Anthony Ayon
Ooduah Orafidiya

Case Study 2 Report

Introduction

Year after year around 3000-4000 enroll in statistic courses at UC Berkeley where half of the students prepare by taking an introductory statistic course. In order to help students develop their critical thinking skills, a committee of faculty and students developed a series of computer labs which many claimed correlated with video games. The labs consisted of not following the traditional norms and offer an alternative method to learning statistics and probability, which in some way many video games offer. In aim of developing well rounded labs, the committee decided to conduct a survey to all lower division students enrolled in a statistics class. The overall intent of the survey was to determine the extent to which students play videogames and aspects of it they liked or disliked.

Data

For the class, Statistics 2, consisting of 314 students, 95 were randomly selected to participate in the survey. Overall 91 out of 95 completed the survey. This completion rate was due to data collectors attending both discussion sections on Tuesday and Thursday, survey being given when exams where returned and students who did not attend section where given the option to take the survey during class.
Data of the 91 students is either numerical or categorical where each column represents a distinctive question. The '99' within the data represents that the question was not answered correctly by the student.

Table 1: Survey Properties

Column Name	Description
Time	Number of hours played week prior to survey
Like	Like to Play 1 = never played 2=very much 3=somewhat 4=not really 5=not at all
Where	Where they play 1=arcade 2=home system 3=home computer 4=arcade and either home computer 5=home computer and system 6=all three
Freq	How often 1=daily 2=weekly 3=monthly 4=semesterly
Busy	Play if busy 1=yes 0=no
Edu	Videogame playing is educational 1=yes 0=no
Sex	1=male 0=female
Age	Student age in years
Home	Have computer at home 1=yes 0=no
Math	Hate math

	1=yes 0=no
Work	Number of hours worked week prior to survey
Own	Own PC 1=yes 0=no
Cdrom	PC has CD-Rom 1=yes 0=no
Email	Whether student has email 1=yes 0=no
Grade	Grade expected by student 4=A 3=B 2=C 1=D 0=F

.

A second survey was incorporated to determine the reason why students like or dislike playing videogames. These are not attached to the question in the first survey and the tables below provide a scope of the data results.

Table 2: Type of Videogames Played

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

Table 3: Reason For Playing Videogames

Reason for playing ?	Percent
Graphics/Realism	26%
Relaxation	66%
Eye/Hand Coordination	5%
Mental Challenge	24%
Feeling of Mastery	28%
Bored	27%

Table 4: Reasons Not Like About Videogames

Did not like about playing?	Percent
Too Much Time	48%
Frustrating	26%
Lonely	6%
Too Many Rules	19%
Costs too Much	40%
Boring	17%
Friends Don't Play	17%
It is Pointless	33%

Results

Scenario 1

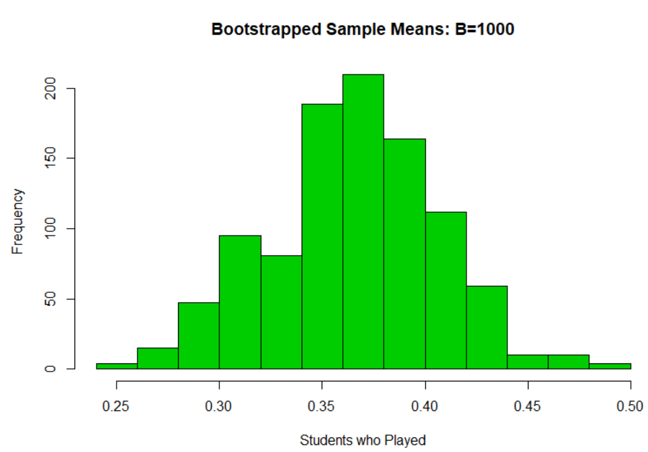
For this procedure, we will take into consideration the students who claimed they played videogames prior to the survey.

- CLT-Only Students Who Played
  In order to get a feel for the filtered data we applied the Central Limit Theorem. Again only the data where students claimed they played was used. Even though this is a small dataset that can drastically be affected if more data were to be introduced, we can say it gives us a general scope of our sample.
  `Average Time Played: 0.3736264`

  Through the first scenario we did a point estimate which returned 37% of the class played video games in the week prior to the survey with a 95% confidence interval of 0.274 as the lower bound nad 0.473 as the upper bound percentage of students.
  `Confidence Intervals: (0.2742318, 0.4730210)`

- Mean Bootstrapped Results
  After using bootstrapping with iterations being B = 1000, that essentially resample and compute new each calculation, we took an overall average. This average is to compare with the computed Central Limit Theorem above.
  `Average Time Of Students that Played: 0.3647253`

Figure 1:



- Bootstrapping: Confidence Intervals
  This method gives us an overall clearer view of the confidence interval. Since bootstrap also randomizing it sampling selection, this gives us a more accurate confidence interval that may have not drastically changed to the CLT one but entails a clearer picture of

where the data lies. Hence, it can be stated that with 95% confidence, the fraction of students who played videogames lies within this interval.

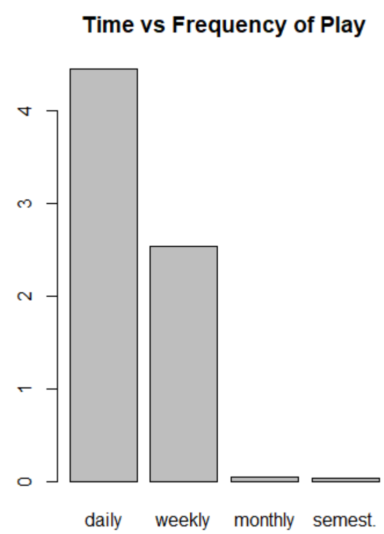
```
Bootstrap Confidence Interval: (0.2857143, 0.4395604)
```

<u>Scenario 2</u>

The overall time to frequency comparison demonstrated that there is a significant difference for those who claimed they played daily or somewhat. In comparison with frequency data, there were a lot of students who did not play at all (time = 0) a week prior to the test. Just as mentioned in the top section, students may have altered their schedules knowing there was a test coming up. In the other hand, below we analyzed the students that claimed that they did play and tracked whether there was any association between frequency and time played.

The bar plot above is an average of the time played by each student a week prior to the survey based on the level of frequency they dedicate to videogames. Hence, we can see out of the four possible categories, students who stated that they played videogames on a 'daily' basis dedicated almost double the time playing than those who stated that they play 'weekly'. In comparison with 'monthly' and 'semesterly' responses, which obtained an average of about 0.05555 and 0.04347 respectively, correlates well considering those students are not frequent gamers and less likely motivated to start playing prior to an upcoming exam.

Figure 2:



In order to further understand any direct impact between time spent playing and having a test in the week prior, we decided to include expected grade for the course. This may give us an interesting perspective on the overall effort or gaming time a student is willing to sacrifice prior to a test. One aspect to point out is that not many responses where given for C grades which may be attributed to the fact that most students start a course with high expectations and not all students need the same study time to perform well. Also, the survey was taken prior to students receiving their test results which makes it harder to determine a final grade. On the contrary, interesting results were obtained and interesting trends where found.

One of the biggest trends we find in these results are in the average times found within the 'Daily' column. Here we see average hours played substantially lower in students expecting an A in comparison to a C. This can be a small-scale indication that students who had lower grade expectations maintained a certain level of playing time regardless if there was a test or not. Again, only 8% of the sampled students answered a C grade, so these average times may not be representative of the entire student population. This tendency does continue with A and B grades, where the playing time is almost double for B grades in the 'Daily' and 'Weekly' categories.

Table 4: Frequency vs Grades Comparison

	Daily	Weekly	Monthly	Semesterly
Grade : A	1.800000	1.507143	0.00000000	0.00000000
Grade: B	5.666667	3.571429	0.08333333	0.07142857
Grade: C	14.000000	NaN	0.00000000	0.00000000

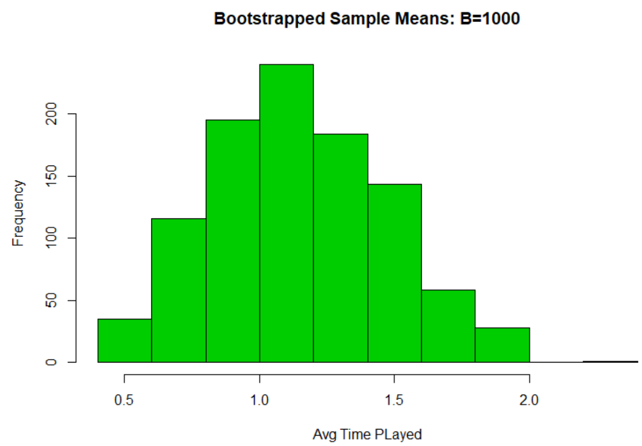
<u>Scenario 3</u>

- <u>CLT- Overall Time</u>
  For this case we focused on the entire student population instead of focusing on only the students who dedicated some playing time. Just like in *scenario 1,* we applied the Central Limit Theorem to the data just so we could get a glimpse of our estimated range.
  ```
  Average Overall Time: 1.242857

  Confidence Interval: (0.4668263 2.0188880)
  ```
  Our confidence interval has a 95% certainty that the value lies within 0.466 and 2.018 which is a little more distributed when comparing to *Scenario 1*.

- <u>Mean Bootstrapped Results</u>
  In our bootstrap we used B = 1000 iterations then calculated an average of all values.
  ```
  Average Overall Time: 1.145062
  ```

Figure 3:

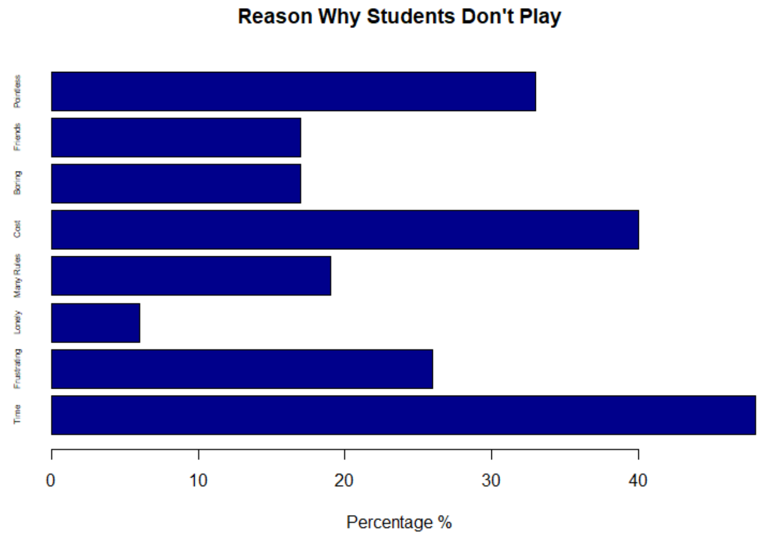


- <u>Bootstrapping</u>
  Similar to the procedure done in scenario 1, for bootstrapping, the sample was now taken from all 91 samples. We can see the relative decrease in the interval estimates considering all the zeros present in students who did not play any video games. Our interval, in comparison to the CLT done above, is a little smaller giving us 95% certainty the value lies within this interval.
  `Bootstrap Confidence Interval: (0.5821978 1.8197802)`

<u>Scenario 4</u>

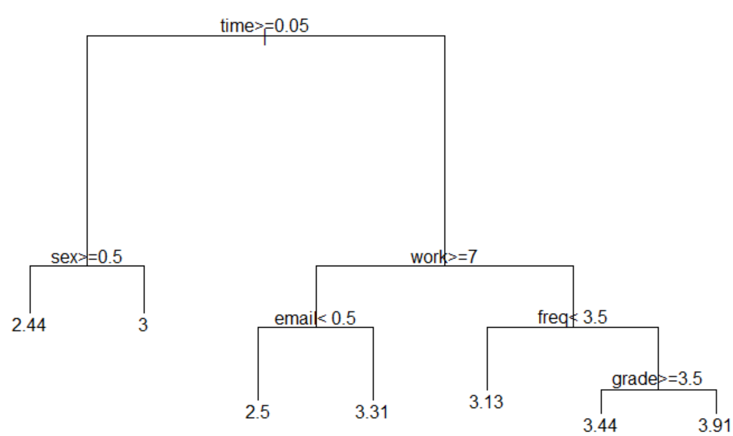
We will now focus on potential factors of why students may like or dislike playing videogames. As we saw in *table 4*, “Too Much Time” overwhelmingly wins as the number one reason students disliked videogames. In the other hand, we believe “Too Much Time” along with “Cost too Much” and “Friends Don’t Play” could be the most significant reasons for people disliking videogames because they were the only answers that didn’t address how the survey participants emotionally felt about them. The total percentage of people who marked these answers where 48% for “Too Much Time”, 40% for “Cost Too Much” and 17% for “Friends Don’t Play”.

Figure 4:



Next we applied the CART algorithm on all 91 samples to see how different features affected someone likelihood of either disliking or liking videogames. Not only will this give us a better visual of all factors that affect a student's position toward playing videogames but, also potential connections to cost, time and social attributes.

Figure 5: Regression Tree (All Data)

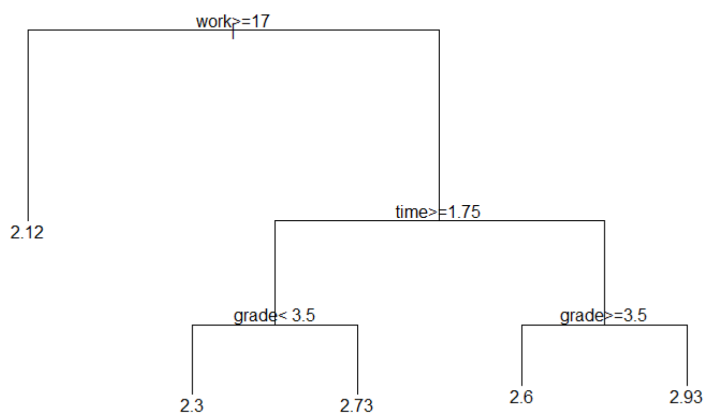


In *figure 5*, each leaf at the end represents the average of the categorical values used for "Like to Play". The root node depicts a value of 0.05, this is the overall mean for "time" or hours played and is this low because it also considers the students who played zero hours.

Hence, the left split node will focus on students who dedicated more than 0.05 hours playing while the right split node denotes students who played less than 0.05 hours, essentially zero hours. From the left we can see a fairly simple split with males averaging 2.44, indicating that most students in that group liked videogames very much, while females somewhat as their average is a solid 3. The right split node splits further with those who work 7 hours or more and those who work less than 7 hours. It is interesting to point out that those students who worked 7 or more hours and did not have an email had a 2.5 average, meaning they enjoyed playing the most when compared to the other groups in that left node. Also, students where less likely to like video games if they worked less hours and played rarely.

Now, if we filter this same method to only the students who stated that they like videogames very much and somewhat, we get some of the same factors.

Figure 6: Regression Tree: Liked Videogames

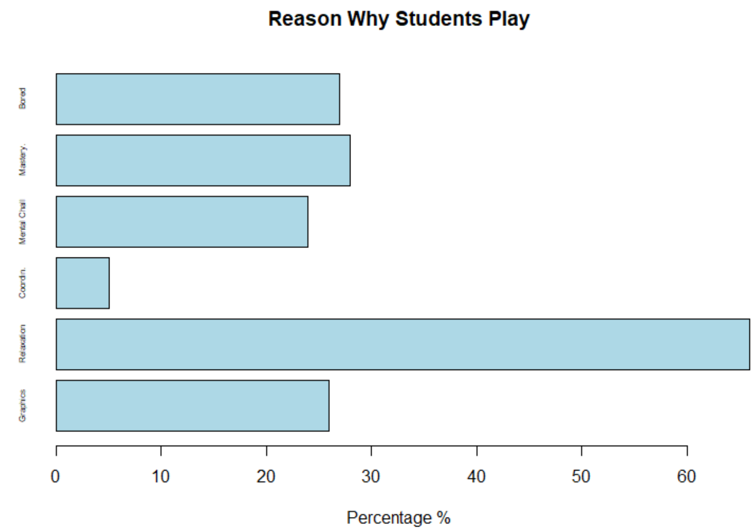


As we can see in *figure 6*, on average, students who worked more than 17 hours really enjoyed playing videogames. On the contrary, those who worked less than 17 hours where influenced by hours played and expected grade. For instance, a student who did not work that week, played less than 2 hours and was not expecting an A in the class, is less likely to enjoy videogames.

In general, all these factors play an essential role in determining why students like or disliked playing videogames. After looking at both regression trees, we can assume students that work the most and have less free time find greater enjoyment in playing. This correlates with the findings in *table 3* because most students playing videogames are seeking relaxation, thus it makes sense that those who work long hours seek less physically stressful activities like videogames. Also, in most cases students who do not work have fewer financial freedoms hence, will be less likely to invest in videogames or consoles. As a result, these sentiments are reflected in cost and time described above, where videogames are not worth the investment if its not a source of relaxation.

On a different note, we can also state that people who work less may prefer to invest their time on other activities that do not involve videogames. If there isn't a social environment within videogames, students typically find other activities that do provide it. In effect, although students may have more free time, it doesn't necessarily mean they will spend it on videogames but rather other group activities that is in the best of interest for everyone.

Figure 7:



Scenario 5

In this section we will have visual representations of some of the comparisons we have been discussing throughout this paper. This will help us better catalog different groups depending on what they responded in the survey. Some of the tools used where cross-tables, bar plots and boxplots.

Comparison 1: Like vs Gender

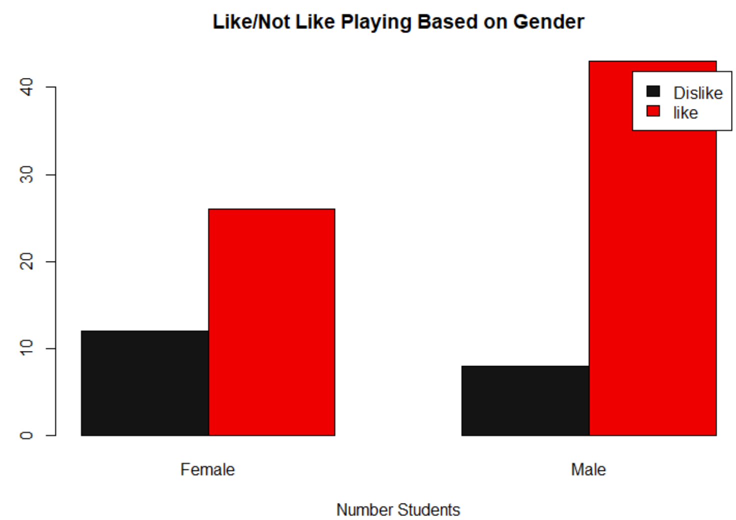
Below in *figure 8* we have a cross-tabulation between students who disliked or liked video games. Two individuals where excluded from this comparison because they claimed they never played, leaving us with 89 samples. As we can see, most students from both genders do enjoy playing videogames while only 22.5% dislike them. Even though there where more male participants, from the results we observe that males are more likely to like videogames. This can be attributed to the stigma of the typical "gamer" who is portrayed as a male rather than a female in many societies. Hence, making it harder for women to adopt this activity and enjoying it especially when it becomes male dominated.

Figure 8:

```
Cell Contents
|-------------------------|
|                       N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  89


               | data.edit$sex
data.edit$like |    Female |      Male | Row Total |
---------------|-----------|-----------|-----------|
       Dislike |        12 |         8 |        20 |
               |     1.402 |     1.045 |           |
               |     0.600 |     0.400 |     0.225 |
               |     0.316 |     0.157 |           |
               |     0.135 |     0.090 |           |
---------------|-----------|-----------|-----------|
          like |        26 |        43 |        69 |
               |     0.407 |     0.303 |           |
               |     0.377 |     0.623 |     0.775 |
               |     0.684 |     0.843 |           |
               |     0.292 |     0.483 |           |
---------------|-----------|-----------|-----------|
  Column Total |        38 |        51 |        89 |
               |     0.427 |     0.573 |           |
---------------|-----------|-----------|-----------|
```



<u>*Comparison 2: Like vs Work*</u>

Even though we briefly covered this topic in scenario 4, we felt these figures would provide a clearer picture of how work may influence student's perspective on videogames. Again, we can see a lower percentage of students who work dislike videogames despite both work proportions being almost equal. For instance, only 14.3% of students who worked disliked videogames in comparison to 31.8% of those who did not work. This is also illustrated in *figure 10*, which seems to support the claim that those who work more hours are less likely to dislike videogames. There is a general satisfaction toward playing videogames regardless of working part-time or full-time hours.

Figure 9::

```
Total Observations in Table:  86


               | data.edit$work
data.edit$like |   No Work |      Work | Row Total |
---------------|-----------|-----------|-----------|
       Dislike |        14 |         6 |        20 |
               |     1.387 |     1.453 |           |
               |     0.700 |     0.300 |     0.233 |
               |     0.318 |     0.143 |           |
               |     0.163 |     0.070 |           |
---------------|-----------|-----------|-----------|
          like |        30 |        36 |        66 |
               |     0.420 |     0.440 |           |
               |     0.455 |     0.545 |     0.767 |
               |     0.682 |     0.857 |           |
               |     0.349 |     0.419 |           |
---------------|-----------|-----------|-----------|
  Column Total |        44 |        42 |        86 |
               |     0.512 |     0.488 |           |
---------------|-----------|-----------|-----------|
```

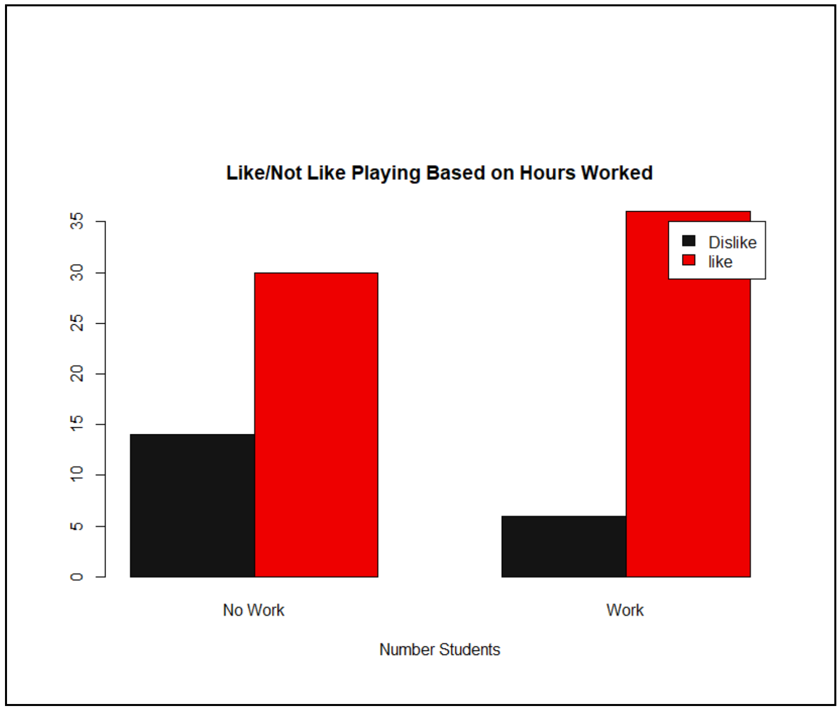
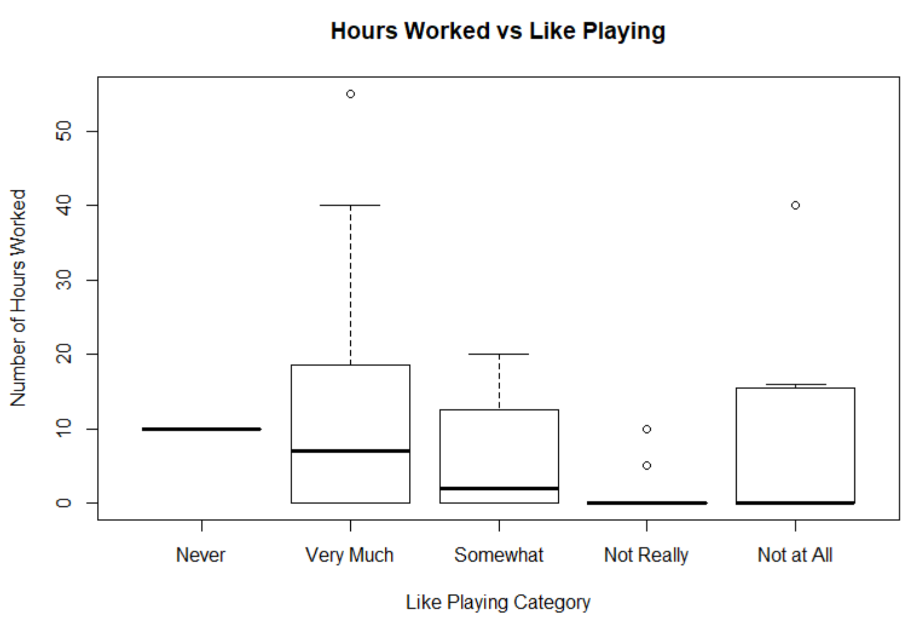


Figure 10: Boxplot



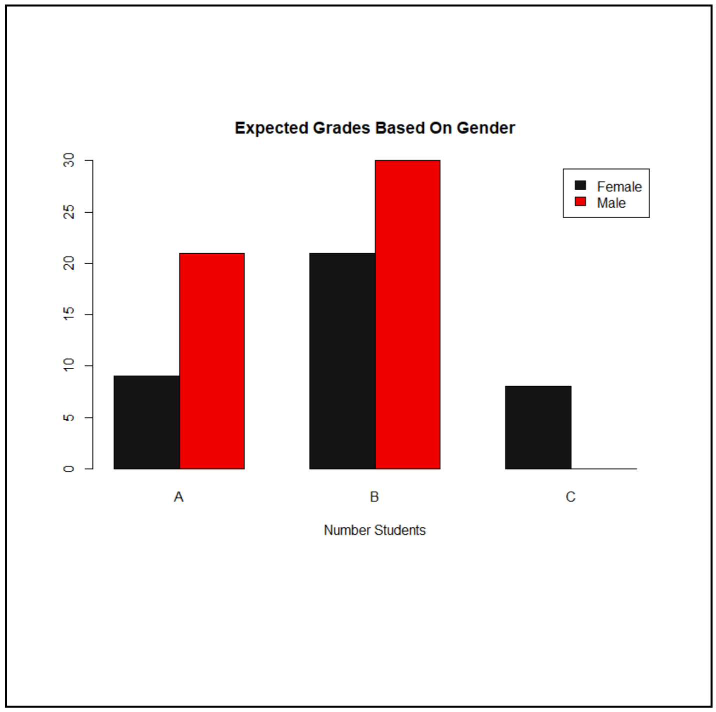
Comparison 3: Gender vs Grades Expected

Lastly, we will see the expected grade distribution among males and females. Because we want to focus on students who played videogames at least once, we excluded the student who never played, leaving us with 89 samples. The purpose of this comparison was to see any possible correlation between expected grades and gender, considering we already compared gender with likeness of videogames. To be clear, none of the students in this sample answered that they expected a D or F. As we saw in *Comparison 1*, males where more likely to like videogames when compared to women. In this comparison we can see that males have higher grade expectations when compared to females as 70% of students expecting an A where men while 30% where women. This may entail that even though males are more likely to play more frequently, it does not affect their hope of obtaining a good grade. On the contrary, based on results found in *Scenario 2* we can conclude that those females expecting a C where also “Daily” videogame players. Their average time spent playing was 14 hours prior to the exam, meaning they encompass the group that like videogames. In other words, all females who disliked videogames expected a B or A in the course. As a result, we can state that playing videogames may have a deeper negative effect on female course expectations than males.

Figure 11:

```
Total Observations in Table:  89


              | data.edit$grade
data.edit$sex |         A |         B |         C | Row Total |
--------------|-----------|-----------|-----------|-----------|
       Female |         9 |        21 |         8 |        38 |
              |     1.133 |     0.028 |     6.153 |           |
              |     0.237 |     0.553 |     0.211 |     0.427 |
              |     0.300 |     0.412 |     1.000 |           |
              |     0.101 |     0.236 |     0.090 |           |
--------------|-----------|-----------|-----------|-----------|
         Male |        21 |        30 |         0 |        51 |
              |     0.844 |     0.021 |     4.584 |           |
              |     0.412 |     0.588 |     0.000 |     0.573 |
              |     0.700 |     0.588 |     0.000 |           |
              |     0.236 |     0.337 |     0.000 |           |
--------------|-----------|-----------|-----------|-----------|
 Column Total |        30 |        51 |         8 |        89 |
              |     0.337 |     0.573 |     0.090 |           |
--------------|-----------|-----------|-----------|-----------|
```



Theory

Central Limit Theorem
Let X1, X2, . . . , Xn be an i.i.d. random sample from a distribution with mean μ and standard deviation σ.

Bootstrapping
Method used on finite samples to improve accuracy among estimations. This is done by iteratively resampling a dataset with replacement. Meaning sample of size n is taken from a dataset and resampled until a variety of combinations are created with the original dataset properties. There statistical calculations, that include mean, median and standard deviation, are done on each sample. The user can choose any B iterations where B number of samples will be created. Then a sampling distribution can be used to visualize each sample mean and calculate a new overall mean as well.

Regression Trees
Type of decision tree focused on continuous target variables. Decision trees is a predictive modeling tool used to make decisions based on their possible consequences. The internal nodes represent a trial done on an attribute and the branch the outcome of that trial. The properties of decision trees include, root node, splitting, decision node(sub-node), leaf(non-split node) and branch(entire sub-section of tree).

Boxplots
The box plot displays the distribution of the data. It is separtaed into first quartile, median, third quartile, and maximum with each approximately being 25%.

Interpretation/Discussion

After doing several analyzes on the data, just as scenario 1-5 demonstrated, we can interpret several significant factors that would be useful for a lab design. One of the main factors that may have affected the outcome of the lab results was running the survey a week before the test, considering how lightly skewed left the data turned out to be when illustrating average time spent playing video games. Meaning students are less likely to spend time playing video games or would only play for an average of about 1.1 hours. In order to aid students with course material but not make them feel like it's another math session, these labs need to focus on making the gaming experience interactive. Have students work together in groups, both males and females, so labs could have videogame aspects such as feedback, shared knowledge and teamwork. As mentioned in this paper, students see in videogames relaxation, hence reducing overloads of course material per session would be crucial in maintaining student's interest. Also just like in videogames where individual know consistency will make them better players, labs should incorporate this and encourage constant practice of math, so students become better mathematicians. We have demonstrated that students who expect higher grades tend to decrease their playing time, hence if they have an upcoming exam they are discouraged to engage in the games and opt to study. What these labs need to focus is on demonstrating that by coming into lab and actively participating in the activity, grades can be increased. For instance, these labs should serve as the solution for someone who just worked 6 hours but really wants to get an A in the course and must choose between studying or relaxing with videogames.

Conclusion

In the first section we found the average time played by only those students who had previously played videogames. We used the Central Limit Theorem to get a general perspective of the distribution and used bootstrapping to tune our average and confidence interval. The end result was 0.36 hours played with 95% confidence that the average was within 0.28 to 0.43 hours. In scenario three we performed the same procedure but, with the entire dataset. We calculated that students played an average of 1.24 hours with 95% confidence that it was within 0.58 to 1.81 hours. For scenario 2 and 5 we compared several of the attributes from the survey to find any potential trends as to why some students enjoyed videogames more than others. Some factors dealt with how often a student plays, grade expected in the course, hours worked and gender. One of the most interesting findings was that people who worked more tended to play videogames more. This claim was also visualized in scenario 3 where depending on hours worked, determined if you liked videogames or not. This in fact helped us understand the percentage distribution behind the posterior surveys asking students to name reasons why they liked or disliked videogames.

Overall 91 samples are not enough to make statistical accurate hypothesis about a general student population. In order to improve the accuracy of these findings, replicating the experiment with a different random sample from either the same 314 students or a different group would be ideal. This will help us get a better view of different trends present in this sample group and determine if they are representative of the student body. Also expanding the population sample to different courses non-related to math would be an interesting approach.