**Case Study 2: Statistics Lab vs. Video Games**

Dukki Hong (A98058412)

Karl Rummler (A12405136)

Nick Roberts (A11705541)

John Diez (A14751991)

Yiwen Li (A13959913)

## Introduction

The exponential growth of a digitalized society highlights why playing video games has become an integral part of households around the United States. A total of 63% of US households have at least one person who plays video games approximately 3 - 4 times per week (Fuller, 2017). Likewise, the video game market in the United States is valued at 18.4 billion USD, with some of the value coming directly from the increasingly popular esports scene (Fuller, 2017). Video games come into many people's lives, even starting when they are younger as young as 3 years old. Playing video games can add flavors into life, but sometimes, addiction into video games may cause other issues. With just this information alone, there is no denying that people's attraction to video games is a worldwide phenomenon; moreover, this has led people, especially in education, to try and find ways to implement it into the common curriculum.

In University of California Berkeley Statistics program, professors were interested in developing a new educational program/lab that mimics a video game in an  intro to statistics class. To do this, they developed a questionnaire that attempted to answer the question of why do people play video games. They sampled current statistic students with an array of questions, so the professors/statistics students could create a program that greatly mimics the positive reasons why people play video games.

In this paper, we will be discussing and analyzing the data they've collected and developing possible pathways and advice for how they should structure their lab. Specifically we'll be delving deeper with how attitudes of video games influence the rate of engagement or playing, how different genres of video games influence engagement/or likelihood for games, how gender and other participant characteristics are involved in the frequency or attitudes towards specific types of games, and finally, we'll also talk about possible issues or limitations that could stem from psychological issues of addiction, and ways we can limit or be aware of it in the design of the lab for an intro to statistics class.

**Background**

The first question that we need to answer is what are the reasons people play video games. A study surveying typically developing children found that some of the main reason they play games are for fun, excitement, boredom, competitiveness, relaxing, story, and exploring a new world. This was a combination of both quantitative and qualitative analyses from focus groups and surveys of children from the ages of 6-16. With these factors in mind, it's important to pay attention to how to apply these motivations into a lab program. To build it up, an online research study conducted on gamers found that players spend more time playing video games if they have the ability to work with other people, especially in cooperative style games. These game genres are ones like MMORPGS (massive online multiplayer games) and MOBA (6v6 battle game) (Johnson, Gardner & Sweetster, 2016).

An important factor of games is the genre in which it's categorized under. Most people are familiar with First Person Shooters (FPS) as it's one of the most popular games, but it only spans about 35% of the playerbase. A lot of games are multi-genre, but games like shooter games are typically classified as Action. With this understanding of multi-dimensionality of genre for games, we need to delve deeper into what genres of video games do people play, and which one do they actually prefer. In a study conducted by Lee & Colleagues (2007), Korean highschool and middle school students were recruited for the study and were asked to take a survey. Games were separated into 8 categories, but we'll only focus in 4: Action, role-playing, adventure, and simulation. Majority males (77%). What they found was that more people preferred simulation (36%) with the two runner ups being action and RPG (~16). However, one main finding they found was that participants also preferred games where they were able to play with others, or multiplayer games.

One factor that should always be considered when designing a program is the gender of the people who take it. Do females and men have a difference in their preference for video games, and what they like? In a study by Hartmann and Klimmt (2017), they conducted two major studying Gender differences in video game preferences. The first study was a pen-and-paper survey taken by females in different universities in Germany. Their results found that they didn't like games that demonstrated lack of meaningful social interaction, and too many violent content and sexual gender role stereotyping of game character. The second study conducted online surveys in different German Websites, and they compared differences in competitive aspects of different genres of games. They found that females are less attracted to competitive elements in video games compared to males. If we were to create a lab, we might want to consider making it less competitive and more cooperative if it's more female dominated. It is essential to consider different types of games based on gender, because studies have shown that type of game influence user preference and experience for different genders.

**Data Description**

Out of 314 students in Statistics 2, Section 1, during Fall 1994, 95 were selected at random to participate in the survey. Complete surveys were obtained from 91 out of 95 students. The data available here are the students' responses to the questionnaire. The survey asks students to identify how often they play video games and what they like and dislike about the games. The limitations of the data are that how we collected the data is dependent (we accounted for this in our analysis). Also this sample is only from undergraduate statistics students from UC Berkeley, leaving our interpretations to a larger scope at less confidence. Also, the study or questionnaire was created by upper-division college statistics students, where the questions asked might have not been psychologically accurate or valid for their measures of enjoyment.

**Missing Data:**

If a question was not answered or improperly answered, then it was coded as 99. Those respondents who had never played a video game or who did not at all like playing video games were asked to skip many of the questions. In this case, there are data missing issue in corresponding columns. Moreover, the second part of the survey that covers whether the student likes or dislikes playing games and why. These questions are different from the others in the more than one response may be given.

**Definition of Column Names in Data:**

| | | |
|---|---|---|
| Time | # of hours played in the week prior to survey | Numerical Data |
| Like to play | 1=never played, 2=very much, 3=somewhat, 4=not really, 5=not at all | Categorical Data |
| Where play | 1=arcade, 2=home system, 3=home computer, 4=arcade and either home computer of system, 5= home computer and system, 6=all three. | Categorical Data |
| How often | 1=daily, 2=weekly, 3=monthly, 4=semesterly | Categorical Data |
| Play if busy | 1=yes, 0=no | Categorical Data |
| Playing educational | 1=yes, 0=no | Categorical Data |
| Sex | 1=male, 0=female | Categorical Data |
| Age | Students age in years | Numerical Data |
| Computer at home | 1=yes, 0=no | Categorical Data |
| Hate math | 1=yes, 0=no | Categorical Data |
| Work | # of hours work the week prior to the survey | Numerical Data |
| Own PC | 1=yes, 0=no | Categorical Data |
| PS has CD-Rom | 1=yes, 0=no | Categorical Data |
| Have email | 1=yes, 0=no | Categorical Data |
| Grade expected | 4=A, 3=B, 2=C, 1=D, 0=F | Categorical Data |

***Table1**: Description of the dataset columns.*

**Methods**

When calculating the confidence intervals of the amount of time playing video games, we utilized both the approximately normal interval and the finitely corrected confidence intervals to adjust for the dependently sampled data. We also utilized the correct standard deviation for both the proportional and average calculations. To make sure that these intervals were valid, we did bootstrapping, created Q-Q plots and calculated both the kurtosis and skewness.

When bootstrapping the data, we modified our methods to adjust for the dependent data by sampling upwards towards 309 data points while maintaining the proportions of time played of the original 91 data points. Afterwards we sampled 91 data points from the 309 and calculated the mean for 1000 trials to create a bootstrapped distribution.

Additionally, in order to determine the important variables for determining whether a student likes playing video games or not, we used an information based decision tree to classify students on whether they liked video games or not. First we pruned the dataset to remove rows with many data points missing and trained the classifier on our data. The variables used in the decision tree when making decisions are the variables that provide the most information on predicting whether a student likes video games or not.

**Numerical Summaries**

Considering that the survey sample size is relatively small and the data comes from a specific course at an exact timeslot, we can say that the data are approximately independent and non-correlated to each other. Hence, we assume the data is independent, identically distributed(*i.i.d*).

Question: what types of games do you play?

| Type | Percent |
|---|---|
| Action | 50% |
| Adventure | 28% |
| Simulation | 17% |
| Sports | 39% |
| Strategy | 63% |

***Table2****: summarizes the types of games played*

Students who did answer this question were also asked to provide reasons why they play the games they do. They were asked to select up to three such reasons.

Question: Why do you play the games you checked above?

| Why? | Percent |
|---|---|
| Graphics/Realism | 26% |
| Relaxation | 66% |
| Eye/hand coordination | 5% |
| Mental Challenge | 24% |
| Feeling of Mastery | 28% |
| Bored | 27% |

***Table 3**: Summarizes reasons for playing the game*

Students who did answer this question were also asked to provide reasons why they play the games they do. They were asked to select up to three such reasons. Their responses are presented in Table 3.

Question: What don't you like about video game playing?

| Dislike | Percent |
|---|---|
| Too much time | 48% |
| Frustrating | 26% |
| Lonely | 6% |
| Too many rules | 19% |
| Costs too much | 40% |
| Boring | 17% |
| Friend's don't play | 17% |
| It is pointless | 33% |

***Table 4**: Summaries what students didn't like about games*

Table 4 summarizes what the students do not like about video games. All students were asked to answer this question, and again they were asked to select up to three reasons for not liking video games. Notice that the third part of the survey collect general information about the student: age, sex, etc.

**Scenario 1**

**Question:** *What is the proportion of students who play video games*

**Point Estimate:** The point estimate is the fraction of students who played video games one week prior to the survey. We found the proportion of people who played video games by asking participants to log how many hours they played video games in the week prior to the exam. Anyone who marked 0 hours were considered non-gamers, while those who were greater than 0 were marked as gamers. The point estimate of the 314 students was approximately 0.374 based on our 91 sample. This means that 37.4% of the students played video games a week prior to the survey.

Point Estimate: $\overline{X} = 0.374$

**Confidence Interval (CI) with Finite Correction:** Due Given a 95% confidence interval (Z-value = 1.96), the lower bound for the CI turned out to be approximately 0.289 and approximately 0.458 for the upper bound of CI. Hence the 95% CI for the students that played video games a week prior to the survey is approximately (0.289, 0.458).

Point Estimate: $\overline{X} = 0.374$

Interval Estimate: (0.289, 0.458)

Formula Used: $\left(\overline{X} - Z_{\alpha/2} \cdot \frac{\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n-1}} \cdot \sqrt{\frac{N-n}{N}}, \overline{X} + Z_{\alpha/2} \cdot \frac{\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n-1}} \cdot \sqrt{\frac{N-n}{N}}\right)$

**Confidence Interval (CI) with Bootstrap:** Bootstrap resampling can be used to make an estimate to construct the CI, this is usually done when the sample size is finite. 2000 sample means were bootstrapped and the point estimate turned out to be approximately 0.336, which was slightly lower than the point estimate of 0.374 above. With a 95% confidence, the CI interval was approximately (0.238, 0.433), which is also close to the CI above. Figure is a graph of the bootstrap of the sample proportion of gamers.

Point Estimate: $\overline{X} = 0.336$

CI: (0.238, 0.433)

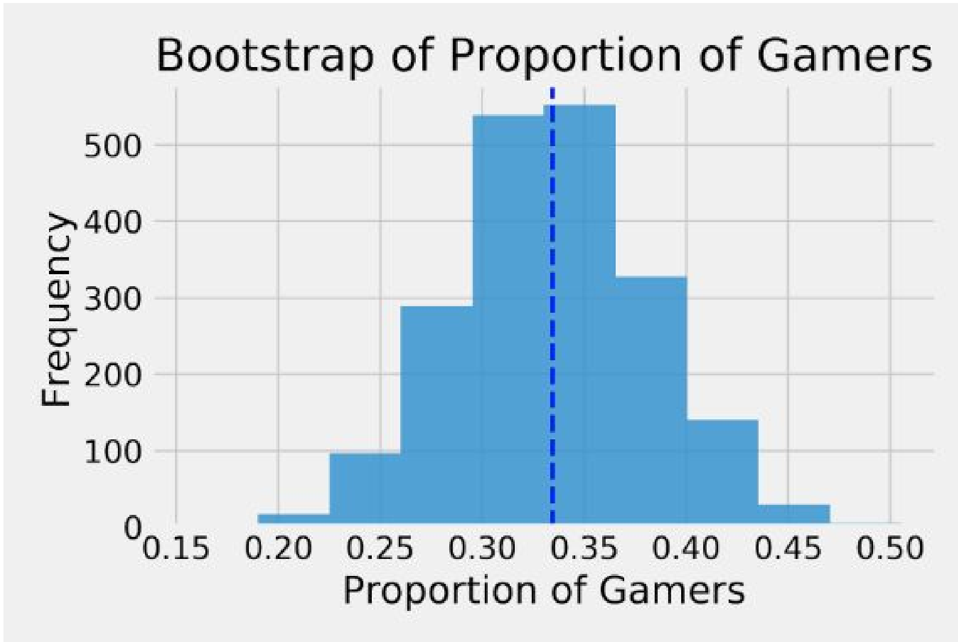



***Figure 1:*** *Distribution of 2000 Bootstrapped Sample Means*

Comparing the CI from the bootstrapped sample and CI from finite correction, CI with finite correction had a wider CI. It is important to note that CI of bootstrap resampling method is more precise and accurate.

**Scenario 2**

**Average game playing time by frequency:**

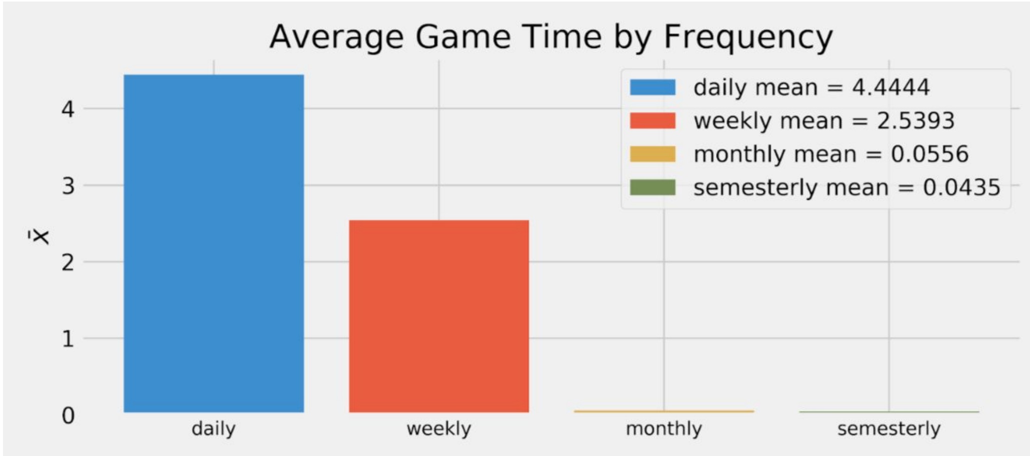



***Figure 2:*** *Histogram of Playing Time vs. Frequencies ($\overline{x}$ = hours)*

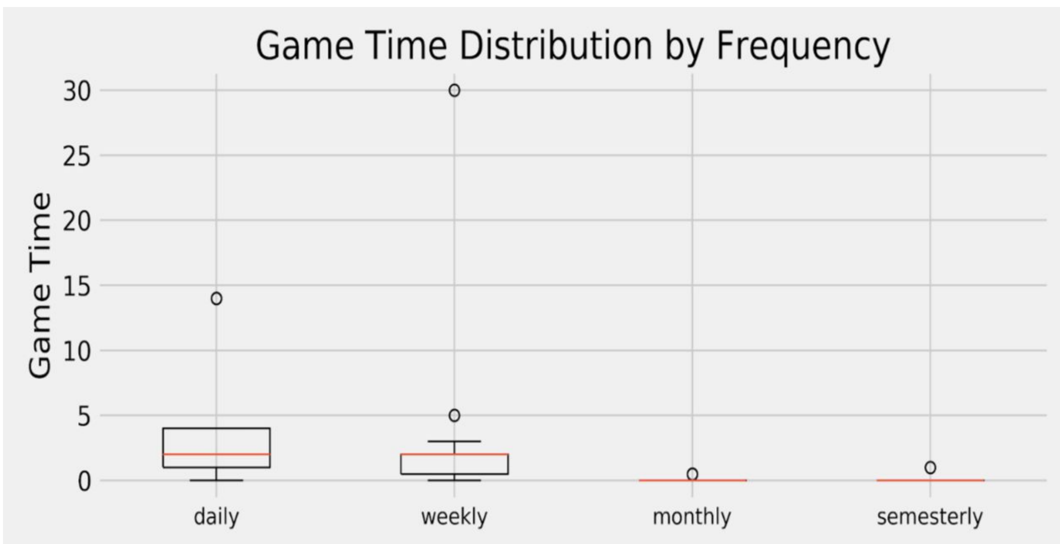



***Figure 3:*** *Box-plot of Playing Time vs. Frequencies*

The histogram above (Figure 2*)* shows the average amount of time students spent on playing games prior to the survey based on four different frequencies: daily (blue), weekly (orange), monthly (yellow) and semesterly (green). Figure 3 shows how the medians of the first two frequencies, daily and weekly, are greater than those of the other two frequencies. This results indicate that students who play more frequently tend to play longer hours than those who play less frequently. It is also important to note that the third quartile of the first box-plot (daily) is greater than that of the second box-plot (weekly). This indicates that more than 75 percent of the people who play video games daily are more likely to spend more time playing than those that play games weekly. Based on the box-plot from monthly and semesterly game players, it is hard to make any conclusions since the median and quartiles are at 0.

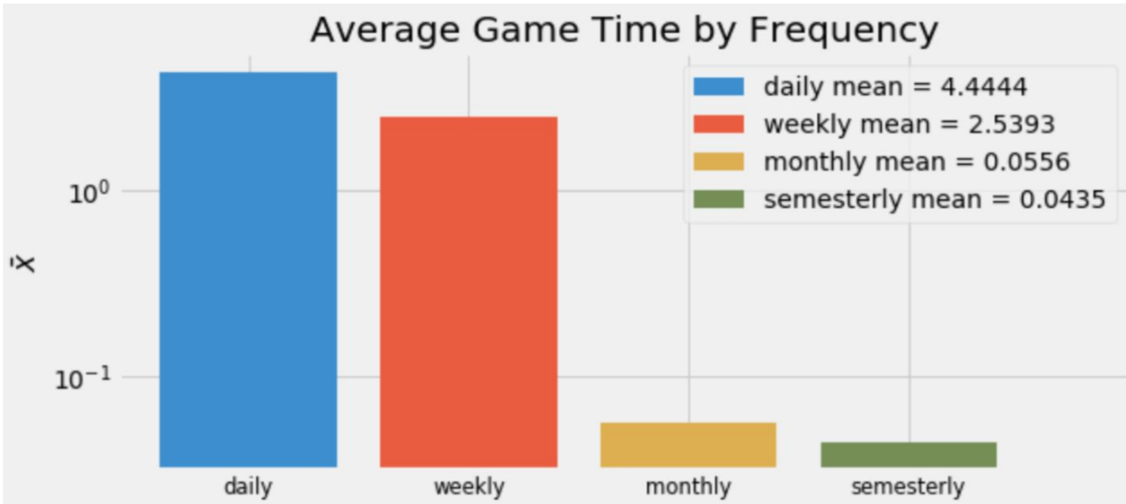



***Figure 4:*** *Adjusted Histogram of Playing time vs. Frequencies*

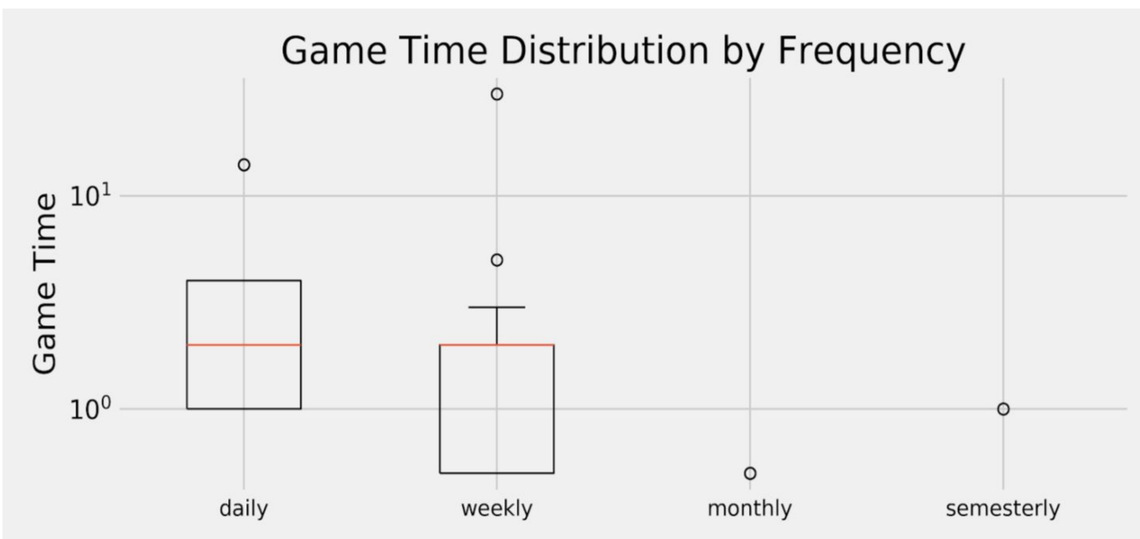



***Figure 5:*** *Adjusted Box-plot of Playing Time vs. Frequencies*

Further analysis was performed where the y-axis was log-scaled to observe the differences in average game time played by frequencies. Figure 4 shows how there is a positive association between the frequency of playing games and the average amount of time spent on playing games: students that played games daily (blue) had the highest mean (approx. 4.44 hrs) followed by students that played games weekly (orange, approx. mean 2.54 hrs), monthly (orange, approx. mean=0.056 hrs) and semesterly (green, approx. mean=0.04hrs). Figure 5 shows a better picture of the box-plots: The box-plot for daily users had the 1st quartile much greater than that of the weekly users and similar result can be seen with the third quartile as it was shown in Figure 4.

**What happens if there was an exam in the week prior:**

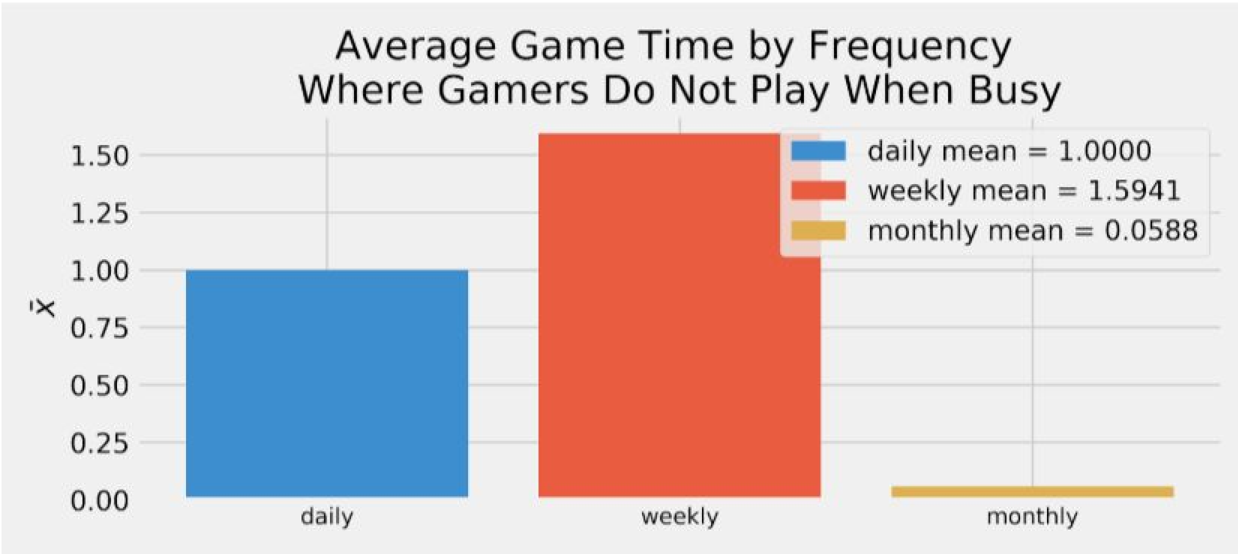



***Figure 6:*** *Playing time vs. Frequencies for students who do not play when busy ($\bar{x}$ = hours)*

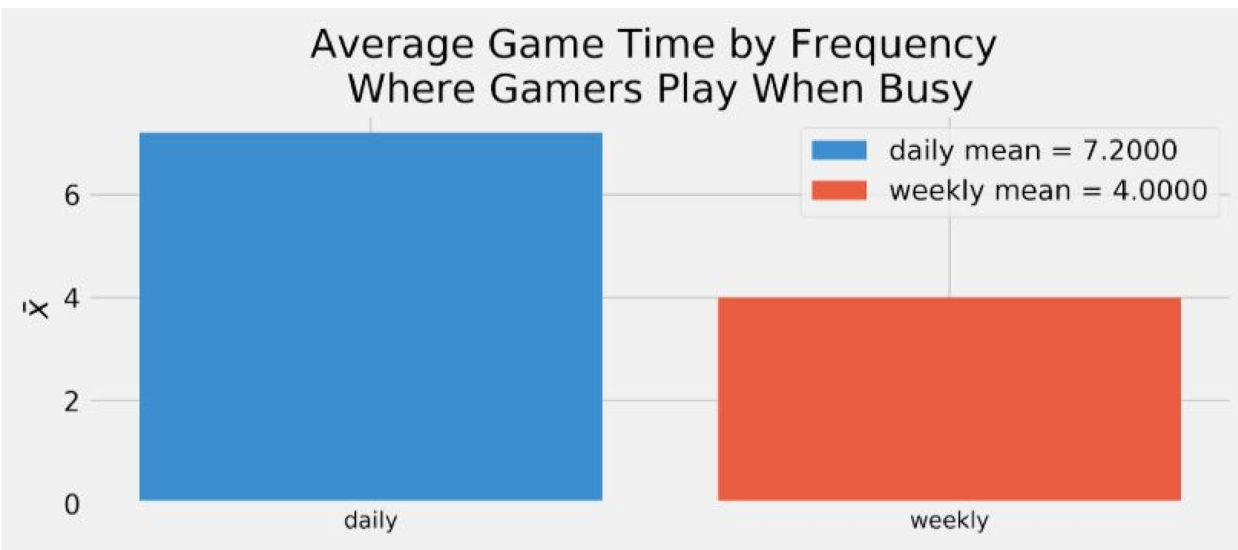



***Figure 7:*** *Playing time vs. Frequencies for students who play even when busy ($\bar{x}$ = hours)*

Figure 6 shows the average number of hours students who do not play video games when they are busy. This shows that students who play regularly, those that play daily or weekly, still showed a mean value higher than those of monthly and semesterly (students who played semesterly had a mean value of 0). Likewise, Figure 7 shows students who play regularly also have higher means (7.2 hours and 4 hours accordingly) compared to those of monthly and semesterly (monthly mean had a mean value of 0 and semesterly mean had a mean value of NA due to rows with values of 99).

In general, one can observe that despite whether students have exams one week prior to the survey, students that play video games daily and weekly had a play time average much greater than those of the students that play monthly and semesterly. Furthermore, students that play video games daily had a greater average play time despite the presence of an exam.

**Scenario 3**

**Interval Estimate:** The point estimate of the average time spent on playing video games in the week prior to the survey was approximately 1.243 hours. With 95% confidence, the lower bound for the CI turned out to be 1.113 and the upper bound for the CI turned out to be 1.373.

Point Estimate: $\overline{X} = 1.243$

Interval Estimate: (1.113, 1.373)

**Interval Estimate with Finite Correction:** Figure 8 shows a simulation study of the average time students spent on playing video games in the week prior to the survey. Notice the distribution is right skewed, meaning that the point estimate is below the median. Also the point estimate of the distribution was 1.161 and the CI was (0.595, 1.727) given a 95% confidence level.

Point Estimate: $\overline{X} = 1.161$

Corrected Interval: (0.595, 1.727)

Sampling Fraction: 91/314

**Simulation:**

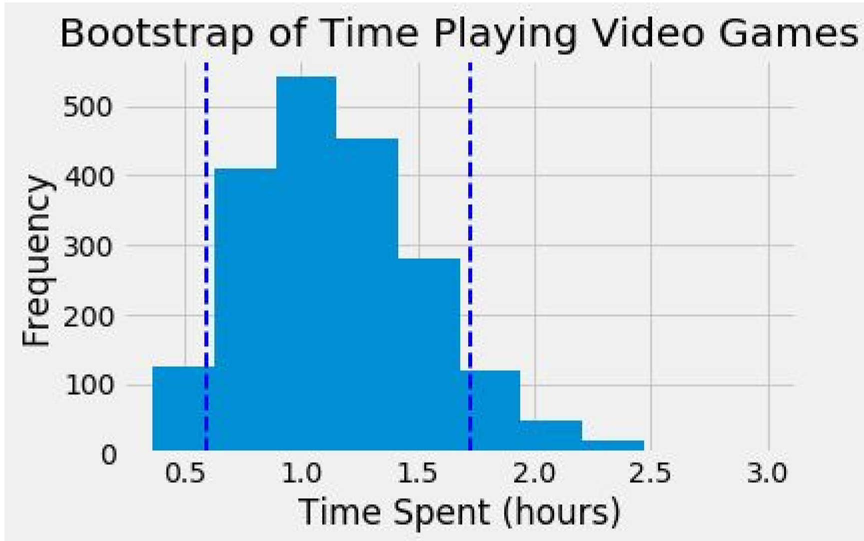



***Figure 8:*** *Distribution of 2000 bootstrapped sample means*

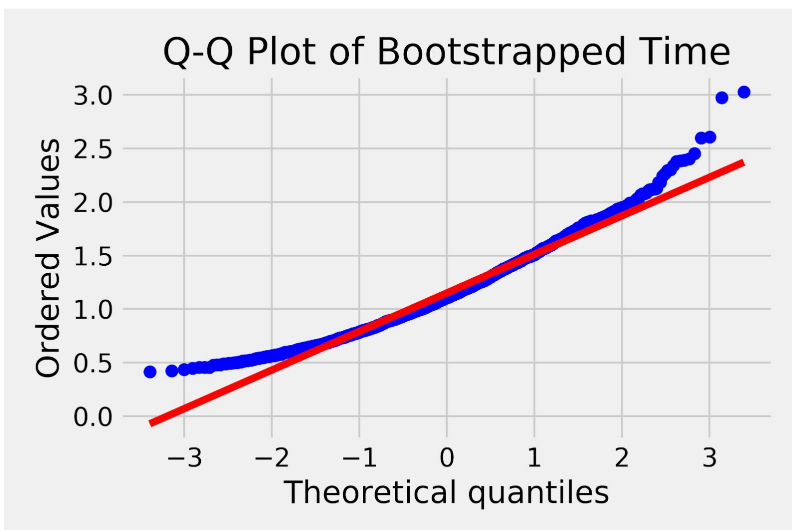



***Figure 9:*** *Q-Q Plot of the Bootstrapped Time Playing Video Games*

Kurtosis Value: 3.559

Skewness: 0.707

**Scenario 4**

**Question:** *Do students enjoy playing video games?*

We find that students have an inclination toward liking video games. From Figure 8, we can see that the mode attitude toward gaming corresponds to students somewhat liking video games, with 46 out of the 91 respondents answering in this way. Furthermore, we see that the second most popular response is that which corresponds to liking video games very much, with 23 such responses. Figure 8 shows that the remaining responses decline somewhat proportionally to the degree of dislike for video games.

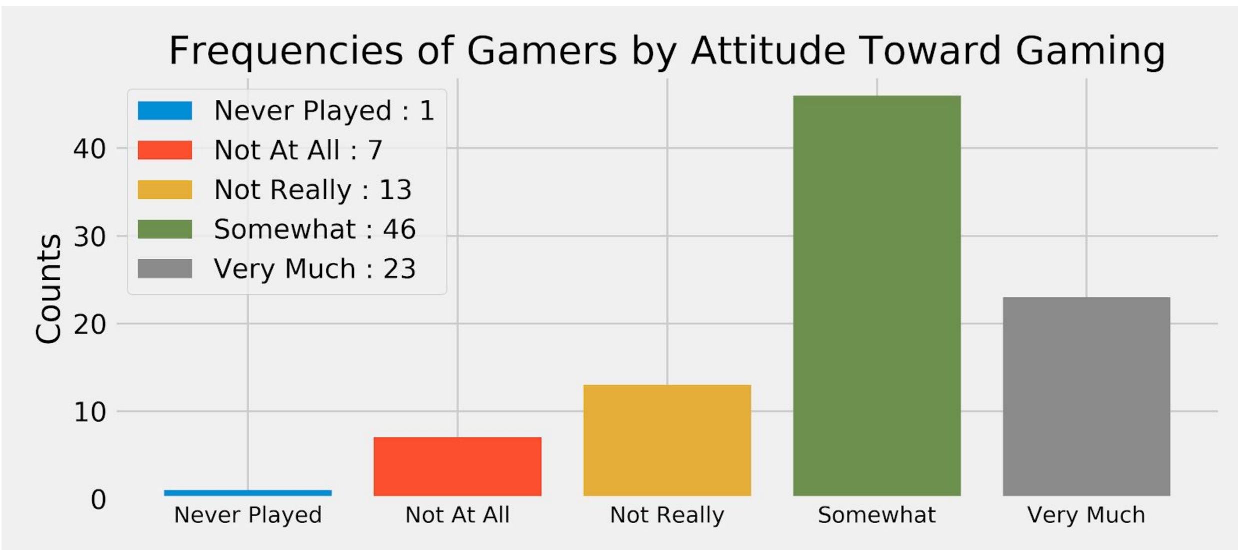



***Figure 10:*** *Number of Gamers by Attitude Toward Gaming: shows ordered by their positive sentiment toward video games and we see that the distribution is heavily skewed toward liking video games.*

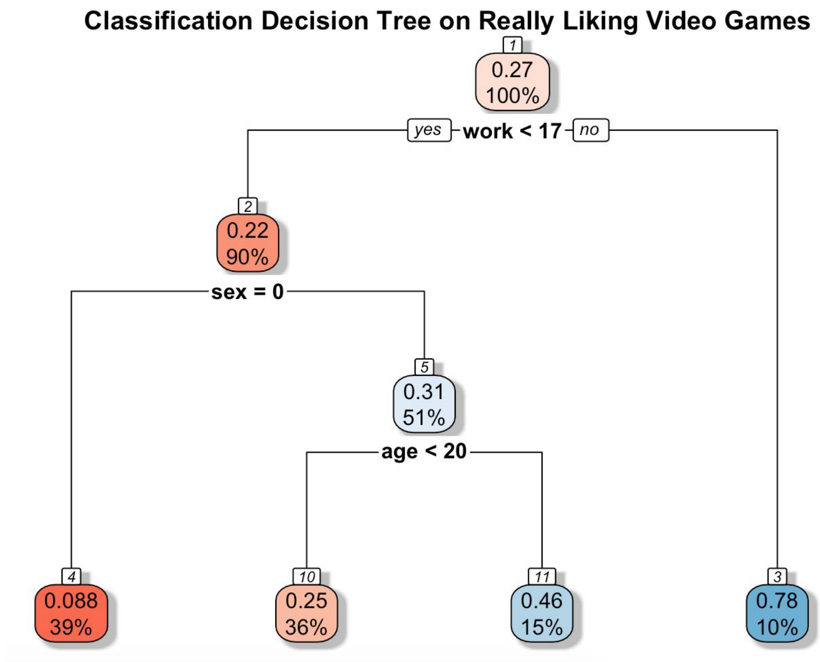



***Figure 11:*** *Classification Decision Tree on Really Liking Video Games: Shows the decision tree resulting from the binary classification task of whether or not a student 'really likes' video games.*

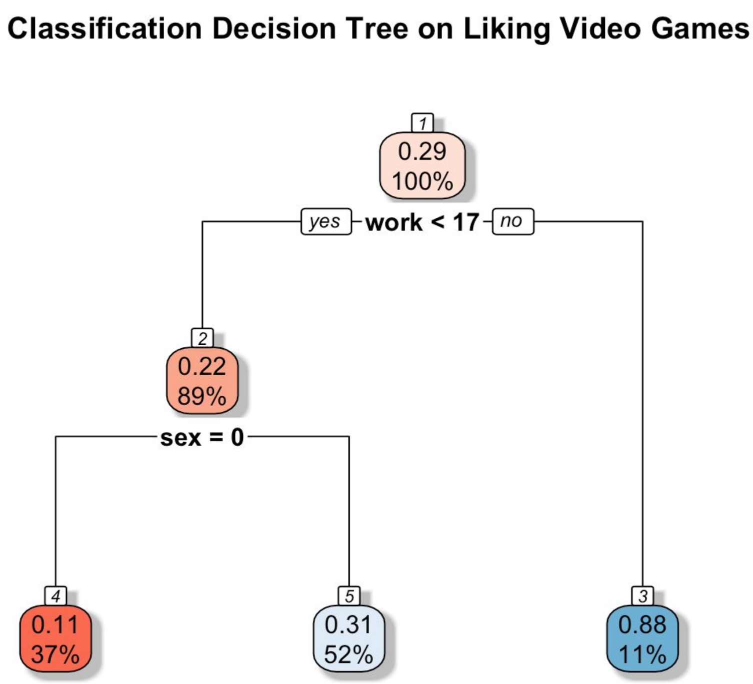



***Figure 12:*** *Classification Decision Tree on Liking Video Games: shows the decision tree resulting from the binary classification task of whether or not a student 'likes' video games, in which the variable in question is whether the student responded with "somewhat" to the question of how much they like video games.*

We built a decision tree to analyze significant variables for predicting whether or not a student likes video games. Figures 11 and 12 show the decision trees resulting from predicting positive sentiment toward video games. Our initial attempts at classification indicated that missing entries (i.e. a value of 99) in the 'work' and 'education' columns were erroneously deemed as important. We did not include the corresponding rows in this part of our analysis due to the insignificant number of affected rows. Training the decision trees on the remaining data rendered the trees shows in Figures 11 and 12, the nodes from which can be said to be the important predictive variables for 'liking' and 'really liking' video games. In considering a more general sentiment toward video games, we examine the overlap in the predictive variables used in both tasks, namely that they both contain the 'work' and 'sex' variables. This indicates that these variables are predictive for both tasks. In noteable contrast, we see that 'age' is only used in the predictive task corresponding to 'really liking' video games. From this we can conclude that 'work' and 'sex' are the most significant variables for predicting a general positive sentiment toward video games.

**<u>Scenario 5: Differences of Gamers vs Non-Gamers</u>**

We're interested in exploring whether or not differences exist between gamers and non-gamers. This will help understand if we should being a gamer could create drastically different expectations for simple demographic variables.

*Gender and Gamers*:

Question: Do the proportion of gamers versus non-gamers differ by gender?

We've done two visualizations of gamers and gender, one table and one bar graph. Looking visually at Table 5, we see that there is more gamer males than female gamers (46 vs 26), but there are about an equal proportion of female and male non-gamers (12 vs 10). We defined video gamers as we have in the previous scenario, where anyone who marked that they played 0 hours of video games are non-gamers, and gamers are anyone with >0 hours. When looking at figure 13 places the values that we have found from the table into a more visual bar graph. From there, we can see more visually that the ratio of male gamers to male non-gamers is higher than that of females, possibly indicating that men are more likely to be gamers than females (4.3 males vs 2.16 females ratio).

| isgamer<br>gender | Gamer | Non Gamer |
|---|---|---|
| Female | 26 | 12 |
| Male | 43 | 10 |

***Table 5:*** Top columns represent the two data types of gamers: non-gamer and gamer. The left hand rows represent gender titles: male and female. Those who played 0 hours of video games were considered non-gamers, anyone >0 are gamers.

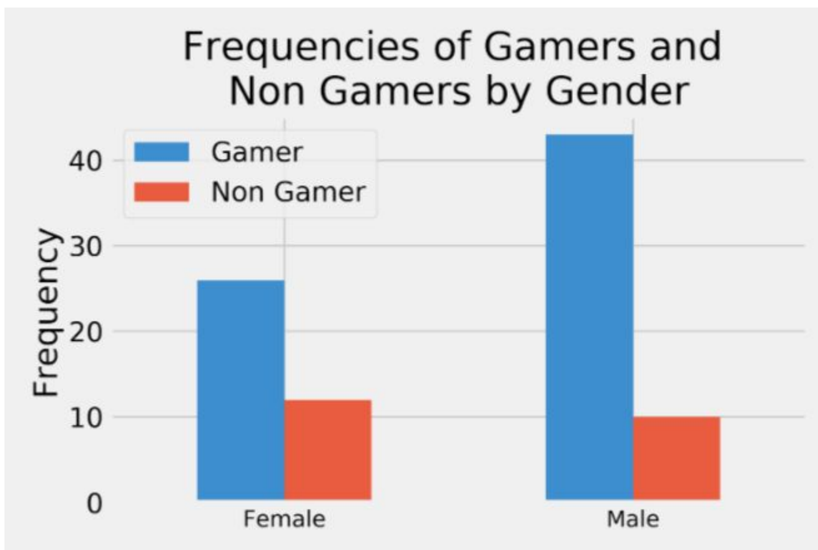
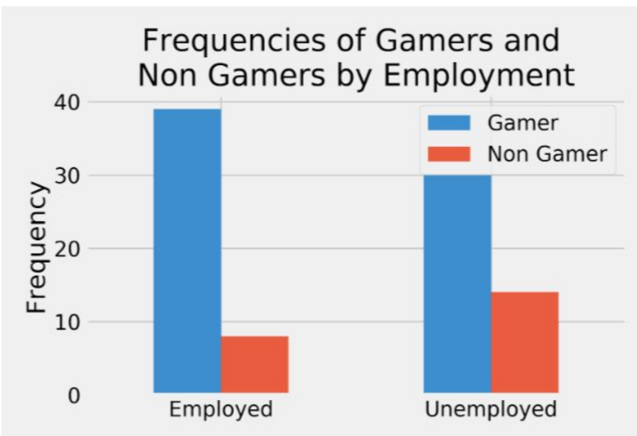



***Figure 13:*** The X axis the two groups of females versus males. The blue bars represent those within that group (male vs. female) that are gamers, while red represents non-gamers  Y axis, frequency, represents the number of people in the sample that fall under that description

Employment and Gamers

Question: Do the proportion of gamers versus non-gamers differ by employment?

We've done two visualizations of gamers and employment, one table and one bar graph. We measured employment by asking participants whether or not they have a job, and separated our data based on those who do and don't. Visually, there seems to be a similar amount of gamers for both employed and unemployed (39 vs 30) and those who are non-gamers (8 vs 14). When we visualize the tabulation with figure 14 we notice that there is a similar pattern of shape between the employed and unemployed groups. However, it seems that there is a higher ratio of employed gamers than unemployed gamers (4.88 vs 2.14). This could mean that there is a higher likelihood that employed people play games than those who don't.




| isgamer | Gamer | Non Gamer |
|---|---|---|
| **employed** | | |
| **Employed** | 39 | 8 |
| **Unemployed** | 30 | 14 |

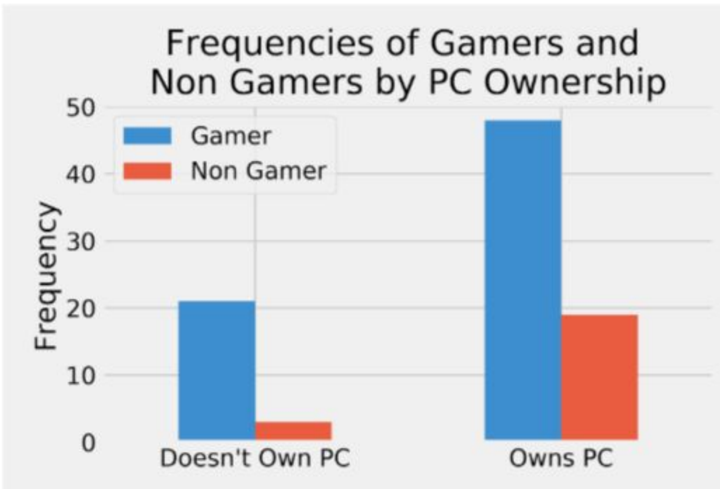
***Table 6:*** Top columns represent the two data types of gamers: non-gamer and gamer. The left hand rows represent employment status: employed, not employed.

***Figure 14:*** The X axis the two levels of employment: employed and unemployed. The blue bars represent those within that group (employed vs unemployed) that are gamers, while red represents non-gamers Y axis, frequency, represents the number of people in the sample that fall under that description

PC ownership and Gamers

Question: Do the proportion of gamers versus non-gamers differ by PC ownership?

We've done two visualizations of gamers and PC ownership, one table and one bar graph. We measured pc ownership by asking participants whether or not they own one and separated our data based on those who do and don't. Looking at both the table and figure, we notice that there seems to more gamers who own a PC compared to gamers who don't own a PC ( 59 PC gamers vs 22 non-PC gamers). This could mean that our sample of gamers are mostly ones more comfortable with using a PC, or that PC games are more popular, however we cannot establish causation. There also means to be more gamers who own a PC compared to gamers who don't have a PC (48 vs 21). Also more people have PCs than not own a PC ( 67 vs 24), which could indicate that the lab should involve a computer as it is one that most people are familiar with.




| isgamer | Gamer | Non Gamer |
|---|---|---|
| ownpc | | |
| Doesn't Own PC | 21 | 3 |
| Owns PC | 48 | 19 |

***Table 7:*** Top columns represent the two data types of gamers: non-gamer and gamer. The left hand rows represent ownership status: yes, or no.

***Figure 15:*** The X axis the two levels of Ownership: doesn't own PC and owns pc. The blue bars represent those within that group (doesn't own PC vs Owns PC) that are gamers, while red represents non-gamers  Y axis, frequency, represents the number of people in the sample that fall under that description

**Scenario 6: Expected Grades vs Expected Distribution of Grades**

**Question***: How does our target distribution of grades (20% A, 30% B, 40% C, 10% D or lower) differ from the expected grades that students expect*

When we compare the expected grades by students versus the expected distribution, there seems to be a great difference with the score. We can do this by using Figure X. With theGrade of A, only 8.7% of students said they would expect it, compared to the 20% in the distribution. 57% of participants expected a B, when we only expect the distribution to have 30%, a 27% difference. The closest one would be C, where there's only a 6% difference between student's expectations and the expected distribution. For D, not a single respondent expected to get a D or lower, which would be a sign that everyone wants to not fail the class. However, if the non-respondents (4 people in the sample) were the people who were failing the class, then we would have the expected value of 4/95 to be 4.6% to be people who expected a D or lower which is closer to the 10% expected average for D's and lower.

| | *Expected Grades by Students* | *Expected Distribution* |
|---|---|---|
| *A* | $\frac{8}{91}$ = .087 = 8.7% | 20% |
| *B* | $\frac{52}{91}$ = .57 = 57% | 30% |
| *C* | $\frac{31}{91}$ = .34 = 34% | 40% |
| *D* | $\frac{0}{91}$ = .0 = 0 % | 10% |

***Table 8:*** Expected Grades by Student vs Expected Grades for the class. A, B, C, D represents the grade that they could receive. The ratio or division bars in "Expected Grades by Students" represents the number of participants who expected that grade (numerator) and the total number of respondents (denominator).

**Chi Squared Test**

Since we're working with categorical data, and we have the actual expected distribution of scores, we can conduct a one-sample chi squared goodness of fit test. More on the theory behind this can be found in the theory section. When we conduct the test, we get a Chi-squared value of 37.97, df 3, p <.0001, indicating that the expected grades by the student is greatly different than the expected distributions. Limitations include having data that contains a 0 for the observed frequency (D), and having the B response having a 4.73 residual, which is greatly deviated from its average. When we conduct the test with the non-respondents, X(3) = 30.22, p <.0001, meaning can reach the same conclusion with less strength. B remains the same in terms of having a higher residual at 4.4

**Conclusions**:

We were tasked by the Statistics program of berkeley with the goal of answering the question of what aspects of video games do people like so we can apply it in the process of creating a more interesting and successful statistics lab. To do this, upper division students developed a survey and asked participants in the intro stats class about their video game usage and different aspects about it. I will now lay out the conclusions that we can make based on the analysis that we have done.

**Why do People play?**

From our major analyses, we have a couple of recommendations for future directions for creating the future intro to statistics lab. IN an overall view, people who do play games very much enjoys playing the game. Some of the reasons why people play games based on past research have shown that competitiveness, competence, relaxation, and graphics were motivations. IN our analysis, we were able to replicate relaxation, with its effects greatly amplified in this sample (being the most popular compared to being the 4th most popular).We would suggest to the creators to try and make the game simple or relaxing and not make it a stressful endeavor to play the lab.

**Business/Work**

Business seems to matter for whether or not people do play video games, where people who are busy tend to play a smaller amount of hours compared to those aren't. This could mean that time for the lab should be accounted for in terms of how long it should be. Busy gamers still play video games in their free time, but if the lab wants to create something that needs to be worked outside of class, we recommend that they keep in mind people's busy schedules

**PC Ownership**

We found that a larger proportion of people who played PCs were gamers, which could represent that people prefer computer games or that gamers are typically ones who use the PC. We would recommend it to the creators to develop their game in the PC, instead of other consoles. This is better as PC's are generally easier to share and access between people compared to just using a console for a game.

On average, there seems to be a larger proportion of male gamers compared to female gamers. We recommend reading our "accounting gender" section to understand how this could interplay with the structure of the game that will be made for the class.

**Accounting Gender**

Previous research has found that females prefer other genres and rationale of games compared to males. Specifically, females like games where it's less competitive and more cooperative, and involves less female sexualiazation and violence. Part of this was supported by our analysis, where very much prefer sports video games compared to women, and a little bit more of action. We would recommend that games stray away from the structure of sports games, and possibly full on competition if the statistic labs wants to be inclusive to at least two genders of male and female.

We also found in our study that females greatly prefer strategy games over everything else. So we would recommend that if the class is majority female, then the game should have a lot of strategic elements to it to challenge and interest the class.

**Accounting Genre**

Building up on it, genre preference has been shown in research to have different ratios of enjoyment. Our analysis found similar trends and saw that action games have the most counts of people who very much or somewhat likes playing it. In addition, Strategy games had less "very much" likes for this genre of video games, but had a significant amount of "somewhat", making it still a very reliable and well liked genre. We recommend to the makers of the lab to focus on an action or strategy focused approach for the game. This works perfectly with gender as strategy was the popular genre for females, and still a popular one for males.

**Final Words**

Even though we've developed these recommendations, limitations still need to be reminded for possible issues. First, the sample was dependent and was only on Berkeley students, which limits its external validity to another statistic groups. Second, the study was conducted in 1995, which means it's outdated as the world of video games have expanded greatly since then. Finally the survey was developed by upper division statistics students, and not by trained questionnaire developers (possibly a psychologist), so the validity of the questions can not be fully addressed or be confident in.

**<u>Bonus Question: Gender and Video Game Preference</u>**

Question: *Does Video game genre preferences differ between genders*

From past research, we know that there are some differences in preferences for why males or females play video games. Although we don't have data for what the person actually prefers to play, we can compare the choices of genres that males and females play and compare them (Figure 16). The X axis of the figure represents the 5 different genre types that individuals were able to select as games that they play (Action, Adventure, Simulation, Sports, and Strategy). The different bars correspond to different genders of Male (Red) and female (blue). The Y axis represents the frequency or amount of participants that chose that choice (counts). The counts don't add up to 91 because participants were able to choose multiple genres when asked the question.

When looking at the figure, we find that in most cases (⅘) male counts are higher for the genre, however, females actually have higher counts when we consider strategy games. Strategy games might be the best for both worlds. To see a better visualization as the number of female vs male responses

Proportion Graph:

Figure 17 represents a more fair comparison between the two groups of male and female. We've calculated the proportion of each gender by themselves for each genre. For example, we calculated that ~50% of all choices for females were captured by strategy, when just looking at females. With a more standardized figure, we can find that females greatly prefer strategy games compared to male (over double the size of males). Males seems to have a larger preference for sports and action games compared to female, however, strategy games is still one of the top 3 games for males, and they have a larger spread of genre variability. From this, it would seem that the genre of game that would appeal best for both genders would be some sort of strategy game, with the second being action games.

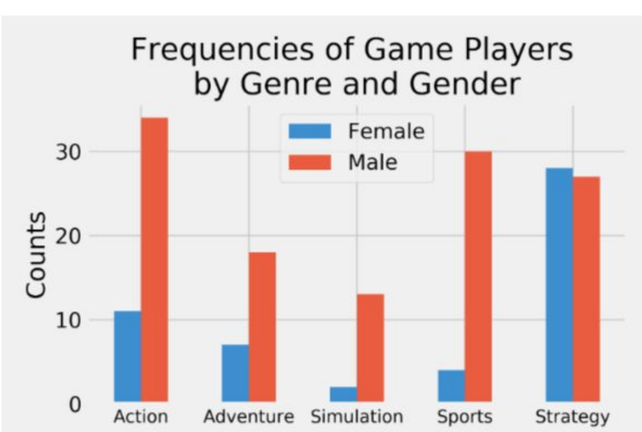



***Figure 16:*** The X axis of the figure represents the 5 different genre types that individuals were able to select as games that they play (Action, Adventure, Simulation, Sports, and Strategy). The different bars correspond to different genders of Male (Red) and

female (blue). The Y axis represents the frequency or amount of participants that chose that choice (counts)

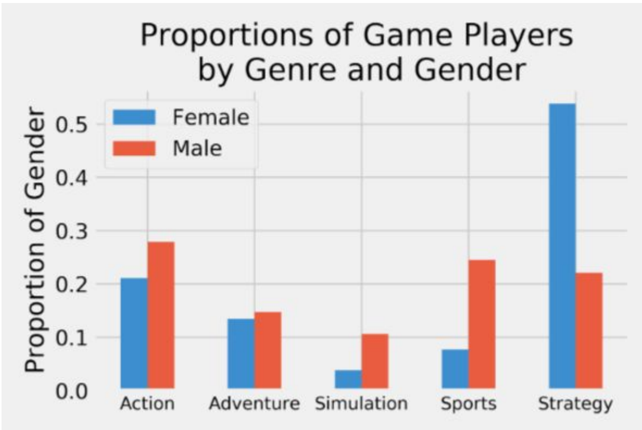



***Figure 17:*** The X axis of the figure represents the 5 different genre types that individuals were able to select as games that they play (Action, Adventure, Simulation, Sports, and Strategy). The different bars correspond to different genders of Male (Red) and female (blue). The Y axis represents the proportion of participants in that specific gender who chose that choice.

**Second Bonus Question:**

**Question:** *Does the Genre of video games influence the attitude of gamers? Do certain genres appeal to more players compared to others*

Based on our previous research and our analysis, we know that genre of video games do influence the likelihood that people will first play the game, and also how long they spend time on it. To create the best lab environment, we would want to see which specific genre leads to a higher ratio of people who enjoy actually playing it. In figure 18, the x axis represents the different genres that the participants got to answer. The different bars represent their response to the survey on their attitude on gaming: very much (blue), somewhat (red), and not really (yellow). We got rid of two answers of “don’t play” and “not at all” because for most groups, the values were 0s. Visually, we see that a large proportion of the sample somewhat likes strategy games. Action is the one where people like it the most (with sports, strategy, and adventure approximately tying for second). Interestingly, strategy games had the highest “not really” responses, but in terms of proportion to the people who liked it, it’s small (48 to 6).

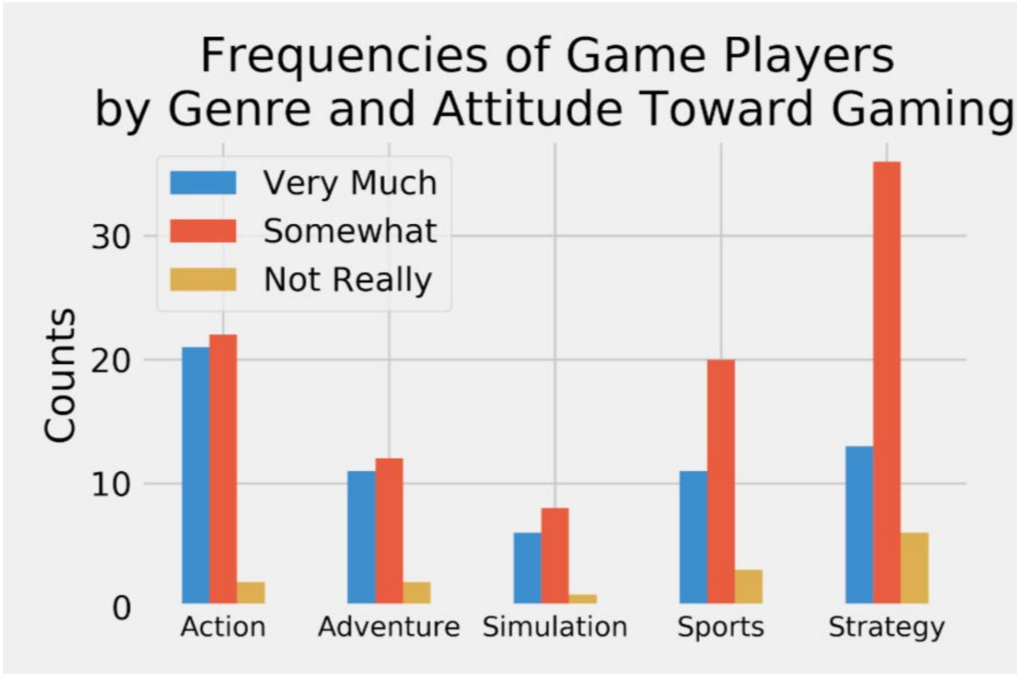



***Figure 18:*** The X axis of the figure represents the 5 different genre types that individuals were able to select as games that they play (Action, Adventure, Simulation, Sports, and Strategy). The different bars correspond to different attitude responses of Somewhat like video games (Red),  very much like video games (blue), and not at all likes video games yellow). The Y axis represents the frequency or amount of participants that chose that choice (counts).

**Theory**

Population units: Total number of group that will be investigated.

Population size : Total number of units in the population, usually denoted as 'N'. The population size for this case study is 314.

Unit characteristic: A piece of information from each member of the population. The characteristic we are concerned in this study is the amount of time the student played video games in the week prior to the survey.

Population parameter: Summary of the characteristic for all units in the population, i.e.

average amount of time the student played video games in the week prior to the survey.

Sample units: Proportion of the population selected for the sample.

Sample size: Number of population units chosen for the sample, usually denoted as 'n'. The sample size for this case study is 91.

Sample statistic: Numerical summary of the characteristic of the units being sampled. The average time student spent on playing video games is the sample statistic in this study is an example of a sample statistic. The following is a mathematical expression of the sample statistic:

$$\bar{x} = \sum_{i=1}^{n} x_i \cdot \frac{1}{n}, \text{ where } n = \textit{sample size}$$

Estimators for standard error: SE (Standard Error) of an estimator is the standard deviation of an estimate of the standard deviation. SE is an indicator of the size of the deviation of the estimator from its expectation.

Central Limit Theorem: Central Limit Theorem states that given a large sample size from a finite population variance, the sample mean will approximately equal to the mean of the population. In a more rigorous sense, given n i.i.d. random samples from a distribution with mean $\mu$ and standard deviation $\sigma$, then the mean of the sum of all the variables will have a normal distribution.

QQ-Plot: Quantile-quantile plot is a probability plot used to compare the similarities and differences between two probability distributions by plotting their quantiles against each other. A Q-Q plot is used to compare the shapes of distributions, providing a graphical view of how properties such as location, scale, and skewness are similar between two distribution. A normal Q-Q plot comparing standard normal data

with

standard normal population, for which the linearity of their relation suggests that the data are normally distributed.

Skewness: Skewness is a measure of asymmetry property of the probability distribution by comparing the median with the mean of a data sample. Most of the time, skewness is used to check the normality of distribution. For a standard normal distribution, the skewness coefficient is approximately equal to zero. Skewness is calculated by the third power of the average of the data sample.

Chi-Squared Tests: Chi-square tests are a family of significance tests that give us ways to test hypotheses about distributions of categorical data. This topic covers goodness-of-fit tests to see if sample data fits a hypothesized distribution, and tests for independence between two categorical variables.

Bootstrap: In statistics, bootstrapping is any test or metric that relies on random sampling with replacement. Bootstrapping allows assigning measures of accuracy to sample estimates. This technique allows to estimate the sampling distribution of almost any statistics using random sampling methods. Generated by bootstrap method, the distribution is approximately to the population. Bootstrapping is the practice of estimating properties of an estimator by measuring those properties when sampling from an approximating distribution(wiki). It is a straightforward way to derive estimates of variance and confidence intervals for complex estimators of complex parameters of the distribution.

Confidence intervals: In statistics, a confidence interval is a type of interval estimate, computed from the statistics of the observed data, that might contain the true value of an unknown population parameter. The interval has an associated confidence level that quantifies the level of confidence that the parameter lies in the interval. If confidence intervals are constructed using a given confidence level from an infinite number of independent sample statistics, the proportion of those intervals that contain the true value of the parameter will be equal to the confidence level. The confidence level is designated prior to examining the data. Most frequently, the 95% confidence level is used.

## VII. References


[1] Johnson, D., Gardner, J., & Sweetser, P. (2016). Motivations for videogame play: Predictors of time spent playing. Computers in Human Behavior, 63, 805-812.

[2] Olson, C. K. (2010). Children's motivations for video game play in the context of normal development. Review of general Psychology, 14(2), 180-187.

[3] Lee, M. S., Ko, Y. H., Song, H. S., Kwon, K. H., Lee, H. S., Nam, M., & Jung, I. K. (2006). Characteristics of Internet use in relation to game genre in Korean adolescents. CyberPsychology & Behavior, 10(2), 278-285.

[4] Hartmann, T., & Klimmt, C. (2006). Gender and computer games: Exploring females' dislikes. Journal of Computer-Mediated Communication, 11(4), 910-931.

[5] Fuller, S. (2017). Video Game Industry- Statistics & Facts (n.d.). Retrieved February 25, 2018, from https://www.statista.com/topics/868/video-games/