, 0.4587514) for a 95% confidence interval.

*Analysis*: By using the point estimate of 0.37, we can estimate with 95% confidence that our confidence level of 28.85% and 45.88% will hold the population true value, which is the fraction of students in the class that plays video games.

*Scenario 2:* Amount of time spent playing in the previous week compared to reported frequency of play.

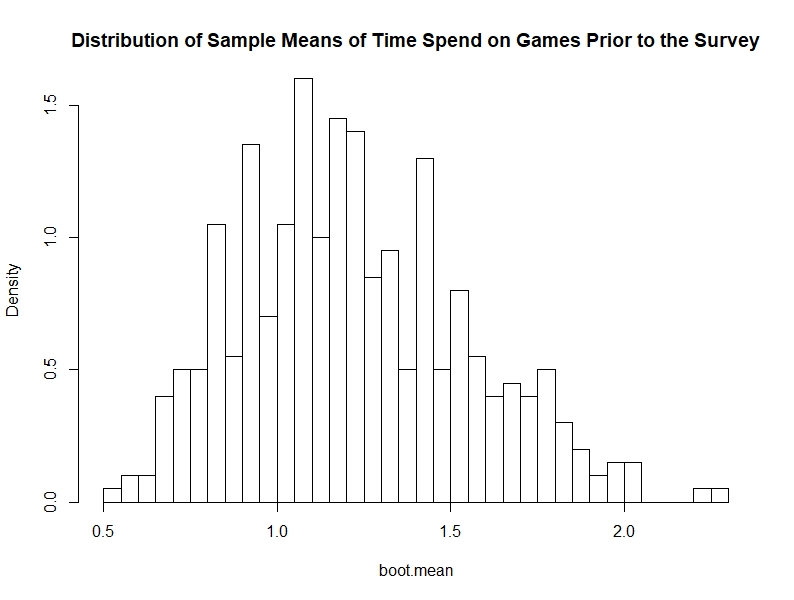
*Method:* We plotted a boxplot relating the time played during this past week with the frequency of students, and then we plotted a histogram showing the distribution of time played this week with frequency on a regular basis. 

*Analysis:* The weekly box plot shows a general increase in median as the frequency of their regular game habits increase. However, their time spent is observed to be in the range of 0 to 5 for the majority of students, disregarding the outliers. This can be shown with the distribution on the left being skewed slightly to the right. Moreover, we can see that the median is observed to be 0 from the graph for students who answered monthly or semesterly.

The main findings of this relation is that the majority of students are reducing the amount of time spent gaming or not playing at all due to the upcoming exam, except for a few outliers. We have to take these assumptions into account and into play with scenario 1 as it holds a lot of value understanding the context. Given the context, it was harder to find a noticeably large difference between frequency and their time spent in the past week.

*Scenario 3:* Average amount of time spent playing in the previous week. Analyze overall shape of dist. Simulation study to estimate the appropriateness of the interval estimate.

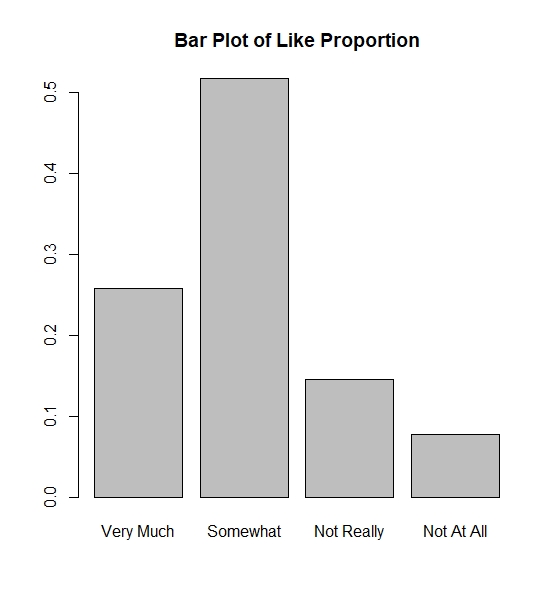
*Method*: We used bootstrapping to sample means 400 times from our original sample. Then, we plotted the distribution of the sample means on a histogram relating the bootstrapped mean with its density. The mean of this distribution is 1.242857. Then, we calculated the 95% confidence interval estimate to be (0.6035585, 1.8821558).



*Analysis*: The distribution of the sample means is roughly normal with a slight skew to the right. The hours played range from 0.5 hours to 2.6 hours played prior to the survey, with a mean of 1.24 hours. We can estimate with 95% confidence that our confidence interval of 0.6 hours and 1.88 hours will hold the true population value, the mean population of students in the class spending time playing video games . Since the sample size is large and the distribution is approximately normal, we can conclude that our original distribution is independent and random. Therefore, our estimate is appropriate for this study.

*Scenario 4:* Attitude: Do students like video games?

*Method*: We used a bar plot and proportion tables to analyze whether the students like gaming or not. In the bar plot, we omitted those who never played games because those people do not tell us whether they like games or not. In the proportion table, we omitted the missing values in the video. Multiple dataset to not assume results in parts of the survey.



*Analysis*: By looking at the bar plot, we could see that there is a higher proportion of students that like playing games, if we split the groups into two. The most popular reasons that students enjoyed were relaxation, feeling of mastery and being bored. (1.1) These are characteristics of video games that could be incorporated into the labs making labs intuitive yet challenging. On the other hand, popular reasons why students disliked video games were too much time, costs too much and it being pointless.(1.1) Moreover, the other section of the follow up survey allowed students to voice other reasons, 9% of students filled this out for favorable reasons, while 10% filled it for unfavorable reasons. Although a small difference, this could show another perspective where people have more to say when it comes to dislike reasons towards video games.

Some aspects of this follow-up survey to keep in mind is that those who disliked video games were asked or chose to skip the questions for liking video games and vice versa. This could create bias based on looking at findings from groups within the samples. Furthermore, this was answered in close succession to their exams, and could lead to a biased increase of students choosing relaxation seeing that there was over 66% in the follow-up survey results. (1.1) Moreover, some students chose not to uphold the restriction of choosing 3 choices and chose more than 3 choices, however, this data was still utilised in our analysis.

*Scenario 5:* Differences between those who like and dislike games.

*Method:* We plotted double bar graphs of the data to compare those who like video games and those who dislike them based on pay, gender, and owning a PC, and then we used the proportions split into the specified categories to have a relative comparison between one another when analyzing the profile of gamers.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Proportion Table of Liking Games VS Gender**   |  |  |  | | --- | --- | --- | |  | Female | Male | | Very Much | 13.157895% | 35.294118% | | Somewhat | 55.263158% | 49.019608% | | Not Really | 21.052632% | 9.803922% | | Not At All | 10.526316% | 5.882353% | | **Proportion Table of Liking Games VS Paying**   |  |  |  | | --- | --- | --- | |  | Not Paid | Paid | | Very Much | 20.454545% | 33.333333% | | Somewhat | 47.727273% | 52.380952% | | Not Really | 22.727273% | 7.142857% | | Not At All | 9.090909% | 7.142857% | | **Proportion Table of Liking Games VS Owning a PC**   |  |  |  | | --- | --- | --- | |  | Not Own | Own | | Very Much | 20.833333% | 27.692308% | | Somewhat | 66.666667% | 46.153846% | | Not Really | 4.166667% | 18.461538% | | Not At All | 8.333333% | 7.692308% | |

*Analysis:*

The bar plots helped enable us to understand the population of people who answered the sample. For example, it is clear to see from the observed graph that the majority of students owned a PC. (1.2) However, proportions were needed to analyze the relative difference between these categories.

Working: The proportion tables show a difference of 12.88% when it comes to looking at students who ‘Very Much’ liked video games out of the groups, in favor of those who worked. The opposite is stemmed for those who don’t like it ‘at all’. This shows that those who work are more inclined to enjoy video games, and this makes sense when looking at scenario 4 and their reasons, such as relaxation and the costs of time.

PC ownership: The responses within the categories are quite varied and hard to make any strong claims from. Therefore, this bar graph does not supposedly suggest any correlation between PC ownership and love for video games.

Gender: Grouping the top two like categories together, 84% of males enjoy video games compared to 68% for females. This difference could show that typically male tend to enjoy video games.

This helps engage the profile of students we are trying to target when designing labs and to understand the important factors, such as access to PC and time.

*Additional Analysis:* Is video games/labs a viable method to improve students who are struggling in the class.

*Method:* We started to observe the relation between the student’s expectation for their grade and how often they play video games.

Covariance: -0.1404

*Analysis*:

Looking at the barplots, one could interpret that there are many people who enjoy video games and therefore this could correlate with many students enjoying the design of labs. (1.4) However, it would be better to observe those who need an alternative method to learn, those with lower grade expectations. First, we utilized the covariance, a measure of joint variability between two variables, over grade expectations and their usual frequency play of video games. A negative value helps to observe a general negative relation between one another. Therefore, we could make a general claim that the lower the grade expectation, the higher the amount of games played. This validates the idea of designing labs that are tailored towards characteristics of a video game.

Further analysis showed ,however, that the mean GPA expected for this class for those who enjoy playing video games and those who do not are 3.280488 and 3.0, respectively. Therefore, the distribution of expected GPA for those who dislike video games is symmetrical. (1.4)

# Conclusion

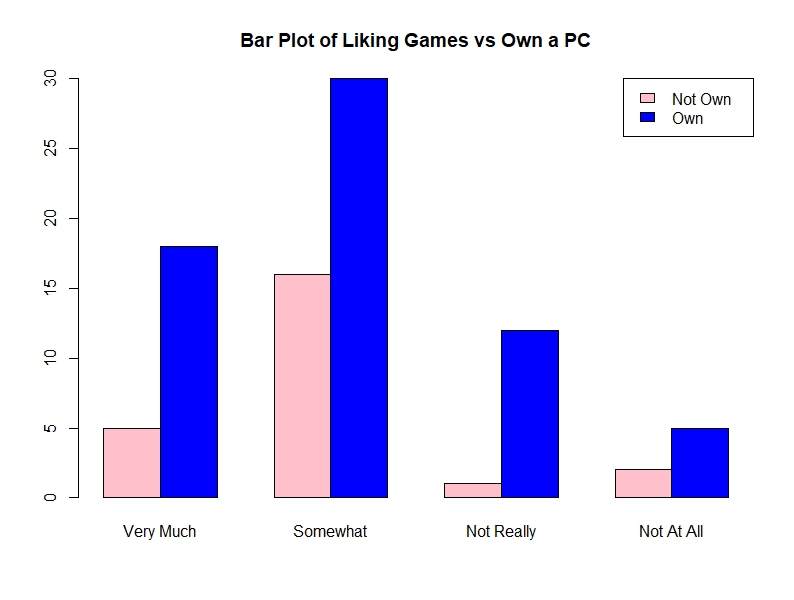
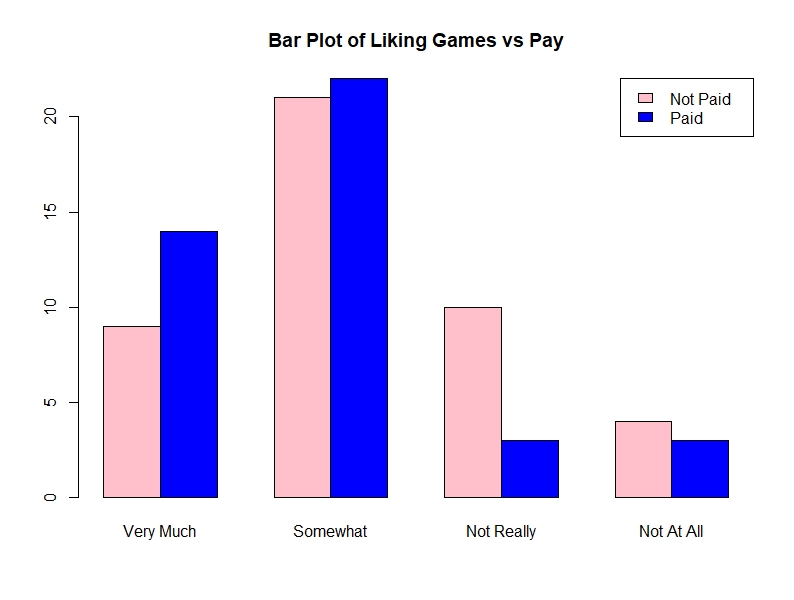
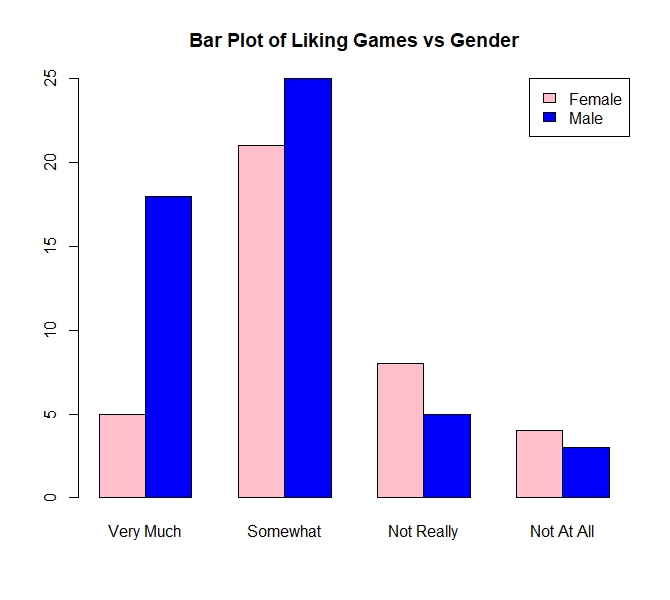
Our analysis was a study on video games in order to gain further insight on how the design of labs for a statistics class can help as an alternative method to learning. Our analysis showed that a week before exams that only 37% of students played video games prior to their studies. Therefore, this demonstrates a sense of priorities and how the labs had to involve the content of the class more heavily. We also looked into characteristics of video games that appeal to students and how making the labs relaxing, yet progression-based, more appealing to a group of statistics students. Finally, we looked into profiling who the labs would be useful for, and our best estimate would be towards an audience who is working, and potentially have access to a PC because of the nature of the student body. Another finding we found interesting was that more students selected 3 choices when looking at reasons why they enjoyed video games, noticing an increase in voicing their opinion. If we were to further investigate this topic, it would be critical to have a greater foundational understanding of the association of video games between labs. Is it even viable to connect the two as the entities with similar characteristics? Furthermore, feedback on an alpha version of the lab is more crucial as it gains insight from everyone, looking back on this survey there were many students who have never played video games and could not give us insightful information.

# Appendix

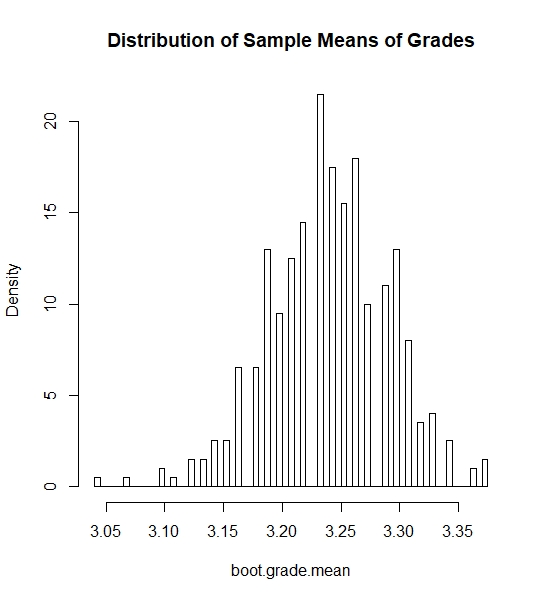
1.1 Proportions based on choices students chose on reasons of enjoyment or dislike for video games.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Proportion Table of “Why?” Portion in**  **Survey**   |  |  | | --- | --- | | **Why?** | **Proportion** | | Relaxation | 66.666667% | | Hand/Eye Coordination | 4.597701% | | Mental Challenge | 24.137931% | | Feeling of Mastery | 28.735632% | | Bored | 27.586207% | | Graphics/Realism | 26.436782% | | Other | 9.195402% | | **Proportion Table of “Dislike” Portion in**  **Survey**   |  |  | | --- | --- | | **Dislike** | **Proportion** | | Too Much Time | 48.275862% | | Frustrating | 26.436782% | | Lonely | 4.597701% | | Too Many Rules | 19.540230% | | Costs Too Much | 40.229885% | | Boring | 16.091954% | | Friends Don’t Play | 2.298851% | | It is Pointless | 33.333333% | | Other | 10.344828 % | |

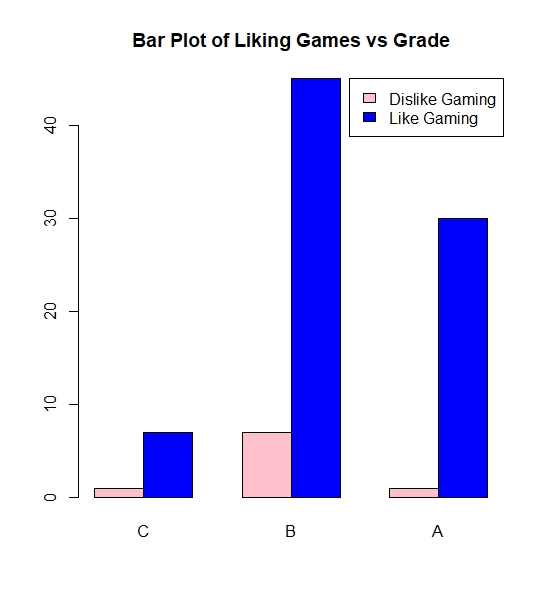
1.2 Bar plots showing a count frequency between the like categories and categories, work, gender, and owning a PC.



1.3 Distribution of sample means of grades



1.4 Bar Plot of analyzing grades and their like for gaming.

Contribution Statement

Introduction: Rick

Analysis Writeup: Wen

Analysis: Everyone

Conclusion: Rick

Code: Hwang