# ANALYSIS OF PLAYING VIDEO GAMES AND STUDENT LEARNING

**Math 189/289C**

**Homework 2 Report**

**Feb. 7th, 2017**

By:
Emma Roth, 4th Year Computer Science/Bioinformatics Major: Section 1, Introduction
Ileena Mitra, 1st year Bioinformatics PhD: Section 6, Hypothesis, Conclusion
Megan Lee, 3rd Year Mathematics/Economics Major: Section 2, Introduction
Jing Gu, 2nd year Chemistry MS: Section 3
Keven Nguyen, 4th Year Mathematics/Economics Major: Section 4
Adithya Bharadwaj Balaji, 1st year Graduate student in Electrical and Computer Engineering: Section 5

# INTRODUCTION

Traditional discussion classes predominantly involve what is called passive learning, in which students listen to the instructor for the duration of class and internalize the information they receive. The current educational debate often focuses on this issue and as a result, active learning has arisen as a common teaching method in the classroom. Active learning is rooted in student engagement in learning activities, such as debates, group projects, and simulation activities. Professor Scott Freeman of the University of Washington, Seattle led a study to analyze the differences in undergraduate teaching methods in STEM classes. The study found that implementing active learning, "reduced failure rates and boosted scores on exams by almost one-half a standard deviation" (Bajak 2014). Professor Michael Prince of Bucknell University examined the effectiveness of active learning, and found that students retain more information when activities are simulated in the classroom. Furthermore, student attitude and critical thinking skills have been linked to teaching methods that utilize problem based learning (Prince 2004).

Game-based active learning is not a new concept. As the popularity of video games amongst youths has increased in recent decades, the potential for their use in the learning environment has drawn great attention. Much research has been done on the application of skills developed from playing video games to education because video games are highly engaging and entertaining and require problem-solving and critical thinking. Kurt D. Squire in his 2008 paper writes that video games are excellent for education because players get instantaneous consequences for the choices they make, and games can be customized to the specific wants and needs of the player (Squire 2008). This individualization is another key draw of video games for education as it allows users to explore "above and beyond the boundaries of a given material….[and] allows the student to become a self-reliant learner" (Taradi 2005).

In an article for Education and Health Journal, Dr. Mark Griffiths of Nottingham Trent University examined the ways in which the positive aspects of playing video games could be extrapolated and utilized in a learning environment. He notes that the mass variety of types of video games attracts a wide demographic of students and that the diversity of video games can help teachers "measure performance on a very wide variety of tasks, and can be easily changed, standardized and understood" (Griffiths 47). Overall, Griffiths concludes that the negative aspects of video games, in particular the violent nature present in a number of popular video games, can be combated by careful design and utilized in a thoughtful learning environment in balance with effective traditional learning methods.

Video games for education have shown promise in practice. A highly successful Stanford University study examined the improvement of math skills in 3rd graders using the mobile math game "Wuzzit Trouble" and found that those who played the game in addition to attending regular lectures given by the teacher showed considerable improvement in understanding concepts over students who only received lectures (Pope and Mangram, 2015). Additionally, a University of Minnesota study published in 2010 studied probabilistic inference skills in people who played video games regularly and people who never played video games found that not only did video game players exhibit higher probabilistic inference, but also that giving non-video game players 50 hours of video game training produced qualitatively similar results to regular video game players when testing for probabilistic inference (Green 2010).

It is widely acknowledged that there are neurological differences between females and males and that these differences cause females and males to approach learning in different ways. A study done by Angela Josette Magon for the Department of Educational Psychology and Leadership Studies at the University of Victoria sought to understand how female and male brains develop differently, and how those differences impact learning processes. She notes that the scientific community has conducted extensive research throughout the past decades and has concluded that, in general, female brains are more adept to process language activities while male brains are better at activities involving spatial-mechanical and motor skills. In particular, these specialties lead boys to "gravitate toward physical activities and video games" (Magon 11). However, Magon notes that the brain's neuroplasticity, its ability to change, shows that through training the brain can adapt to function in different ways.

In this report we analyze data, collected from a survey, from college-age students taking an introductory statistics class at UC Berkeley to determine if introducing a games-based statistics lab will be beneficial to all students. We use a variety of statistical methods to extend estimates about the potential characteristics a lab based discussion could offer to students' learning experience. Based on our evidence, we conclude that a game-based statistics lab has the potential to be a more academically engaging than the traditional lecture and positively change students' attitude towards math.

## HYPOTHESIS

Through surveying a sample of statistics students, we aim to gather information about the best learning methods for students. We plan to use the information gathered from surveying statistics to design a better computer lab class. student Our hypothesis is that an interactive, games-based statistics lab will positively promote learning statistics concepts for all students. In general, we believe all students will benefit from interactive learning.

## DATA

Data was collected from a population of undergraduate students enrolled in Introductory Probability and Statistics, Section 1, at University of California, Berkeley during Fall Semester, 1994. Introductory Probability and Statistics was a lower-division prerequisite for students intending to major in Business. The list of students who had taken the second exam of the semester was used to select the students to be surveyed. This ensured our sample contained students who were invested in the class and intended to remain enrolled. Data was collected in both the Tuesday and Thursday meetings of the discussion sections the week after the second exam, and students who had not been reached during the discussion section were located on Friday in lecture. The second exam was returned the week the data was collected, allowing students to assess how well they were doing in the class. Responses remained anonymous to ensure honesty.

Our data reflects analysis of the sample of 91 students that completed the survey. To choose 95 students for the study, each student in the class was assigned a number between 1 and 314. A pseudo random number generator selected 95 number within this range.

The data collected from the questionnaire examined the following fifteen variables:

*Table 1: Description of variables in the dataset.*

| Variable | Description | Variable Type | Numerical Type |
|---|---|---|---|
| Time | # of hours played in the week prior to the survey | Numerical | Continuous |
| Like to play | 1=never played, 2=very much, 3=somewhat, 4=not really, 5=not at all | Categorical | N/A |
| 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 | N/A |
| How often | 1=daily, 2=weekly, 3=monthly, 4=semesterly | Categorical | N/A |
| Play if busy | 1=yes, 0=no | Categorical | N/A |
| Playing educational | 1=yes, 0=no | Categorical | N/A |
| Sex | 1=male, 0=female | Categorical | N/A |
| Age | Students age in years | Numerical | Discrete |
| Computer at home | 1=yes, 0=no | Categorical | N/A |
| Hate math | 1=yes, 0=no | Categorical | N/A |
| Work | # of hours worked the week prior to the survey | Numerical | Continuous |

| | | | |
|---|---|---|---|
| Own PC | 1=yes, 0=no | Categorical | N/A |
| PC has CD-Rom | 1=yes, 0=no | Categorical | N/A |
| Have email | 1=yes, 0=no | Categorical | N/A |
| Grade expected | 4=A,3=B,2=C,1=D,0=F | Categorical | N/A |

Questions that were unanswered or improperly answered were coded as 99. For our report, we removed all data entries coded as 99. Students who reported to never play video games or to not not like video games were instructed to skip several questions.

## THEORY

In statistics, a **confidence interval** (**CI**) is a type of interval estimate of a population parameter. It is an observed interval within which the estimate lies with the specified degree of probability.

According to the central limit theorem, if $X_1, ...., X_n$ are independent and identically distributed with mean $\mu$ and variance $\sigma^2$ then for large values of n, the probability distribution

$$Z = \frac{\overline{X} - \mu}{\sigma}\sqrt{n}$$

is approximately normal.

But in our case we use random sampling, the samples are identically distributed but they are not independent. But the normal approximation holds if the sample size is large but not too large relative to the population size.

The normal distribution is used to provide the confidence intervals for the population parameter. The 68% confidence interval is given by,

$$\left(\overline{x} - \frac{\sigma}{\sqrt{n}}, \overline{x} + \frac{\sigma}{\sqrt{n}}\right)$$

and the 95% confidence interval is given by,

$$\left(\overline{x} - 2\frac{\sigma}{\sqrt{n}}, \overline{x} + 2\frac{\sigma}{\sqrt{n}}\right)$$

In practice since we don't know the value of $\sigma$, we find an estimate s and substitute in place of $\sigma$. The confidence interval obtained in this case is called the approximate confidence interval.

Let $x_i$ be the time spent playing videogames by student i, where i=1 to 314.

The population average is given by,

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

which is our population parameter.

We then use $X_{I(j)}$ to represent the values of the characteristic for the jth unit sampled,

$$X_{I(j)} = x_i,\ if\ I(j) = i$$

$$X_{I(j)} = 0,\ if\ I(j)\ is\ otherwise$$

$X_{I(j)}$ represents the value of the time spent playing video games by the jth unit sampled, j=1,2,3,...,91.

$$E(X_{I(j)}) = \sum X_{I(j)} P(X_{I(j)})$$

$$E(X_{I(j)}) = \sum_{i=1}^{N} 0.P(otherwise) + \sum_{i=1}^{N} x_i.P(X_{I(j)})$$

$$= \sum_{i=1}^{N} x_i.P(I(j) = i) \ = \frac{1}{N}\sum_{i=1}^{N} x_i = \mu$$

The sample average is the sample statistic that estimates the population parameter $\mu$,

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{n} X_{I(j)}$$

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{n} \frac{n}{N} x_i = \frac{1}{N}\sum_{i=1}^{N} x_i = \mu$$

Hence, the sample average is an unbiased estimator of the population parameter.

Bootstrap simulation can be helpful to obtain an approximate interval estimate for the estimate of a parameter of interest with unknown distribution. When the sample size for each round of simulation increases, the empirical distribution from bootstrap estimates would converge to the unknown distribution of the estimates. Bootstrap also applies to finite samples and finite population.

## RESULTS

### Section 1

In this section, we examine the fraction of students that played video games in the week prior to data collection. It is important to note that the second exam of the semester was also during the week prior to data collection.

The percentages were calculated by finding the number of students within the specified population who reported playing time greater than 0 and dividing that by the total number of individuals in that population (See Appendix 1.1). The population "Students that like video games" contains all students whose response to the question concerning "like to play" was 2, "very much". Similarly, the population "Students that somewhat like video games" contains all students whose response to the same question was 3, "somewhat". Students who responded 1, "never played", 4, "not really", or 5, "not at all" to "like to play" were excluded from these calculations because they were expected to have a very low, if not non-existent, fractions of students that actually played video games.

Table 2. displays the point estimates for fraction of students who played video games in the week prior to data collection.

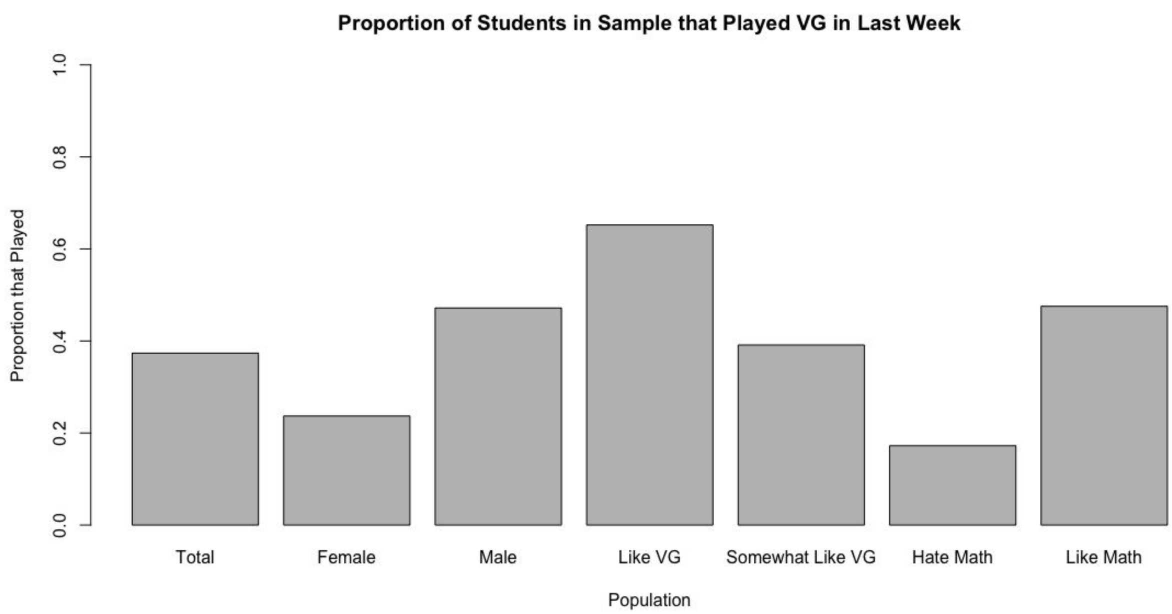
*Table 2: Point Estimates of Fractions of Students that Played Video Games*

| **Population** | **% Played Video Games Week Prior to Study** |
|---|---|
| All students | 37.36 |
| Male students | 47.17 |
| Female students | 23.68 |

| | |
|---|---|
| Students that like video games | 65.21 |
| Students that somewhat like video games | 39.13 |
| Students that hate math | 17.24 |
| Students that like math | 47.54 |

Figure 1, below, represents data from Table 2 in a bar plot.

*Figure 1: Bar Plot of Estimated Proportions*




It's quite clear from Figure 1 that certain subpopulations had a higher fraction of students who played video games than other subpopulations. Less than half, about 37%, of all student respondents played video games during the week of the exam. It is unsurprising that those who reported they liked video games had the highest proportion of students that reported playing, around 65%. It is important to note that the fraction of males who played in the week prior is nearly double the fraction of females who played, and that the fraction of students who reported they liked math and played is nearly three times the fraction of students who hate math and played.

Below, Table 3 gives the 95% confidence interval of fraction of each group of students that played in the week prior. Confidence intervals assert that if the population of students was sampled many times and an interval estimate was taken each time, the resulting intervals would accurately represent the population in 95% of all cases. This calculation was obtained by adding and subtracting the standard error of each population multiplied by the finite population correction factor to each percentage point estimate (Appendix 1.2). The finite population correction factor was used in this case because the sampling fraction $\frac{size\ of\ sample(n)}{size\ of\ population(N)}$ for this data set is not considered small relative to the population we are studying (a generally accepted definition of “small” sample is less than 5% of the population, and our sample of 91 students represents roughly 29% of the population of Intro to Statistics students).

*Table 3: 95% Confidence Interval for Fraction of Students who Played Video Games*

| **Population** | **Interval Estimate (%)** |
|---|---|
| All students | (28.63017, 46.09511) |
| Male students | (38.15878,56.18085) |
| Female students | (16.00992, 31.35850) |
| Students that like video games | (56.62005, 73.81474) |
| Students that somewhat like video games | (30.32078 47.94009) |
| Students that hate math | (10.42280 24.05996) |
| Students that like math | (38.52640 56.55557) |

These confidence intervals are a way to estimate the responses of all 314 students in the class without having to obtain responses from all students in the class. Below is a graphical representation of the data in Table 3.

*Figure 2: Plot of Fraction of Video Game Players and 95% Confidence Intervals*

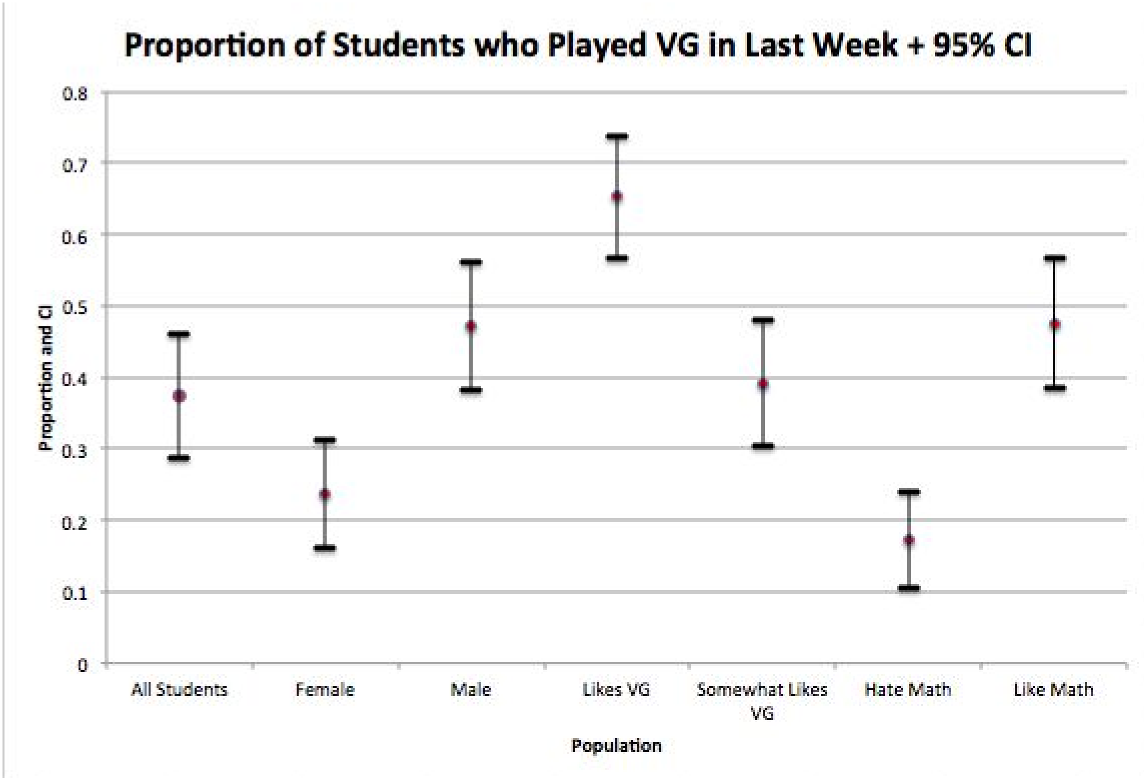



Figure 2 describes 95% confidence intervals for each of the subpopulations. Analyzing confidence interviews gives a sense of how stable the data is. It shows that of all populations being considered, female students and students who hate math appear to have the tightest 95% confidence intervals and are the most stable. The rest of the populations appear to have confidence intervals of very similar range.

The point estimates and confidence intervals together suggest that different groups of students have a varying number of people who play videogames and that when considering how to implement this lab, we must take into account that this surveyed sample as well as the greater population has very many wants, needs, and preferences. However, simply the fraction of each demographic that plays video games is not enough information, further exploration into what draws and deters students from playing video games is necessary for a thorough assessment of whether a game-based lab is appropriate for UC Berkeley Introduction to Probability and Statistics students.  For example, this analysis clearly shows that more male students play video games than female students and many more students who like math play video games than students who do not like math, but this is not directly applicable to our lab until we investigate the correlations and causations underlying these differences. The next sections aim to take a deeper look into these reasons in the hopes of identifying the possibility of creating a lab environment that engages all students.

## Section 2

This section aims to compare the difference between time spent playing video games in the week prior to the survey and reported frequency of game play. We also have taken into consideration differences in attitude towards video games in math between males and females due to the current literature that acknowledges a learning bias in different genders. Since there was an exam in the week prior to the survey, the number of hours recorded for being played may be lower than usual since students may have dedicated more time to studying rather than playing video games. Examining the variables time and frequency provides the following results:

| Time | Frequency |
|---|---|
| Min. : 0.000 | Min. :1.000 |
| 1st Qu.: 0.000 | 1st Qu.:2.000 |
| Median : 0.000 | Median :3.000 |
| Mean : 1.243 | Mean :2.705 |
| 3rd Qu.: 1.250 | 3rd Qu.:4.000 |
| Max. :30.000 | Max. :4.000 |
| | NA's :13 |

The above summary statistics of the time and frequency reveal the average student played video games on about a monthly basis and averaged playing video games for about 1.2 hours in the week prior to the survey. This difference may be accounted for by the exam held in the week prior to the survey. Due to the exam, students likely decreased their video game play time in favor of other activities, likely studying. In terms of the previous estimates, which sought to capture the fraction of students who played a video game in the week prior to the survey, the estimates likely would increase if there were no exam in the week prior to the survey. The time reported would probably increase to a greater average and be more reflective of the reported frequency of playing on a weekly basis.

*Figure 3: Plot of Hours of Video Game Play Reported and Typical Frequency of Play*

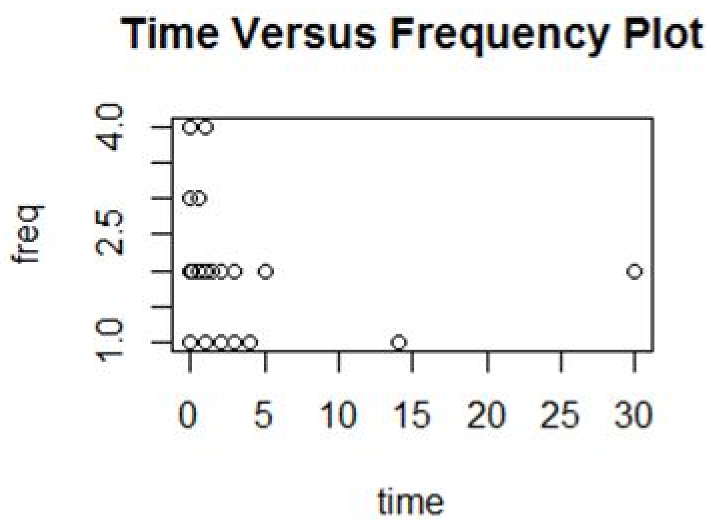



Figure 3 plots the reported hours of video games played by students in the week prior to the exam in comparison to their reported frequency of game play, measured on a scale of one to four (one being daily, two being weekly, three being monthly, and four being semesterly). The graph shows that there was a high concentration of students, regardless of how frequently they report to play video games, who spent between zero and five hours playing that week. This again likely reflects the exam in the week the data was being collected as students probably had less free time to play video games.

Several studies have shown that males and females play video games and learn from them in different ways. Analyzing the differences in the survey data between males and females could provide the designers of the lab with significant factors to keep in mind in creating a learning environment that benefits the maximum number of students. Comparing the means of relevant variables between males to females provides the following summary statistics:

*Table 4: Summary Statistics Comparing Attitude Towards Video Games Between Males and Females*

| | Males | Females |
|---|---|---|
| **Time** (0-30 hrs) | 1.596 | 0.75 |
| **Like** (0-5) (1=never played, 2=very much, 3=somewhat, 4=not really, 5=not at all) | 2.827 | 3.289 |
| **Frequency** (1-4) (1=daily, 2=weekly, 3=monthly, 4=semesterly) | 2.468 | 3.065 |
| **Play if Busy** (0-1) (1=yes, 0=no) | 0.2708 | 0.125 |
| **Hate Math** (0-1) (1=yes, 0=no) | 0.2855 | 0.3684 |

Table 4 reveals several pertinent facts to the design of a potential new discussion lab. In addition to analyzing the variables of time and frequency, it is also worthwhile to examine the variables of liking video games, playing video games if busy, and hating math. These additional variables can help the designers of the lab further understand how students would respond to a lab setting versus a traditional discussion setting in a math class. The survey data shows that females reported to have spent nearly half the time playing games as males reported in the week prior to the survey. Also, regarding frequency, females reported to playing games on a less frequent basis than did males, though both average close to a monthly basis. This could indicate that females do not prioritize video games during exam periods, and thus may not liken video games to intellectual learning. Females also indicated that they enjoy playing video games on a scale between somewhat and not really, while males reported enjoying video games on a level between very much and somewhat. Males were also more inclined than females to report playing video games when busy, which likely affected exam results. Females reported having a stronger dislike for math than males which also likely affected exam results. Data from these various factors indicates that the design of a new computer lab for the purpose of providing an interactive learning environment would most likely benefit the learning experience of males more so than females.

## Section 3

We further investigated into the average amount of time spent playing video games in the week prior to the survey. First of all, the unbiased estimator for that parameter of interest should be sample mean, which is 1.24 hour. Then, we need to find the distribution of sample mean based on the unknown probability histogram of population. As the overall sample distribution shown in figure 4 is not bell-shaped like a normal distribution but extremely skewed, it is doubtable that the probability histogram of the sample average can follow a normal curve. In addition, the survey data are identically but not independently distributed. Therefore, central limit theorem cannot be applied to find confidence interval. However, bootstrap simulation can be helpful to obtain an approximate interval estimate for the average amount of time spent on playing video games.

The general procedure for a bootstrap simulation is as follows:

1) Create a bootstrap population by repeatedly sampling each unit in the sample (91) until it reaches the population size (314).
2) Select 91 random sample from bootstrap population
3) Compute for the sample average
4) Repeat step 2 and 3 for 1000 times to obtain the probability distribution of the sample average

*Figure 4: Histogram of number of hours played in the week prior to survey for all students in the sample.*

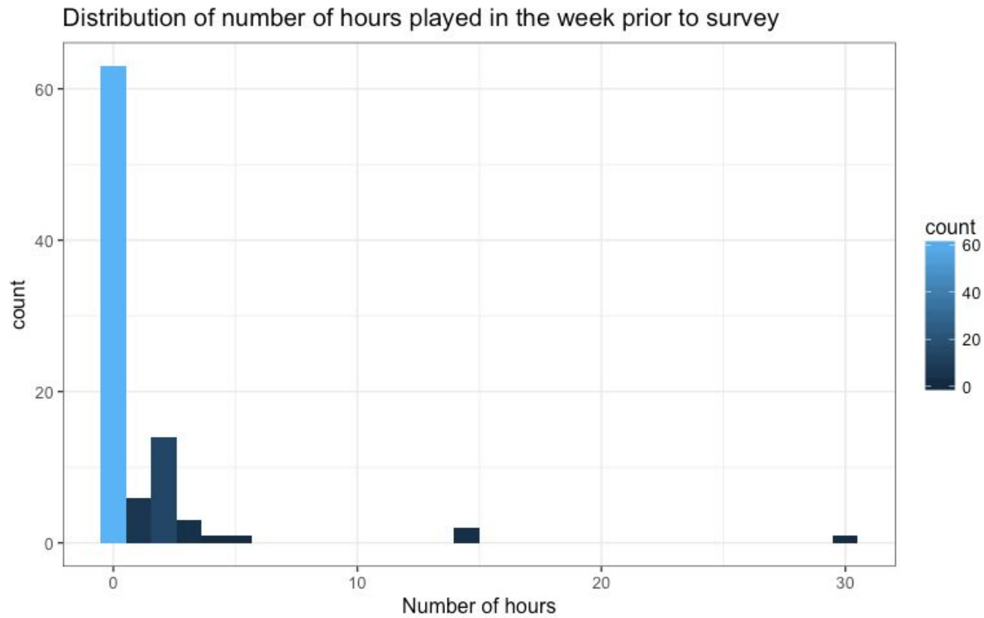



*Figure 5: Histogram of sample averages from 1000 simulated bootstrap sample*

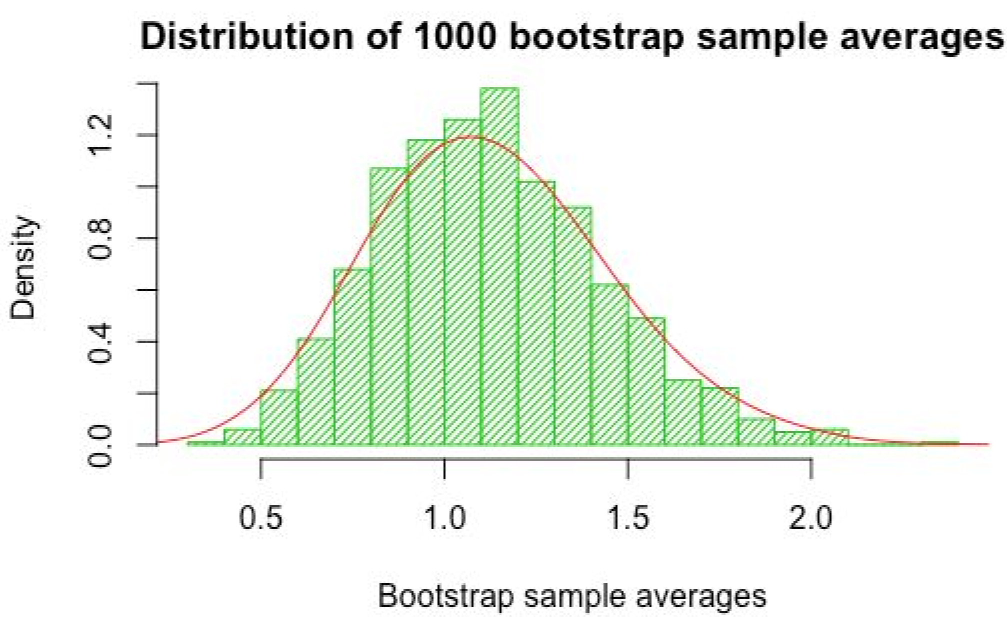
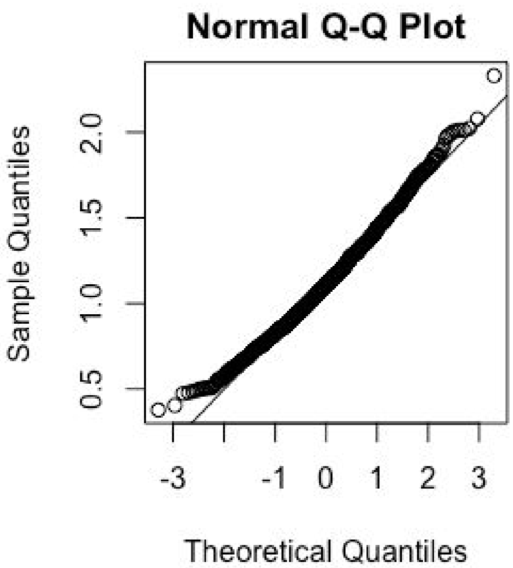


*Figure 6: Quantile-quantile plot used to compare bootstrap sample averages with normal distribution*




A histogram of the bootstrap sample averages for all students in the sample of 91 was shown in Figure 5, which looks like a bell shape. To compute for 95% confidence interval, we first checked for normality. From previous lectures, we learned that we can compare kurtosis and skewness coefficients of samples and simulated sample from normal distribution to better determine if a distribution is normally distributed. The bootstrap sample averages have kurtosis coefficient as 2.569 and skewness as 0.393, which are away from the coefficients distributions that were simulated from normal distribution (shown in fig7). The quantile-quantile plot in Figure 6 also shows that there is some extent of deviation between sample quantiles and normal quantiles. Therefore, even the simulated bootstrap sample averages do not follow a normal curve. Instead, we simply used the 2.5th and 97.5th percentile of the bootstrap distribution as the 95% confidence interval (CI). As the empirical distribution converges to the unknown distribution of estimates , the correct CI for the true parameter (average time) for the empirical distribution should also converge to that for the unknown distribution of estimates(ref1).

*Figure 7: Distributions for both kurtosis and skewness values of 1000 simulated samples from normal distribution.*

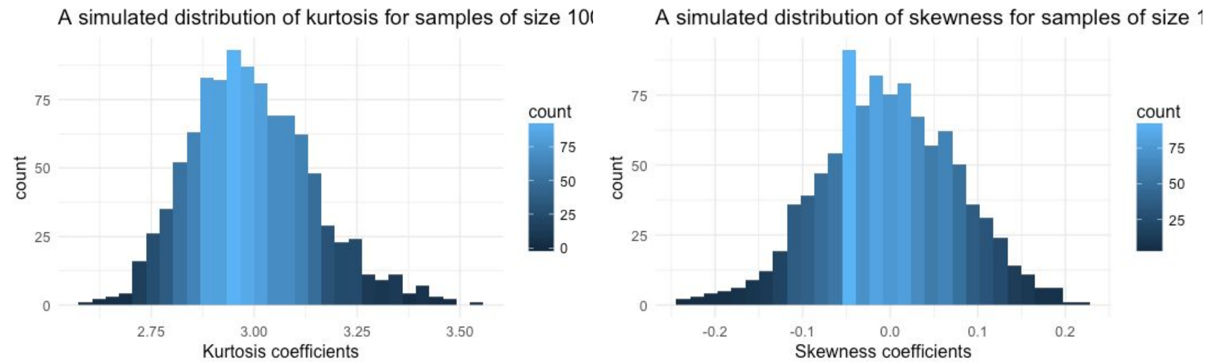



*Table 5: Bootstrap Estimates and 95% confidence interval for all students and their comparisons in students' sex and whether or not they like playing video games or math.*

| **Population** | **estimate(hrs)** | **Interval Estimate (hrs)** |
|---|---|---|
| All students | 1.243 | [0.579 1.783] |
| Male students | 1.596 | [0.681, 2.821] |
| Female students | 0.750 | [0.184, 1.487] |
| Students that like video games | 1.632 | [0.793, 2.384] |
| Students that do not like playing video games | 0.024 | [0.000, 0.071] |
| Students that like math | 1.172 | [0.052, 3.259] |
| Students that hate math | 1.297 | [0.749, 1.952] |

The average amount of time spent playing video games in the week prior to the survey based on 91 sample was determined to be 1.2 hours, with a confidence interval of [0.579, 1.783] hours. This tells us that even though students have a second midterm in the week before the survey, in average they spent 1.2 hours on playing video games. However, this may be biased by the way how data were sampled, such as sex of individuals, attitudes towards video games and etc. Additional variables were investigated to better make a conclusion. As expected, Table 5 shows that male students did spent almost double amount of hours playing video games than female students and student who like playing video games also spent more hours than those who do not like or never play. It is interesting to see that the average time spent on playing video games during exam week does not seem to be affected by whether students like math or not.

*Table 6: For students who like or somewhat like playing video games, this tables shows comparison of bootstrap estimates and 95% confidence interval between student that fall in following categories.*

| **Population of students who like or somewhat like playing video games** | **Estimate (hrs)** | **Interval Estimate (hrs)** |
|---|---|---|
| Students who claimed to play when busy | 4.638 | [1.735, 8.353] |
| Students who claimed not to play when busy | 0.617 | [0.370, 0.890] |
| Students who play for educational purpose | 1.388 | [0.751, 2.274] |
| Students who do not play for educational purpose | 2.020 | [0.452, 4.275] |
| Students who worked for less than five hours during week prior to survey | 1.927 | [0.559, 3.839] |
| Students who worked for more than five hours during week prior to survey | 1.391 | [0.703, 2.312] |

There may also be other confounding factors that contribute to final result for students who have a positive attitude towards video games. From Table 6, students who claim to play when busy did spend a few times higher average time on playing video games during exam week. Other two variables do not provide much information. It need to be noticed that bootstrap simulation may not be accurate for small sample size, so some of the confidence intervals may be biased. To sum up, the average amount of time spent during exam week varies greatly to individuals' sex and attitudes to playing video games. We do not have enough evidence to show that interactive learning may be more appealing to students.

## Section 4

We now investigate whether or not the students like playing video games in general. Naturally, the "attitude" survey provides the exact information needed. Table 7, on reasons as to why students dislike video games, is used first because it is the only subsequent survey that required participation of all 91 students.

*Table 7: Summary of what students didn't like about the games.*

| Reasons to Dislike Video Games | Percent of Total Chosen |
|---|---|
| It is pointless | 33% |
| Boring | 17% |
| Lonely | 6% |
| Too much time | 48% |
| Costs too much | 4% |
| Frustrating | 26% |
| Too many rules | 19% |
| Friend's don't play | 17% |

The cross tabulations are there, but we do not have access to the raw data. Additionally, the survey question was posed as a “Choose up to 3 answers” type of question or a multiple response question. This means that the sum of the `Percent of Total Chosen` column is more than 100%. Without the raw data we can not analyze this survey with depth, but there are still insights to be found. First, we sort the percent column in descending order and add a new feature called `Emotional`.

*Table 8: Ordered summary of dislikes with emotional feature based on contextual knowledge.*

| Dislikes | Percent of Total Chosen | Emotional |
|---|---|---|
| Too much time | 48% | No |
| Costs too much | 40% | No |
| It is pointless | 33% | Yes |
| Frustrating | 26% | Yes |
| Too many rules | 19% | No |

| | | |
|---|---|---|
| Boring | 17% | Yes |
| Friend's don't play | 17% | No |
| Lonely | 6% | Yes |

Immediately we can see a disparity between the top four `Dislikes` and bottom four with respect to the percentage. Although students were able to choose anywhere between one and three reasons, the first four reasons were chosen the most frequently. Moreover, since students are allowed to choose up to three reasons we can safely assume that the average student that answered the survey chose at least one of the top three. Of the three choices, the top two reasons, “Too much time” and “Costs too much”, score a no in the `Emotional` category. A takeaway is that those choices do not necessarily mean the survey respondent disliked playing video games; the choices are aspects of video games they dislike but still overlook when they play games. The third choice is the polar opposite of the first two. “It is pointless” scores a yes in the `Emotional` category. Anyone choosing this reason more often than not will dislike playing video games because a person that likes video games will not find playing video games “pointless” as shown in Table 9.

*Table 9: Summary of reasons for playing the game.*

| **Reason** | **Percent of Total Chosen** |
|---|---|
| Relaxation | 66% |
| Feeling of mastery | 28% |
| Bored | 27% |
| Graphics/Realism | 26% |
| Mental Challenge | 24% |
| Eye/Hand Coordination | 5% |

The main insight here is that from the “attitude” surveys we can feel confident in finding something significant pertaining to the proportion of the students that like or dislike playing video games. We cannot conclude yet that most of the survey respondents enjoy playing video games.

To get closer to a conclusion we will use the raw dataset, “videodata.txt” that contains all of the data for non-attitude questions. The dependent variable of interest is not a feature in the original dataset. As a result, we make a new categorical variable named “dislike”. A student scores a value of 1 if they chose option 1(never played), 4(not really), or 5(not at all). We can also treat option 99(NA) as a dislike because some students chose to not answer that question out of disinterest in video games. We will also want to explore if the expected grade in a class will affect our variable of interest so we create three new dummy columns with respect to the grades A(4), B(3), and C(2). Now that we have our variables of interest, we can implement a model using logistic regression (Logit) with independent variables: busy, educ, sex, home, math, own, cdrom, email, a, and b. We exclude the variable “c” to avoid perfect collinearity. Table 10 shows the regression results.

*Table 10: Regression Results*

**Logistic Regression Results**

| | *Dependent variable:* |
|---|---|
| | dislike |
| sex1 | -0.850 (0.603) |
| | p = 0.159 |
| home1 | 0.989 (1.158) |
| | p = 0.393 |
| math1 | 1.215 (0.588) |
| | p = 0.039 |
| own1 | 1.471 (0.898) |
| | p = 0.102 |
| cdrom1 | -0.331 (0.787) |
| | p = 0.675 |
| email1 | -0.087 (0.739) |
| | p = 0.907 |
| a1 | 16.979 (1,539.551) |
| | p = 0.992 |
| b1 | 17.015 (1,539.551) |
| | p = 0.992 |
| Constant | -19.914 (1,539.551) |
| | p = 0.990 |
| Observations | 85 |
| Log Likelihood | -38.506 |
| Akaike Inf. Crit. | 95.012 |

We can note that the regression deems the variable, math, to be significant at the 5% level. This does not mean much to us currently, but after taking the exponent of all the coefficients we can see how each variable impacts the dependent variable in terms of odds, which is shown in Table 11.

*Table 11: Exponentiated regression coefficients*

| Int | sex1 | home1 | math1 | own1 | cdrom1 | email1 | a1 | b1 |
|---|---|---|---|---|---|---|---|---|
| 2.2e-09 | .427 | **2.688** | **3.371** | **4.353** | .718 | .916 | 2.3e+07 | 2.4e+07 |

If the sex variable gets a value of one, then there are less odds that the student dislikes playing video games. In contrast, if a student has a computer at their home then the odds are 2.6 times greater that that student will dislike playing video games. Liking math is similar in that if it scores a one then the student will have 3.3 times more odds to dislike playing video games. The rest of the odds can be interpreted in a similar way, but if the odds are less than one then we cannot describe it as directly as how we do if the odds are greater than one. To remedy that, we can run the regression again on a new variable, “likeGames”, that is the opposite of “dislike”. This shows us the uninterpretable odds as shown in Table 12 below.

*Table 12: Exponentiated regression coefficients with “likeGames” as dependent variable*

| Int | sex1 | home1 | math1 | own1 | cdrom 1 | email1 | a1 | b1 |
|---|---|---|---|---|---|---|---|---|
| 4.4e+08 | **2.339** | .372 | .296 | .229 | **1.391** | **1.09** | 4.2e-08 | 4.1e-08 |

Now we can interpret “sex”, “cdrom”, “email” the same way like above. We can also see from the regression results that the grade expected is completely off in terms of its standard error and odd ratios.

In summary, the “attitude” analysis and the odds ratio analysis is hinting towards the fact that students like playing video games in general, but the provided “attitude” data is not good enough to come to a conclusion by itself. The odds of each independent variable is useful in that we can see how they affect the dependent variable, “dislike”, in an interpretable way. We still do not have a solid and quantifiable proof that students like videogames in general. In the following sections we follow the trail of breadcrumbs that we have found as a result of the this and the previous sections’ analyses.

## Section 5

The given data can be divided into those who like to play videogames and those who don’t like to play videogames based on the variable “like” which was has 5 options. We can broadly classify the options into two main categories, where option 2(very much) and option 3(somewhat), can be grouped as people who like to play video games. The option 4(not really) and option 5(not at all) can be grouped as people who dislike playing video games. For simplicity purposes, the option 1(never played) and option 99(NA) can be grouped as dislike playing video games.

CLASSIFICATION BASED ON SEX:

The sex of survey respondents can be classified based on the variable “sex”, which has a value 1 for “male” and value 0 for “female”.

*Table 13: Cross-tabulation between liking video games and the sex of a person.*

| | | Like Video games | | |
|---|---|---|---|---|
| | | Yes | No | Total |
| Sex | Male | 43 | 10 | 53 |
| | Female | 26 | 12 | 38 |
| | Total | 69 | 22 | 91 |

*Figure 8: Bar plot showing number of students who like videogames vs sex of the students*

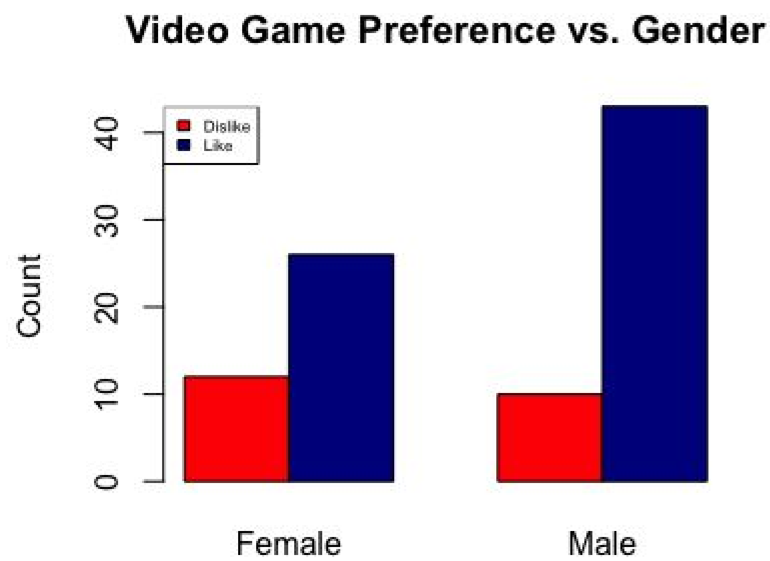



*Table 14: Cross-tabulation between the overall percentage of people who like video games based on their sex.*

| | | Like Video games | |
|---|---|---|---|
| | | Yes | No |
| Sex | Male | 81.13 | 18.87 |
| | Female | 68.42 | 31.58 |

From the cross tabulations and the bar plot, we can observe that the majority of males and females like to play video games. So we cannot substantially associate a particular gender is more likely to play video games compared to the other.

CLASSIFICATION BASED ON OCCUPATIONAL STATUS OF THE PERSON:

Next, we try to classify the respondents into those who work for pay and those who don't work for pay. The people who have answered more than "0" for the number of hours worked last week are grouped as people who work for pay and those who answered "0" for the number of hours worked last week are grouped as people who don't work for pay. Also, people who haven't responded to this question have been classified as people who don't work for pay for computational purposes.

*Table 15: Cross-tabulation between the liking of videogames and occupational status of the person.*

| | | Like Video games | | |
|---|---|---|---|---|
| | | Yes | No | Total |
| Work | Yes | 36 | 8 | 44 |
| | No | 33 | 14 | 47 |
| | Total | 69 | 22 | 91 |

*Figure 9: Bar plot showing number of students who like videogames vs number of the students who work for pay*

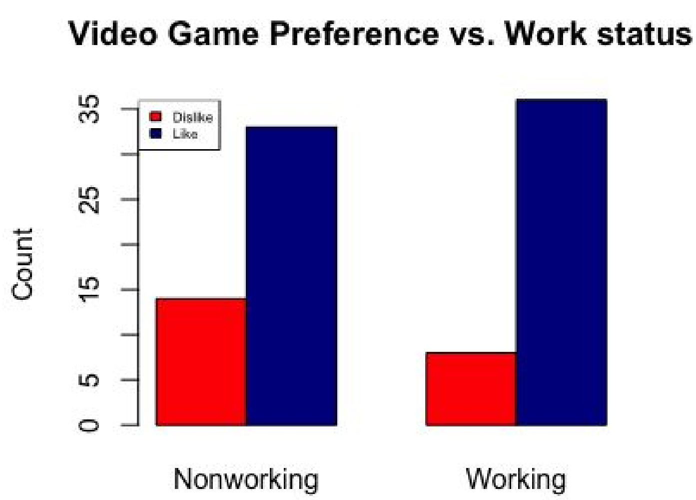



*Table 16: Cross-tabulation between the percentage of people who like video games based*

*on their occupational status.*

| | | Like Video games | |
|---|---|---|---|
| | | Yes | No |
| Work | Yes | 81.82 | 18.18 |
| | No | 70.21 | 29.79 |

From the cross tabulations and the bar plot, we can observe that a majority of the working and the non working students like to play video-games. This shows that the work of the students might affect the number of hours that they play the videogames but doesn't deter their liking towards the videogames.

CLASSIFICATION BASED ON WHETHER THE PERSON OWNS A COMPUTER:

From the survey, the students who own a computer have answered "1" and those who don't own a computer have answered "0". On classifying the students based on this fact, we observe the cross tabulations and bar plot as shown:

*Table 17: Cross-tabulation between the liking of video games and whether a person owns*

*a computer.*

| | | Like Video games | | |
|---|---|---|---|---|
| | | Yes | No | Total |
| Own computer | Yes | 48 | 19 | 67 |
| | No | 21 | 3 | 24 |
| | Total | 69 | 22 | 91 |

*Figure 10: Bar plot showing number of students who like videogames vs number of the students who own a computer.*

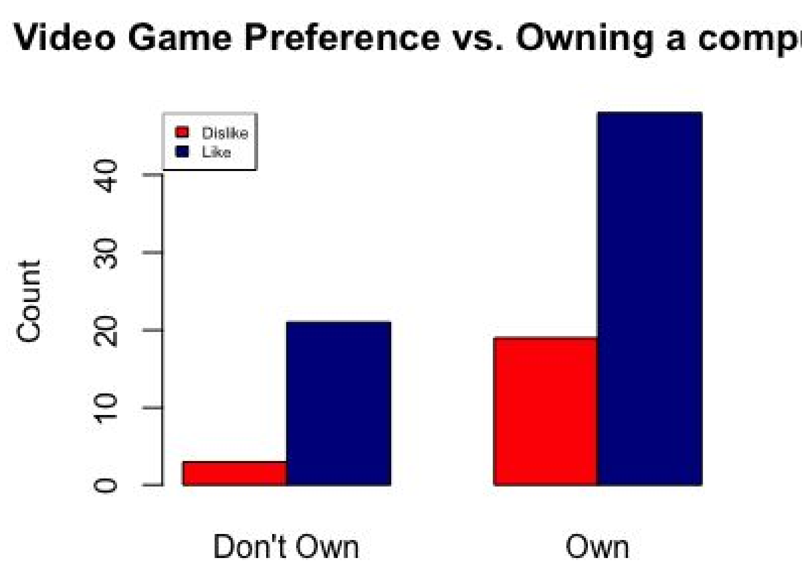



*Table 18: Cross-tabulation between the percentage of people who like video games based*

*on their occupational status.*

| | | Like Video games | |
|---|---|---|---|
| | | Yes | No |
| Own Computer | Yes | 71.64 | 28.36 |
| | No | 87.5 | 12.5 |

From the cross tabulations and the bar plot, once again we can see that the majority of the students who own and those who don't own a computer like to play video games. So, once again we cannot say that owning a computer causes the students to like playing video games more compared to the students those who don't own a computer.

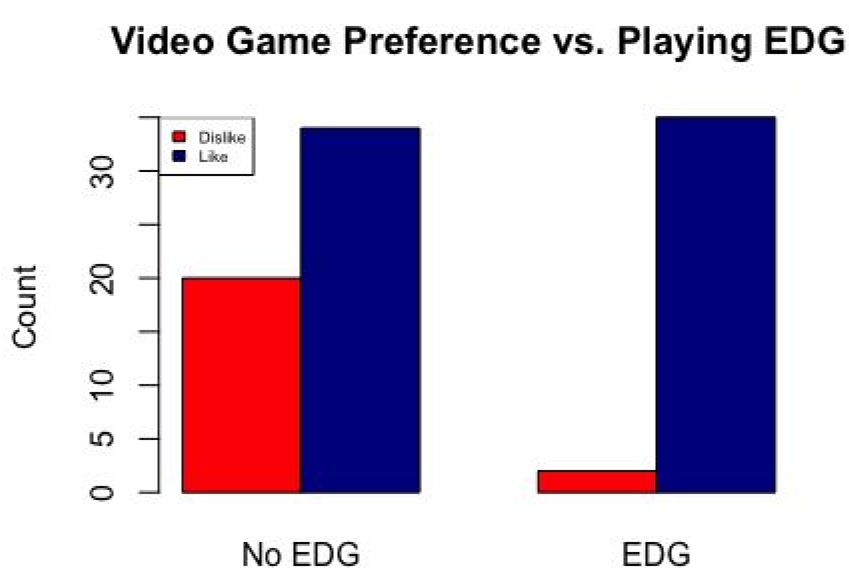
Thus, no particular variable under consideration be it either the gender or the occupational status or whether the person owns a computer can be called as a distinct attribute that causes the students to like playing video games.

Since the main objective of our study is to design mathematical labs which mimic playing video games, we should consider the variable in the survey "educ" which tells whether the person playing video games is actually playing educational games. This is important as the labs that are going to be formed are primarily for educational purposes, so studying this variable gives us a good measure of correlation between the two factors.

*Table 19: Cross-tabulation between the liking of videogames and whether a person plays educational games.*

| | | Like Video games | | |
|---|---|---|---|---|
| | | Yes | No | Total |
| Play Educational Games | Yes | 34 | 2 | 37 |
| | No | 35 | 20 | 55 |
| | Total | 69 | 22 | 91 |

*Figure 11: Bar plot showing number of students who like videogames vs number of the students who play educational games.*



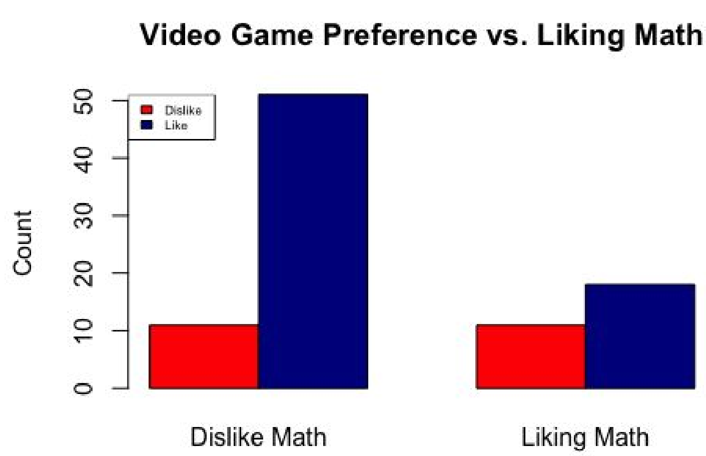
From the table and the bar plot we can observe that of the people who like playing videogames nearly 50% of play educational videogames and the other 50% don't play educational videogames. So designing a lab mimicking video games, don't necessarily guarantee it's success as the material being delivered in educational in nature.

Along the similar lines, considering students who like math and those who don't like math and their relationship with respect to playing video games gives us an exact idea as to how the students would receive a videogame themed lab in comparison to a traditional math lab.

*Table 20: Cross-tabulation between the liking of videogames and liking math.*

| | | Like Video games | | |
|---|---|---|---|---|
| | | Yes | No | Total |
| Like math | Yes | 18 | 11 | 29 |
| | No | 51 | 11 | 62 |
| | Total | 69 | 22 | 91 |

*Figure 12: Bar plot showing number of students who like videogames vs number of the students who like math.*

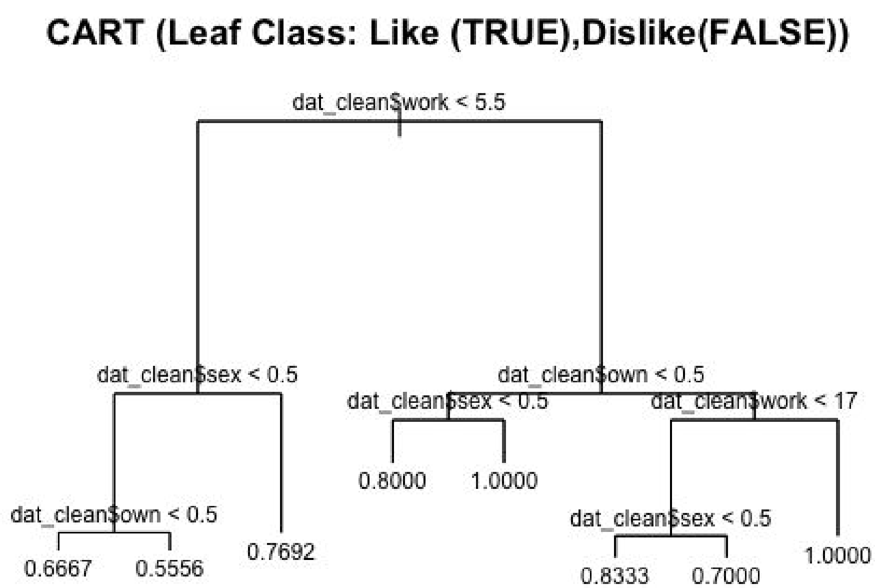


From this table and the bar plot, we can see that majority of the students currently don't like math. So designing a videogame themed lab for math is going to improve the chances of them liking the subject and changing their opinion towards math.

In conclusion, the design of such a lab is a step in the right direction in trying to get everyone in the class involved in the subject and improving their interest in math.

Additionally, when we try to combine the three main variables given in the scenario 5, "sex","own" and "work", using Classification and Regression Tree(CART) technique. We obtain a tree as follows:

*Figure 13: Classification tree based on the variables "sex","own" and "work".*




Summary of Regression tree:

tree(formula = dat_clean$like ~ (dat_clean$own) + (dat_clean$work) + (dat_clean$sex), data = dat_clean, mindev = 0.001)

Number of terminal nodes:  8

Residual mean deviance:  0.1723 = 14.13 / 82

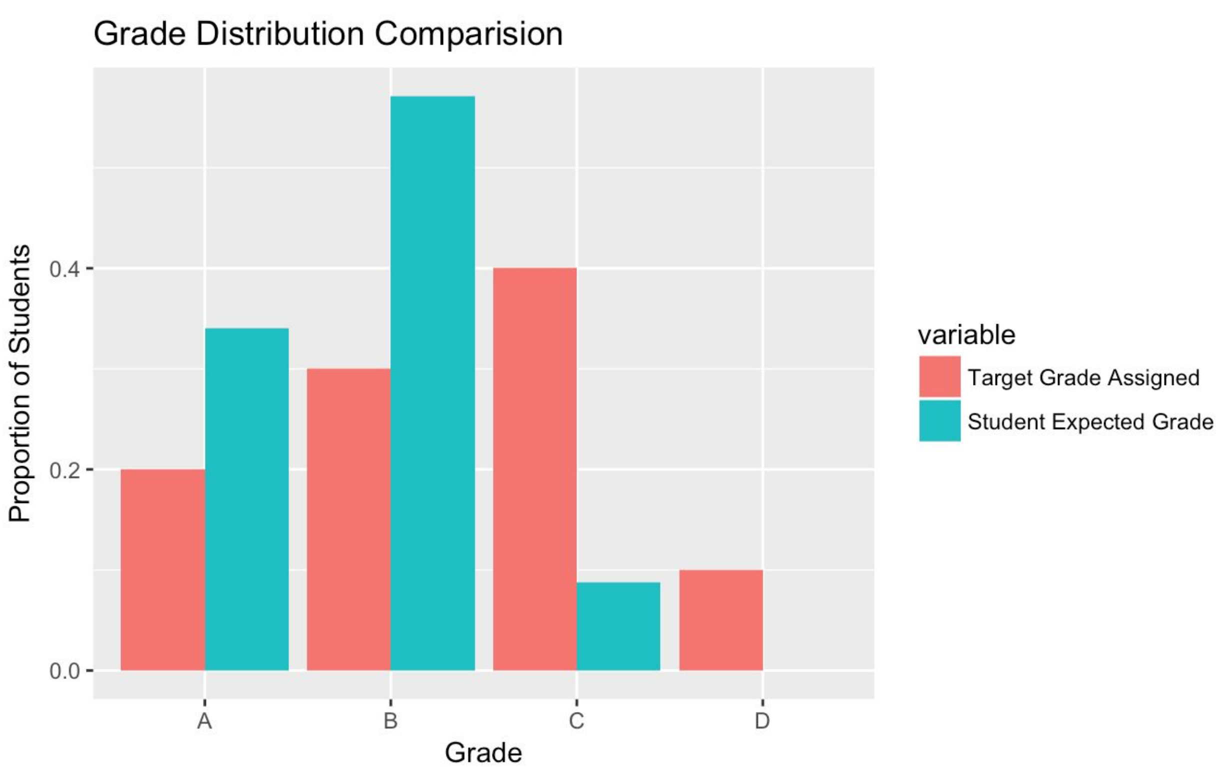
Distribution of residuals:

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|
| -0.8333 | 0.0000 | 0.1833 | 0.0000 | 0.2308 | 0.4444 |

From the tree plot, we can observe that students who worked more than 5.5 hours last week and those who are male and don't own a computer like videogames for sure and those students who own a computer and worked for more than 17 hours a week like playing video games. In all other cases too, we can see that the students are more likely to like videogames, based on the variables taken into consideration.

## Section 6

*Figure 14: Distribution comparison between students expected grade and target grade assigned.*



Lastly, we studied the grade that students expected in the course versus the target distribution used in grade assignment. The target grade distribution consists of the following: 20% A’s, 30%B’s, 40% C’s and 10%D’s or lower. The students self reported expected grade did not follow the target grade distribution (Figure 11). The distribution of grades that students expected to receive consists of the following: 34% A’s, 57%B’s, 9% C’s and 0%D’s or lower. The student’s expected grades distribution is skewed towards receiving higher grades.

We used a Pearson's Chi-squared Test for Count Data to statistically prove there was a difference in the grade distributions. This test for comparing the two distributions is valid because the survey method is a simple random sample and the student’s self reported grades are independent observations. Yates's correction for continuity was applied due to the fact the data collected was discrete, and the Chi-squared distribution is continuous. The null hypothesis for the Chi-squared Test is that there is no difference between the grades students expected and the target distribution. The alternative hypothesis is that there is a difference between grade distributions. The results of the Pearson's Chi-squared Test with simulated p-value based on 2000 replicates are: X-squared = 62.608, p-value = 0.0004998. The p-value is below an alpha of 0.05, therefore we can reject the null hypothesis. This test shows there is a statistically significant difference between the grade distributions.

Seeing that the distribution of grades that students expect does not match the target distribution used in grade assignment, we wanted to understand the potential reasons for this discrepancy. We further investigated the expected grade distribution (“grade” variable) based on various student characteristics (Figures 12-19).

*Figure 15: Expected grade distribution based on variable playing educational games.*

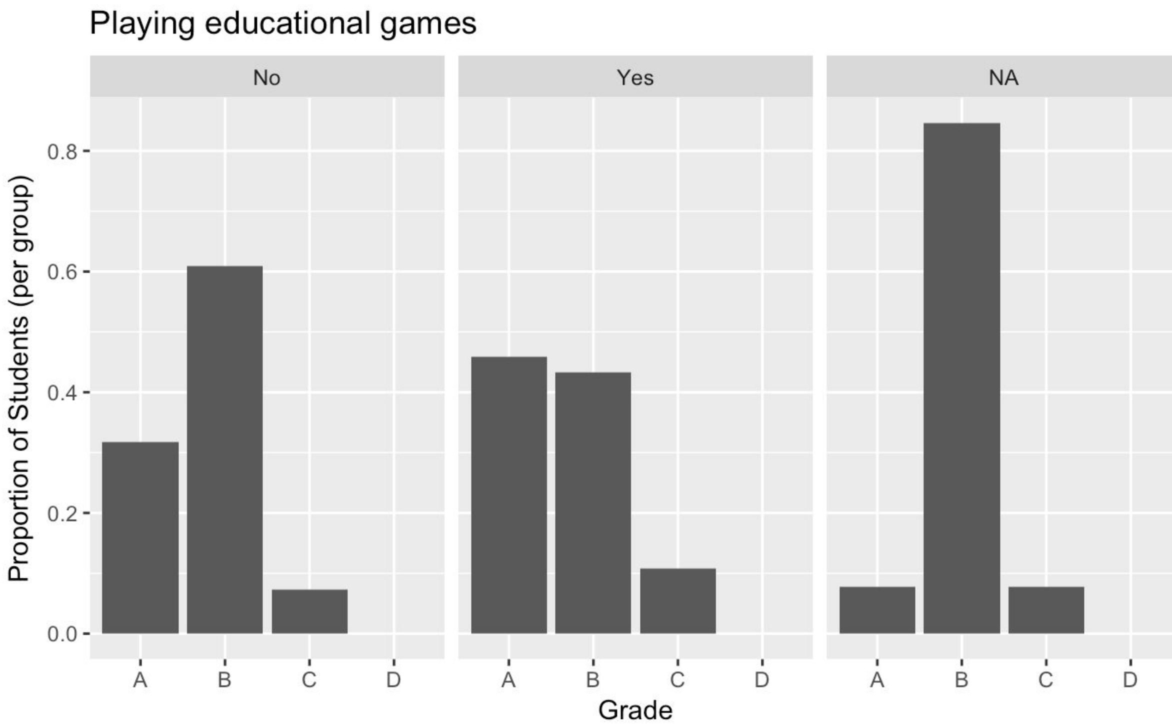



*Figure 16: Expected grade distribution based on sex.*

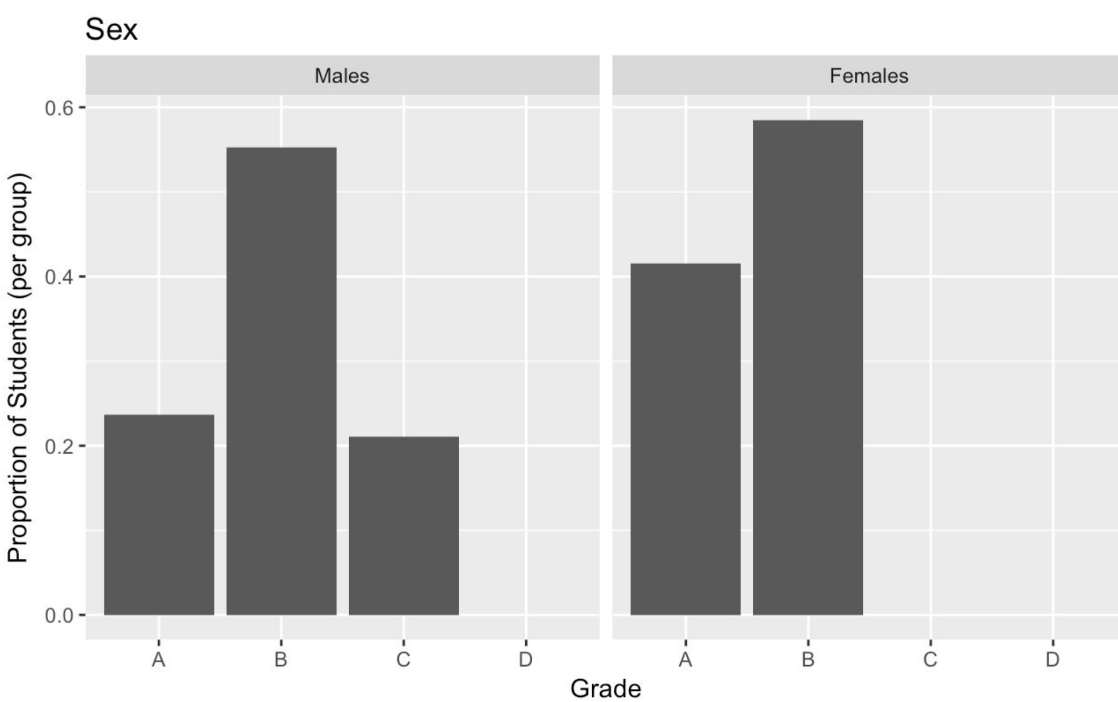


*Figure 17: Expected grade distribution based on variable number of hours worked prior to the survey.*

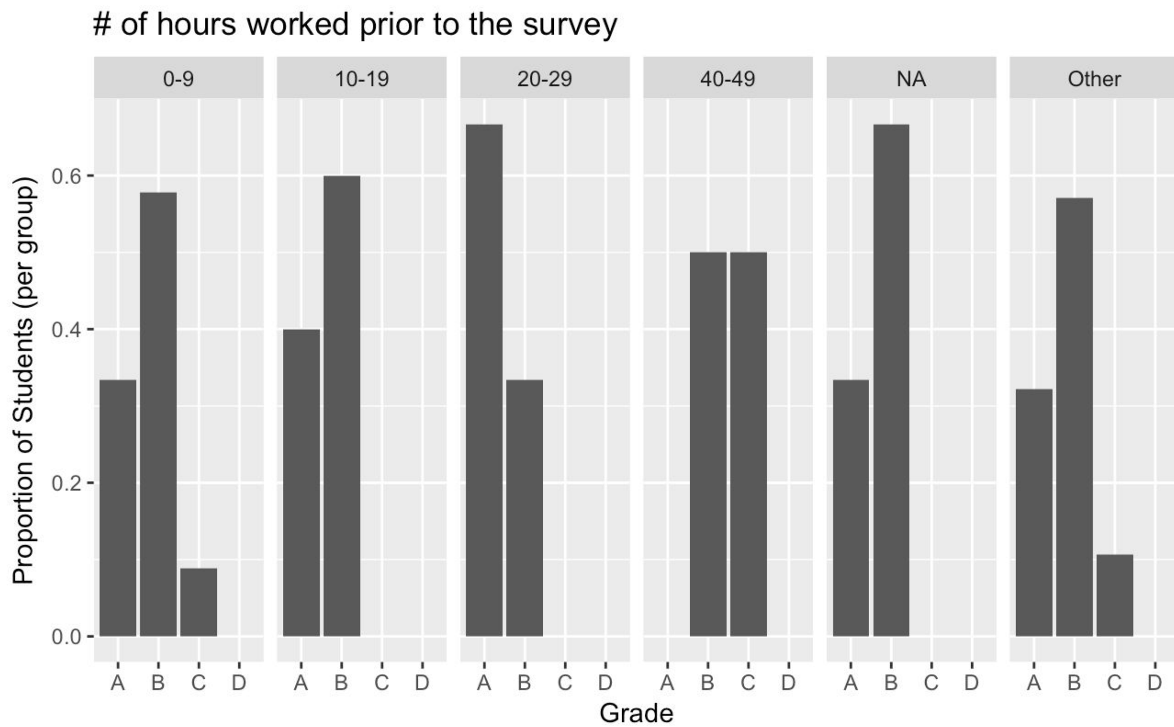



*Figure 18: Expected grade distribution based on age.*

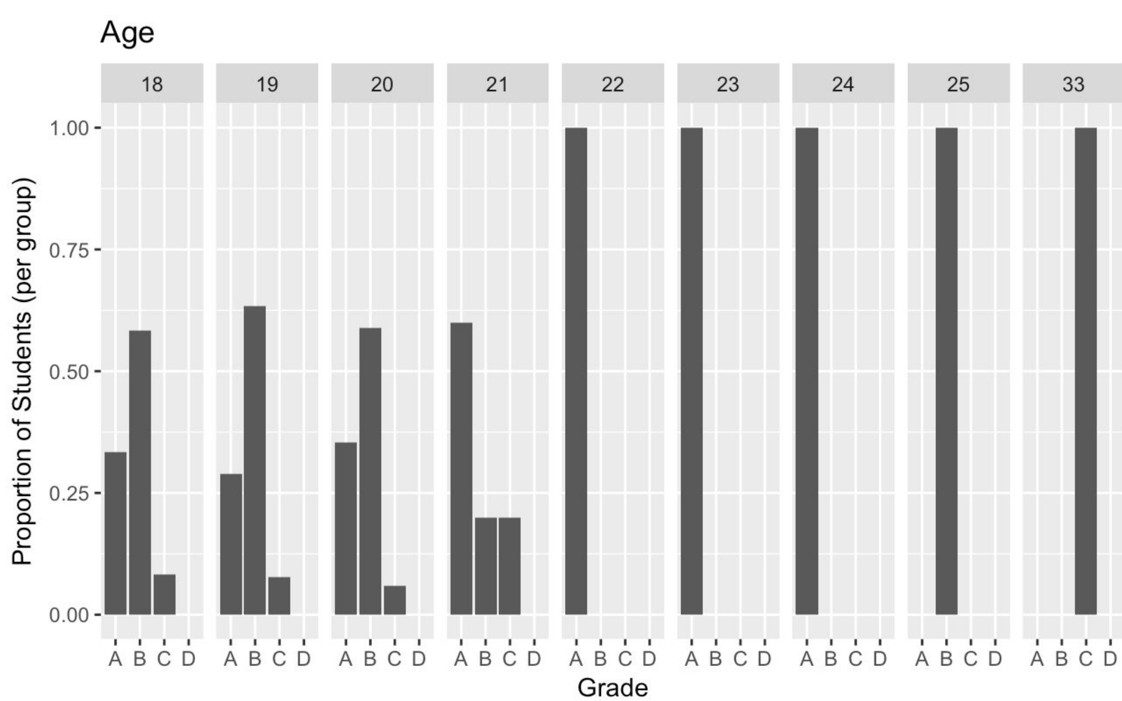


*Figure 19: Expected grade distribution based on variable owns a personal computer.*

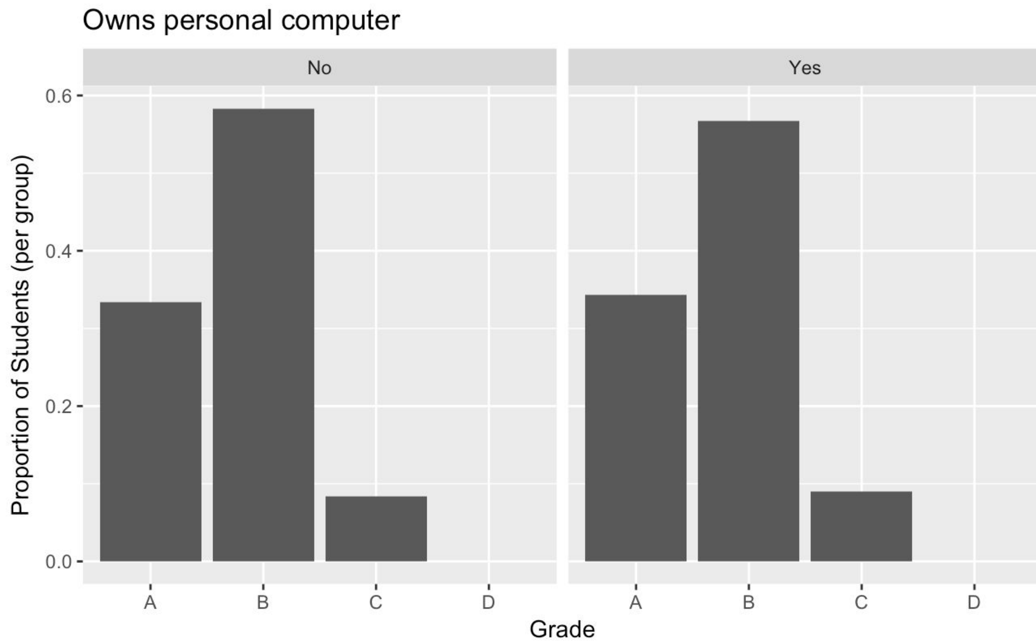



*Figure 20: Expected grade distribution based on variable computer at home.*

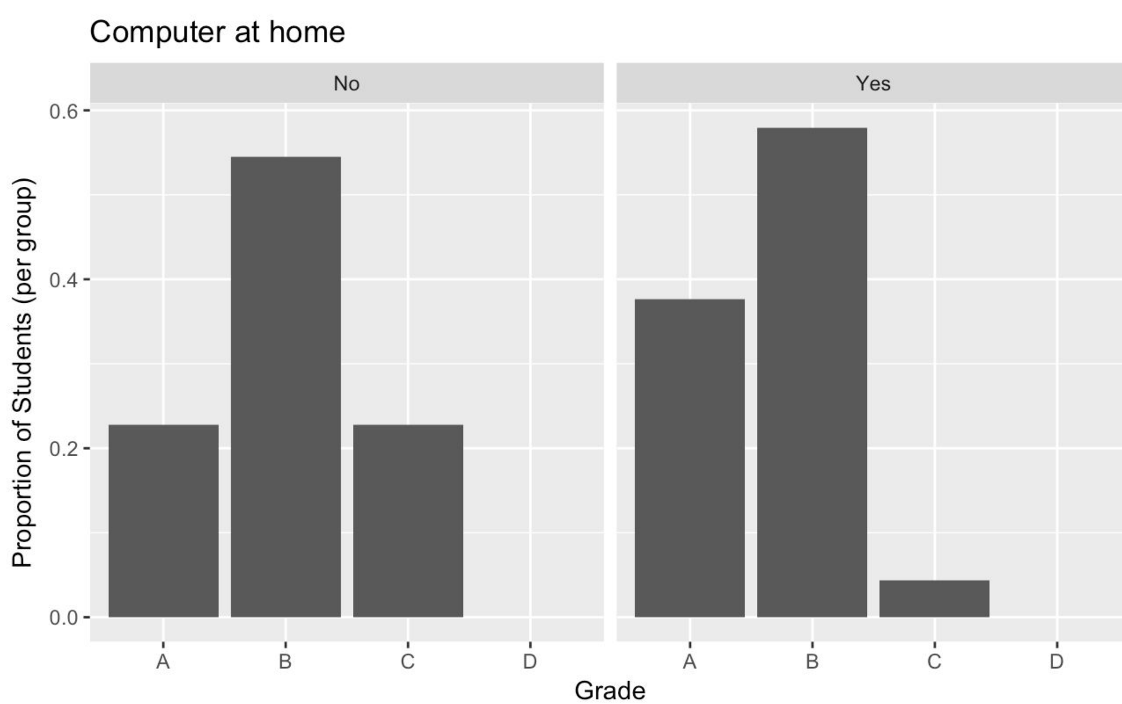


*Figure 21: Expected grade distribution based on variable hates math.*

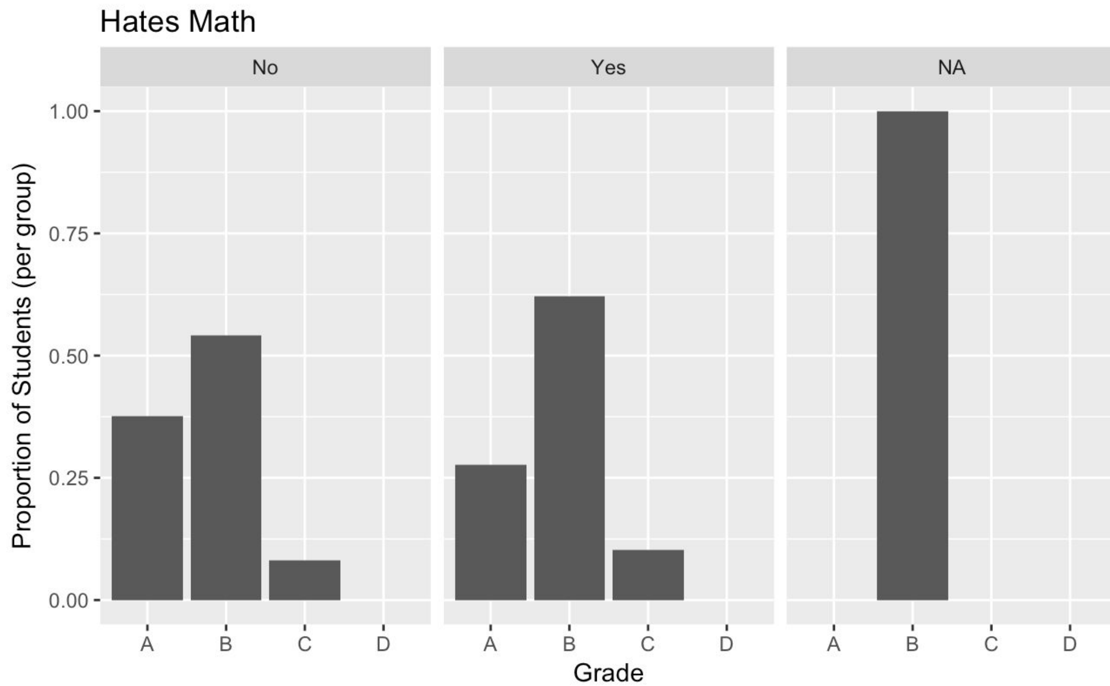
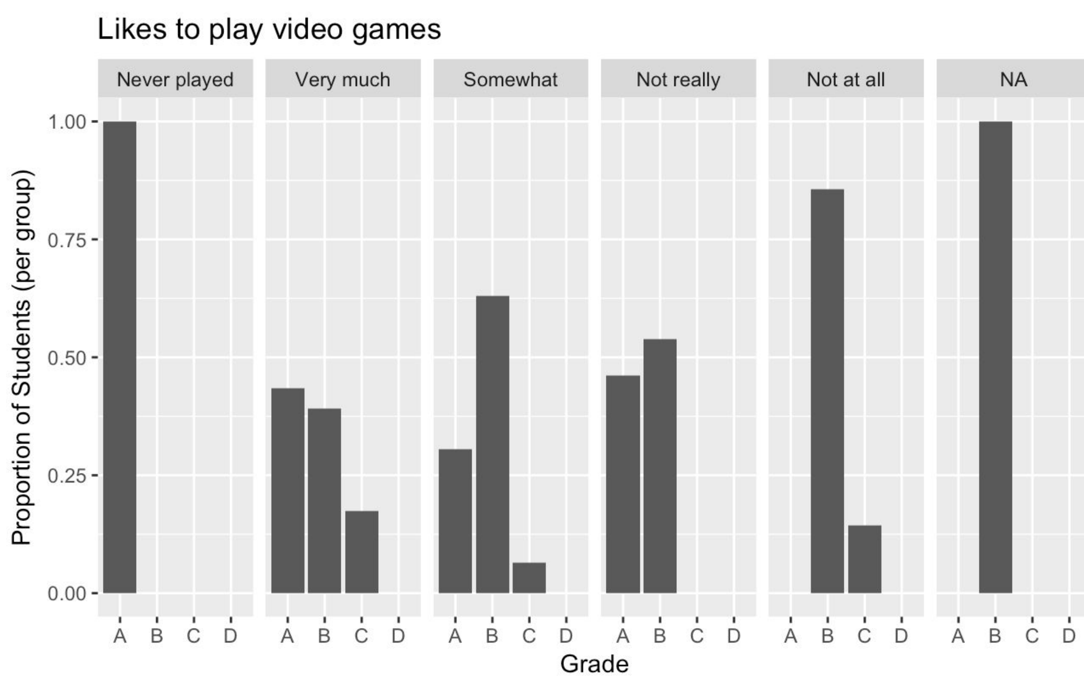
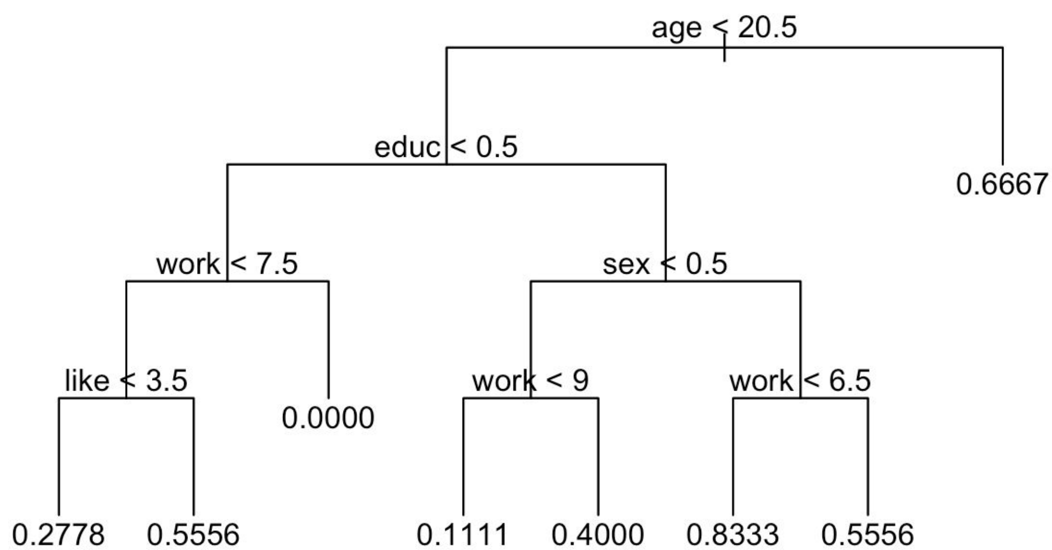



*Figure 22: Expected grade distribution based on variable likes to play video games.*



The student's expected grade distributional breakdown by various characteristics does not show noticeable differences between the groups for "math" (Figure 18), "age" (Figure 15), "like video game" (Figure 19), and most other variables. The characteristics where the most noticeable is observed is between males and females, where more females expect a higher grade than males (Figure 13). There is also a noticeable difference in proportion seen among the groups for "plays educational game"  (Figure 12) and "number of hours worked" (Figure 14).

*Figure 23: Classification tree for expected grade outcome.*




In an attempt to classify students who expected to get an "A" in the course, we created a classification tree based on "grade" as the outcome variable. The students "grade" response was re-coded to a binary classification of "A"=1, "B"-"F" = 0. Using the R "tree" package, we created the binary classification tree. We used the following predictor variables in the model: hate math (math: 1=yes, 0=no), sex (1=male, 0=female), age (age: years), playing educational video games (educ: 1=yes, 0=no), amount of hours worked prior to the survey (work: hours), like to play video games (1=never played, 2=very much, 3=somewhat, 4=not rally, 5=not at all). The tree results

(Figure 20) in 8 terminal nodes. These nodes indicated the characteristics of students likely to expect an “A” if the number is closer to 1 and students likely to expect below an “A” if the node number is closer to 0. The binary classification tree of grade outcome (Figure 20) suggests male students less than 20 years old who worked < 6.5 hours prior to the survey are more likely to expect to earn an “A”.

There could be many reasons for the increased proportion of students expected to receive A’s and B’s. One reason could be due to students who believed they were failing or had done poorly on the first exam stopped attending discussion section or dropped the course. If large number of students stopped attending class for this reason it would bias the expected grade distribution. Another reason the student’s expected grades could be skewed to the higher (A’s and B’s) end is due to optimism bias. This is a cognitive bias where individuals overestimate the likelihood of positive events occurring, and is shown to exist in all humans across demographics (Sharot, 2011). Therefore, students might be more optimistic of their grade outcome.

## CONCLUSION

The data from the survey of 91 students supports our initial hypothesis. A solid majority of students said they liked video games, in Section 5 we reported that 81.13% of male students and 68.42% of female students liked video games. However, we also found in Section 5 that a majority of students, both male and female, reported they dislike math. This data suggests that using video games to teach statistics may change students’ attitudes towards math and additionally engage them more in class and therefore be a beneficial addition to the Introduction to Probability and Statistics curriculum. Within the students that viewed video games positively, male students viewed the games much more positively than female students. Among the students that played video games, there was significant variability in how many students played and how often students played among sub-populations. In addition, we found the majority of students who disliked video games viewed them negatively because “it is pointless”, “takes too much time”, or “it is frustrating”. This points us to the importance of the ability to customize video game learning, as mentioned in the introduction. Our data suggests that implementing a single version of an educational game-based lab may not be beneficial to all students, and when designing the lab, designers must do additional research into the wants and needs of the various subpopulations of students in order to create a program that is beneficial and engaging to a variety of learners. Overall, the survey results suggest that interactive, customizable, games-based statistics lab will positively promote learning statistics concepts for all students.

## APPENDIX

*1.1: Point Estimate of Fraction of Students Who Played Video Games*

Formula: $\frac{(\text{\# of students within subpopulation that like video games})}{(\text{\# of students in population})}$

Example: $\frac{\text{\# of females who like video games}}{\text{total number of females in sample}} = \frac{9}{38} \simeq .23684$

*1.2: 95% Confidence Interval of Fraction of Students Who Played Video Games*

Formula: $y - (2 * \sqrt{\frac{y(1-y)}{\sqrt{n-1}}} * \frac{\sqrt{N-n}}{\sqrt{N}})$ , $y + (2 * \sqrt{\frac{y(1-y)}{\sqrt{n-1}}} * \frac{\sqrt{N-n}}{\sqrt{N}})$ , where N is number of population units, n is sample size, and y is the fraction of the sample that played video games.

Example: % of females that played video games $\simeq 0.23684$

$$0.23684 - (2 * \sqrt{\frac{0.23684\,(1-0.23684\,)}{\sqrt{91-1}}} * \frac{\sqrt{314-91}}{\sqrt{314}}) ,\ 0.23684 + (2 * \sqrt{\frac{0.23684\,(1-0.23684\,)}{\sqrt{91-1}}} * \frac{\sqrt{314-91}}{\sqrt{314}})$$

95% CI $\simeq$ (16.00992, 31.35850)

## REFERENCES


Bajak, Aleszu. "Lectures Aren't Just Boring, They're Ineffective, Too, Study Finds." *Science | AAAS*. N.p., 12 Jan. 2016. Web. 04 Feb. 2017. http://www.sciencemag.org/news/2014/05/lectures-arent-just-boring-theyre-ineffective-too-study-finds

Green, C. Shawn, et al. "Improved Probabilistic Inference as a General Learning Mechanism with Action Video Games." *Current Biology* vol. 20 , no. 17, 2010, pp. 1573-1579. Web. Accessed 3 Feb 2017. http://www.sciencedirect.com/science/article/pii/S0960982210009425

Griffiths, Mark, Dr. "The Educational Benefits of Videogames." *Education and Health Journal* vol. 20, no. 3, 2002, pp. 47-51. Web. 4 Feb. 2017. http://sheu.org.uk/sites/sheu.org.uk/files/imagepicker/1/eh203mg.pdf

Magon, Angela J. "Gender, the Brain and Education: Do Boys and Girls Learn

Differently?"  University of Victoria, Apr. 2009. Web. 4 Feb. 2017. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.456.6637&rep=rep1&type=pdf

Pope, Holly, and Charmaine Mangram. “Wuzzit Trouble: The Influence of a Digital Math Game on Student Number Sense.” *International Journal of Serious Games*, vol. 2, no. 4, 2015, pp. 5-21 Web. Accessed 3 Feb 2017. http://journal.seriousgamessociety.org/index.php?journal=IJSG&page=article&op=view&path%5B%5D=88

Prince, Michael. "Does Active Learning Work? A Review of the Research." *Journal of Engineering Education*. Wiley Online Library, July 2004. Web. 04 Feb. 2017.

http://onlinelibrary.wiley.com/doi/10.1002/j.2168-9830.2004.tb00809.x/abstract

Sharot, Tali. “The optimism bias”. Current Biology , Volume 21 , Issue 23 , R941 - R945. Web Accessed 6 Feb. 2017. http://dx.doi.org/10.1016/j.cub.2011.10.030

Squire, Kurt D. “Video Games and Education: Designing Learning Systems for an Interactive Age.” *Educational Technology*, March-April 2008, pp. 17-26. Web. Accessed 3 Feb 2017. http://website.education.wisc.edu/~kdsquire/manuscripts/02-squire-ed-tech-refchecV3.pdf

Taradi, M., et al. “Blending problem-based learning with Web technology positively impacts student learning outcomes in acid-base physiology.” *Journal of Advanced Physiological Education*, vol. 29, 2005, pp. 35-39. Web. Accessed 3 Feb 2017. https://www.ncbi.nlm.nih.gov/pubmed/15718381

Ref1: http://www2.stat.duke.edu/~banks/110-lectures.dir/bootstrap.pdf