WHO PLAYS VIDEO GAMES: DESIGNING A COURSE DISCUSSION

Nathan Liittschwager[1], Arielle Thibeault[1], Katherine Yamamoto[1], Alec Guthrie[1], Alex Liebscher[1], and Mike Ona[1]

***Abstract*— In order to better design educational statistics laboratories at UC Berkeley, a survey of 95 statistics students was drawn from a population of 314. Their responses to various questions regarding video game and enjoyment were recorded, so that the same characteristics that make video games enjoyable may be applied to the design of a statistical laboratory. Confidence intervals and proportion tests were conducted to estimate the true values of students who enjoy video games in various demographics, as well as to estimate the average duration and frequency of video game play. We found that approximately 36.3% of students played video games in the week leading up to the survey, 26.4% percent identified as greatly enjoying video games, and that the average duration of play per week was 1.5 hours.**

I. INTRODUCTION

The structure of courses in higher education has revolved around the traditional teacher-led lecture and discussion for decades. With the advent of more complex, engrossing video games with increasingly realistic graphics and multifaceted narratives, one may begin to question whether incorporating such technology into curricula could improve student learning. Based on a dataset of survey results from undergraduate students at a large Californian university, we investigate the video gaming habits of students and the general student life to determine an optimal video game-focused teaching approach to construct a discussion section for a statistics course. Incorporating advanced technologies into the learning process of a numerical methods course may improve the conceptual learning of students (Coller & Scott, 2009). Additionally, the possibilities that technology in higher education may bring include introducing more effective, next-generation educational models, positively reorganizing the role of educators, and preparing students to face challenging, global, interconnected situations (Fishman & Dede, 2016).

II. DATA

A. Source

Our data are from a collection of survey results from a 1994 Statistics 2 course at a large, public university in California. Of the 314 students enrolled, 95 were randomly selected to complete the survey. Complete surveys were obtained from 91 of those students. All possible permutations of the sample had an equal likelihood of being selected.

B. Variables

The survey collected information about the students' video gaming experiences, both quantitatively and qualitatively, and academic and working lifestyles. Variables collected were hours played in the week prior, the extent to which they enjoy playing, where they enjoy playing, how often they play, if they prioritize playing over other more important tasks, if they think games are educational, if there is a computer at home, if they own a PC, if their PC has a CD-ROM, if they enjoy math, hours worked in the week prior, and if they have an email address. Additionally, sex, age, and their expected course grade were self-reported.

C. Data Summary

7 of the 15 variables had responses from all 91 participants. 58% of respondents were male, and the mean age was 19.5 years. 41% reported that they played video games daily or weekly. Fortunately, it is easier to design a well-received curriculum knowing that students enjoy video games, and 76% of respondents claimed they liked to play video games "very much" or "somewhat."

III. BACKGROUND

Although many modern video games focus on first-person shooting or sports games, and are not directly educational, there is a push from educators and legislators to begin creating *serious games* (i.e. games designed for teaching and training) to supplement courses in higher education (Annetta, 2008). This push is justified by students demonstrating deeper conceptual understanding and higher motivation, and by the newly available understanding of individuals' performance for teachers after implementing video games into the curriculum (Coller & Scott, 2009; Criswell, 2009; Dominguez *et al.*, 2013).

An obstacle that surfaces when bringing technology into classrooms is the implicit preconceptions that we

[1]University of California, Mathematics, San Diego, USA

hold as a society about who is capable of succeeding and worth supporting. It is critical to design a discussion section that is enjoyable and fun for all, but more importantly, is a socioculturally-supportive learning environment that is both inclusive and sustainable (Richard, 2017).

It is far from novel to model course instruction after video games. For example, Schiller has found that instead of dismissing video games as childish and immature, one may look deeper at the foundational game design of certain video games and uncover pedagogical scaffolding and nuanced approaches for teaching new players how to play and succeed. This work was done in particular with the game *Portal*, developed by Valve Software, and finds that specific teaching principles were incorporated into the game, helping players achieve a deep understanding of conceptual topics within the context of the game and its problem-solving requirements (Schiller, 2008). Video games can provide rich environments for *active learning*, which emphasize teacher-student and student-student interaction to help students deeply learn the content they are thinking of (McCarthy & Anderson, 2000). Game design features, such as goals, challenges, narrative, and affirmation of performance, can be implemented into problem-based learning tasks, case studies, and educational games and simulations (Dickey, 2005).

A. *Contextual Importance*

The survey data is important because it demonstrates how frequently students play video games and allows us to determine what it is about video games that students like and dislike. Through statistical analysis of the data, it is possible to design more engaging statistic labs by attributing the engaging aspects of video games to the labs in order to make learning statistics and probability more enjoyable for students. Additionally, the dataset is complete enough for us to determine how to make the discussion most enjoyable to the widest range of personalities and backgrounds.

IV. Statistical Analysis

A. *Time Spent Playing Video Games*

All 91 students that were selected at random to complete the survey reported the number of hours they played the week prior. To find an estimate for the fraction of students who played versus those who did not, students who played any number of hours were classified with 1 and those who did not with 0. In total, 34 students in the sample played while 57 did not.

The mean of the sample serves as a point estimate for the fraction of students who played in the last week. In this case, the sample mean was 0.3736. In addition, a simple 95% confidence interval was calculated

TABLE I: The confidence intervals used to estimate the proportion of students that played video games the week before.

Confidence Interval Method	Sample Mean	Interval
Simple	0.3736	[0.2716, 0.4756]
Finite Sample Population Correction	0.3736	[0.2894, 0.4579]
Bootstrap	0.3630	[0.2857, 0.4505]

as [0.2716, 0.4756] which can be interpreted, by the Central Limit Theorem, as the chance that the sample mean is within two standard errors of the true mean is approximately 95%.

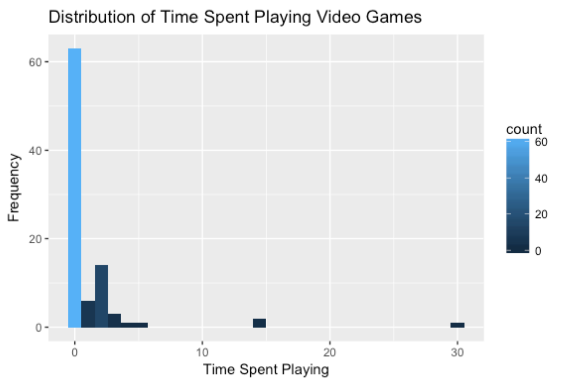
However, the finite population correction factor must be taken into account. It is often ignored since it is very small when the sample size is small compared to the population size. This is not the case with our sample size of 91 and population size of 314. Thus, the new 95% confidence interval with the correction factor is [0.2894, 0.4579].

Bootstrapping was also used to calculate another confidence interval. A bootstrap population of size 314 was created from the sample data. Since the ratio of population size to sample size is about 3.45, each observation in the sample was replicated in the bootstrap population by about the same amount. Then, 100,000 random samples of size 91 were taken from the bootstrap population. The sample means for each bootstrap sample were calculated to create a distribution of sample means. The mean of the bootstrap sample means was 0.3630. Taking the 0.025 quantile and 0.975 quantile of the bootstrap sample means creates the interval estimate, which is [0.2857, 0.4505]. A summary of these confidence intervals is given in Table I.

To further the analysis, we examined the average amount of time students reported playing video games in the week prior. Under the assumption that the data is normally distributed, we computed the 95% confidence interval for the sample mean video game playing time to yield [0.4668, 2.019]. Using the finite population correction factor, another confidence interval under the assumption of normality was constructed, [0.5878, 1.898], which was noticeably smaller in length. A third, bootstrap interval was constructed by generating a population of size 314 from the 91 sample points and then reselecting 91 data points within the generated distribution to become another simulated sample. This process was repeated 1000 times for sufficiency. The resulting simulated confidence interval [0.5333, 1.246] is the most precise of the three interval estimates about the sample mean of time spent playing video games. From this analysis, the data suggests that our assumption of normality was incorrect. A normality test was conducted to further explain the results.

B. Testing Normality in the Time Spent Playing Video Games

Fig. 1: A plot of the distribution of the time spent playing video games from the sample. Note the data is heavily concentrated about zero



- Null Hypothesis:

The time spent playing video games in the week prior from the sample distribution follows a normal distribution, for populations of the same size as the sampled data

- Results:

Since the skewness and kurtosis values lie outside 95% of the simulated data, we reject the null hypothesis and conclude the distribution is non-normal.

- Procedure:

The skewness and kurtosis were calculated from the sample distribution for the time playing video games in the week prior from the sample distribution. To test the trueness in representation of the sample data, a bootstrap simulation is generated from the sample data without replacement to compare distributions. Simulation loops were constructed (for 1000 iterations) to compute the mean, skewness, and kurtosis values of a thousand bootstrap samples of the same size as the sample data. Another Monte Carlo simulation was taken (for 1000 iterations) to compare the bootstrap data with one of a normal distribution. The 2.5% and 97.5% quantiles were taken from the normal skewness and kurtosis distributions similar to the sampled data to get the 95% confidence intervals [-0.2504, 0.2481] for the simulated skewness and [2.575, 3.503] for the simulated kurtosis. This process was repeated testing the same hypothesis with the omission of a possible outlier data point and the result was the same.

Fig. 2: A plot of the Monte Carlo simulations of kurtosis from a random normal distribution

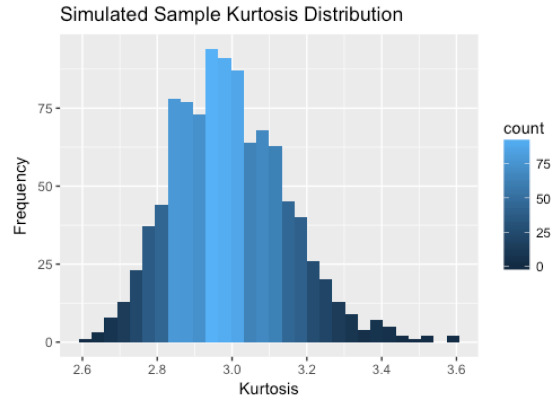
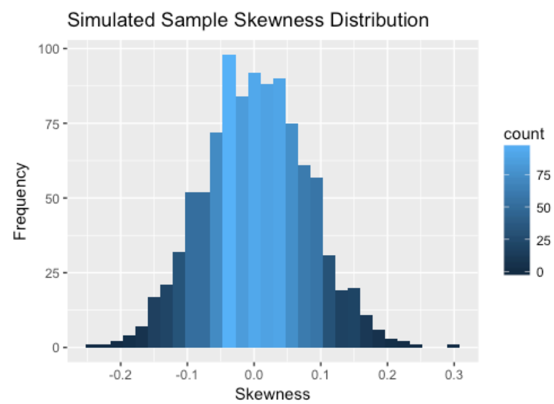


Fig. 3: A plot of the Monte Carlo simulations of skewness from a random normal distribution



C. Relationship Between Playing Time and Playing Frequency

This scenario explores how students spend their time playing video games. A comparison is made between student time playing video games and student frequency playing video games. The results may provide a computer lab designer insight about how long students would be engaged in a lab and how student attendance may be affected when students are more busy. In order to understand this relationship point estimates, cross tabulations, and grouped barplots are presented.

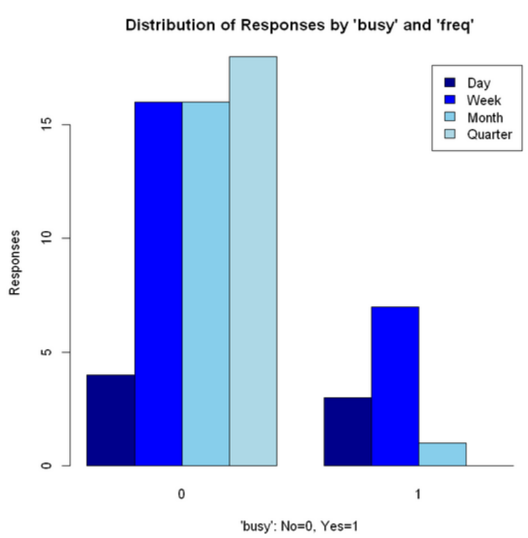
The analysis is performed in a few simple steps. First the data is cleaned. All '99' values are replaced with 'NA' and then all 'NA' responses are removed. This cleaning reduces the number of responses from 91 to 65. Next, responses are separated by the four categories of student frequency playing videogames: daily = 1, weekly = 2, monthly = 3, quarterly = 4. The cross tabulation shows the mean student time playing videogames calculated for each of these categories. Lastly, the frequency of play is separated again by whether the student will play videogames when busy. The grouped bar plot shows the student frequency playing videogames grouped by if the students will play when busy. Results are presented below.

TABLE II: The mean number of hours of video games played during the week of the second exam, also the week prior to the survey, categorized by self-reported frequency of playing.

	Daily	Weekly	Monthly	Semesterly
Mean Hours Played	3.1	1.5	0.06	≈ 0.0

When observing the cross tabulation, it is important to understand that students reported their time playing videogames during a week when they also had an exam. Observing the table, note that no students played more than 3 hours of videogames during the week of the exam. A similar survey of college student videogame use was conducted in 2006 by Sherry et.al with over 1,000 respondents from Midwestern universities. Interestingly, this study found that student time playing videogames averaged 11.5 hours per week (Sherry, 2006). The %300 difference in mean weekly videogame playing time between our data and the study by Sherry et.al is worth noting. Both the small mean in this study and the huge difference between studies might be justified by the fact that students of this study decided against playing videogames when they were busy. This possibility is explored in the figure below.

Fig. 4: A bar plot of the absolute counts by response to the question "Do you play video games when busy" Note that a majority of students do not play video games when busy.



Notice that a majority of students do not play videogames when busy and do not play videogames frequently. First, notice that roughly %83 percent of participants responded that they would not play videogames when busy, while only %17 responded that they would play videogames when busy. Second, notice that not playing when busy is represented by all frequencies of play. This data supports the prior hypothesis that most students of this study decide against playing videogames when they are busy. With regards to designing a lab, this data suggests that lab attendance may attenuate greatly during exam week regardless of how frequently students enjoy lab. Another important point estimate is that students who do not play when busy averaged 34 minutes of videogames while students who do play when busy averaged 2.5 hours of videogames. This point estimate together with the bar plot suggest that most students will not play videogames more than 34 minutes a week when they are busy. This finding might suggest a lab designer keep lab times as short as possible under university requirements, for example 1 hour. Ultimately, the data from the crosstable then supported by the bar plot can be used to infer how long a lab should last and how student attendance of the newly designed labs might change during busy weeks.

D. *Student Attitude Toward Video Games*

In general, it is difficult to claim that a population truly enjoys video games. We make the assumption that if one enjoys video games, they consequently play more hours. However, in the sample, 62% of students did not play any video games in the week prior and only 5.5% played, on average, more than 30 minutes a day.

A follow up survey was conducted to determine what types of games students play, why they play video games, and what they dislike about video games. Students were asked to select all types of games they played and up to three reasons for why they liked or disliked video games. When asked what type of video games they play, 63% of students reported that they enjoy strategy-based games and 50% selected action. It is important to note that not all students responded to this question as those who said they do not play video games or never have where instructed to skip it. These students were also excluded from reporting reasons for playing the game. Amongst the students who answered the initial question, the survey indicated that the primary reason for playing video games was relaxation with 66% of students selecting this response. On the other hand, the data from the survey showed that 48% of the students involved in the survey disliked video games because they are too time consuming while 40% of students reported that video games were too costly.

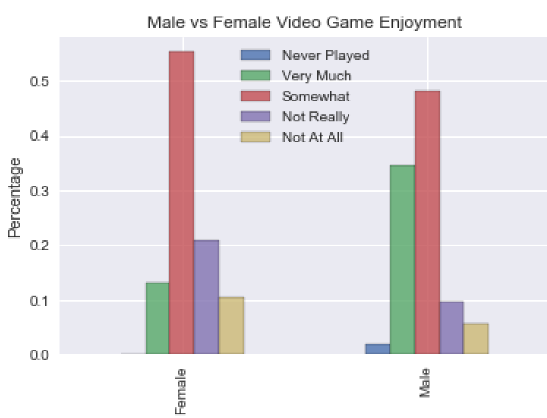
While the follow up survey uncovered some information about the aspects of video games that students like and dislike, the data is inconclusive due to the methodology of the survey which lacks one-to-one correspondence as students were able to choose multiple answers to each question.

E. Difference Between Video Gamers and Non-Video Gamers

Procedure: In order to further understand what demographics play and enjoy video games within the sample data, cross-tabulations between various groups were made. The data was partitioned in several ways. First by sex, then by those students who worked in the week of the survey and those who did not. When applicable, a chi-square test of independence was conducted to determine if the enjoyment of video games was independent from the demographic being analyzed. It should be noted that only one individual in the sample, a male, responded that they had not played a video game before, and in one other case, there was a non-respondent on whether they enjoyed video games or not. The non-respondent was discarded from the sample in the proceeding analysis. Several chi-square tests were performed on the tabulated data.

- Male vs Female Enjoyment of video games

Fig. 5: A bar plot of the proportion of responses by men and women in the survey to the question "Do you enjoy video games?" Note that they are relatively similar, however the green bar in the male group is quite a bit higher. In fact, more than 3 times as many males as females responded "Very much" to the question.



The data was partitioned by reported sex and then a cross tabulation of the data was given. For each row and column, the data is given in a count, then a normalized (against rows) percentage. Figure 1 indicates that a higher proportion of men enjoy video games "very much" when compared to women. This suggests that, within the sample, men and women have similar feelings towards video games - with "video game fanatics" being composed of largely men.

TABLE III: The absolute counts of males and females and their corresponding answer to the question "Do you enjoy video games"? The sample consisted of 38 women and 53 men - with one male eliminated from the following table for being a non-responder.

	Never Played	Very Much	Somewhat	Not Really	Not at All
Female	0	21	8	4	10.5%
Male	1	18	25	5	3

- Additional Analysis on Difference in Sex and Video Games

While Figure 5 and Table III give a visual description of the differences in video game enjoyment derived between men and women, in order to understand whether these differences are really significant, a statistical test for the difference in proportions was conducted. In particular, a bootstrapped test for a difference in proportions.

- Procedure:

In accordance to the phrasing of the survey questions, participants of the survey were classified into a "like" and "dislike" group, but using an indicator function on the survey responses. A "like" was given the value of 1 and included only those that responded with "Very Much" to the question of "Do you enjoy video games?". All other responses, since they varied between lukewarm enjoyment to outright contempt, were given a value of 0, to indicate "dislike". The proportion of students who like video games in the survey was then calculated to be $\hat{p_T} = 0.256$. Moreover, the proportion of male and female students were calculated to be $\hat{p_M} = 0.346$ and $\hat{p_F} = 0.132$,respectively. Thus, the an observed difference was found to be $\hat{p_M} - \hat{p_F} = 0.214$. In order to account for the dependence between data points in the survey, a pseudo-population was created, and 1,000,000 bootstrap samples were drawn from this pseudo population (shifted to simulate the null hypothesis) and passed through the test statistic $t^*(\mathbf{x}) = \hat{p_1} - \hat{p_2}$ and an approximate p-value was calculated. For an in depth discussion on the techniques used here, please see the Methods and Theory section. The proceeding tabulates the results of the test.

- Null Hypothesis ($\alpha = 0.05$):

There is no difference in the proportion of Men and Women who enjoy video games "Very Much".

- Results:

Upon computing the achieved significance level with $B = 1,000,000$ bootstrap samples passed through the test statistic, the approximate p-value was computed to be $p = 0.00006$. Such a small p-value indicates very strong evidence against the null hypothesis, and so the null hypothesis is rejected and the observed differences cannot be attributed to chance alone.

1) Comparison of Video Game Enjoyment of Working and Nonworking students: Students in the sample were partitioned into those that worked and did not work and their response to the question "Do you enjoy video games?" was tabulated into Table 2 and Figure 2. The absolute counts may be seen in Figure 2 with the normalized values given in Table 2.

Fig. 6: A comparison of students who worked in the week leading up to the survey and those who did not. Somewhat surprisingly, working students seem to enjoy video games more than non-working students. Their counts were higher in both the "very much" and "somewhat" categories, and lower in the others, when compared to working students.

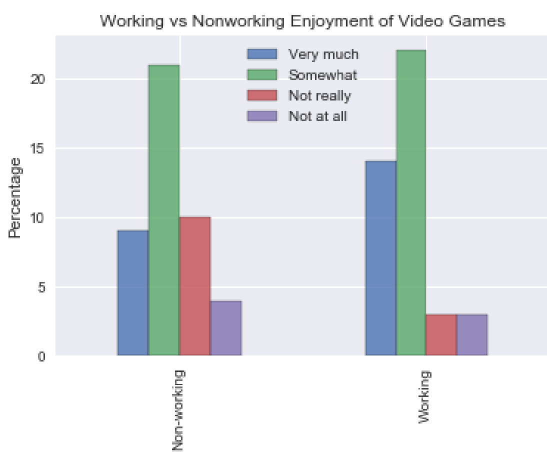


TABLE IV: The absolute counts of the same data given in figure g. It is easy to see that working students appear to enjoy video games more, on average, than non-working students. This is indicated by the relatively few negative responses, and the higher proportion of responses in the "very much" category.

	Very Much	Somewhat	Not Really	Not at all
Non -working	9	21	10	4
Working	14	22	3	3

- Procedure:

Perhaps surprisingly, Figure 6 and Table IV indicate that the students who most enjoy video games are those that worked during the week leading up to the survey. In order to determine whether these visualizations actually indicate significance, a proportion test was conducted to compare the proportions of working and non-working student body whom enjoy video games "very much". (A slight digression: 4 non-respondents had to be eliminated from the survey in order to conduct the proceeding test. This accounts for the slight differences in obtained proportions.) The total proportion of students who enjoy video games was calculated to be $\hat{p_T} = 0.264$. The proportion of working students who like video games very much was computed to be $\hat{p_W} = 0.326$, and the nonworking proportion was found to be $\hat{p_N} = 0.205$, thus giving an observed difference of $\hat{p_W} - \hat{p_N} = 0.121$. A population-bootstrap procedure was fashioned in a similar manner to the difference in proportions between sexes.

- Null Hypothesis $(\alpha = 0.05)$

There is no statistical difference between the proportion of working students who like video games very much, and the proportion of non-working students who like video games very much.

- Results:

Calculating $B = 1,000,000$ bootstrap samples under the null hypothesis (which involved a shifting of the dataset), the achieved significance level was computed and the approximate p-value was computed to be $p \approx 0.023$. The p-value provides reasonable evidence against the null hypothesis. We conclude that the null hypothesis should be rejected, and that there is a statistically significant difference in proportion of students who like video games, between working and non-working students.

2) Comparison of Video Game Enjoyment between PC-Owners and Non-Owners: The advent of the reasonably priced modern personal computer begs the question: does having easier access to video games via a personal computer (PC) affect people's feelings towards video games? In order to further analyze this question, the survey tabulated those students who owned a computer and those who did not, which was subsequently analyzed with the following descriptive statistics.

Figure 3 indicates an interesting dynamic among PC owners and non-PC owners. PC owners appear to fall to the more extreme ends of the spectrum of responses to video games. A somewhat higher proportion of students tended towards the "very much" and "not really" responses to the survey question "Do you enjoy video games?", and significantly less fell on the lukewarm "somewhat" response when compared to those who did not own a PC at home. In fact, the vast majority of responses from non-PC owners fell under "somewhat". While it is outside the scope of this paper to explain exactly why there is ambivalence in non-PC owners, the proportions to each response may be analyzed.

Fig. 7: This figure shows the relative proportions of students' responses to the survey question: "Do you like video games", separated by those who own a pc and those who do not. The relatively higher count of "very much" responses and lowered "somewhat" responses by PC owners seem to indicate that PC owners enjoy videogames more. However, note that in the PC-owning group, there is also a relatively higher proportion of contemptuous responses.

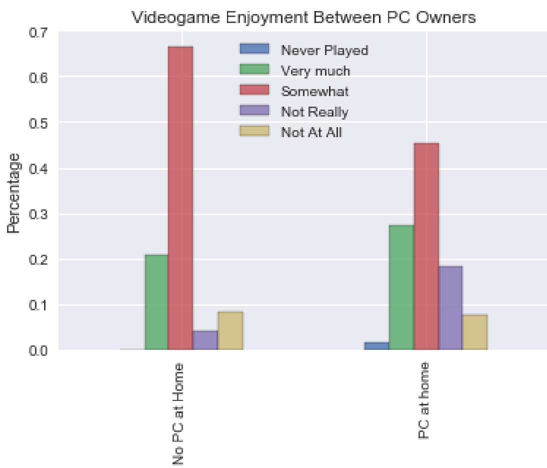


TABLE V: The absolute counts of survey responses to the question "Do you enjoy video games?" Note that despite the fact that PC owners had twice as many responses in the "somewhat" response as compared to non-owners, this indicated a smaller proportion of the population. The group of pc-owners was larger in size than non-owners. Since it is the proportion that is of interest, our analysis focuses there.

PC status	Never Played	Very Much	Somewhat	Not Really	Not at all
Non-Owner	0	5	16	1	2
Owner	1	18	30	12	5

As can be seen in Table V and in Figure 7, the response variables are difficult to analyze intuitively. Once again, to test if there is a statistical difference in the proportion of video gamers between PC owners and non-owners, the population bootstrapping algorithm may be applied.

- Procedure:

The procedure is analogous to that used in the bootstrapping procedure used for comparing video gamers between sexes and video gamers between working and non-working students (see Methods and Theory). Those who responded with "Very Much" were indicated with a 1, and the rest were denoted with a 0. The proportions were calculated. The total proportion of students who enjoyed video games was found to be $\hat{p_T}0.264$. The proportion of video gamers among PC owners was computed to be $\hat{p_O} = 0.281$ and the proportion of video gamers among non-owners was found: $\hat{p_N} = 0.217$. Thus calculating the difference yielded the observation $\hat{p_T} - \hat{p_O} = 0.064$.

- Null Hypothesis ($\alpha = 0.05$)

There is no statistically significant difference in the proportion of students who enjoy video games "very much" between PC owners and non-owners.

- Results:

After fashioning a pseudo-population, $B = 1,000,000$ bootstrap replicates were simulated from the pseudo-population under the null distribution. The achieved significance level was computed and a p-value of approximately $p = 0.2932$ was found. Note that $p > \alpha$ and so there is not sufficient evidence to reject the null hypothesis. Hence, the null hypothesis has failed to be rejected and it's possible the observed differences were due entirely to chance.

3) Deciding Variables in Students who Like Video Games: To further assess the importance of each variables in the overall profile of a student who enjoys video games, we implement a Decision Tree Classifier (DTC). A DTC is a nonparametric supervised learning algorithm which learns simple decision rules from a numerical feature set. Branches of the DTC are if-else types, meaning they separate the data from the previous rule (or from the base dataset, in the case of the first branch) into two child sets. This branching logic continues for a finite number of iterations until the original data has been fully separated.

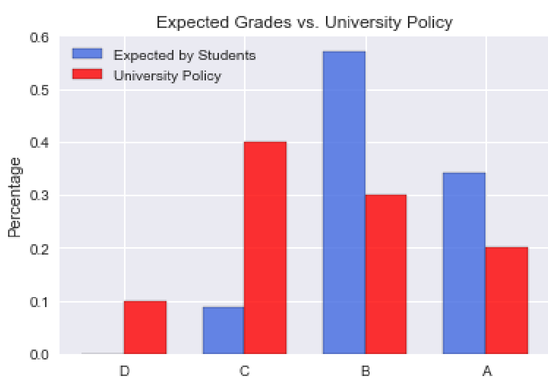
For our DTC, we clean the data to eliminate as many null values as possible, as modeling requires complete data. The *where* variable is broken into three binary-type columns, each denoting whether that student stated the arcade, home computer, or home system as their preferred playing space. For students who claim they do not like to play video games at all, *freq* was set to 5 (less often than semesterly), *busy* was set to 0 (assuming they would not play if they were busy), and *educ* was set to 0 (assuming they would believe video games are not educational). The *where*, *cdrom*, and *email* variables were excluded, since they are not believed to heavily influence whether the student likes video games. After these imputations, rows still containing null values were removed.

We then trained our model over a training split in the remaining dataset ($n = 59$). Part of the dataset was held aside for testing ($n = 20$). Our model achieved a 0.90 mean training accuracy score. The target variable was whether the student reported they liked video games (*like* as ”Very Much” or ”Somewhat”), denoted as 1, otherwise 0.

F. A Further Investigation of Expected Grade Distribution

Berkeley’s University Policy demands that students in their classes be graded on what is commonly called a ”Curve”. That is, grades are assigned to the students in order from best to worst – the top 20% of the class receives an ”A”, the next 30% recieves a ”B”, the next 40% receives a ”C” and the bottom 10% recieves a failing grade, a ”D” or ”F”. In effect, this standardizes the grading procedure between departments, and also gives the students an indication of where they stand relative to the rest of the class. The determination of effectiveness of such a grading policy is well beyond the scope of this paper, but what follows is a cursory analysis of the grades students expected in their classes, in comparison to University Policy. In the survey, students were asked to record what grade they expected to receive at the end of the course.

Fig. 8: This figure contrasts the grade distribution expected by the sample population of students (blue) relative to the actual university policy (red). Note that not a single student expected to fail the course, while university policy necessarily mandates that at least 10% of students will. It is possible that students largely overestimate their abilities.



Again, the proportion of students and their grade expectations was analyzed. Several proportion tests were conducted, first among all students who expected a grade of ”A” or better, then among students who played video games daily, then those who had responded that they liked video games ”very much”. The reason for the division in analyses is to try and answer two questions. First, do students, as a group, over estimate their abilities? Second, do gamers in particular overestimate their abilities? Based on figure 4, it is clear that students certainly do not underestimate themselves, as a whole, when compared to University Policy.

- Null Hypothesis ($\alpha = 0.05$)

There is no statistical difference in the proportion of students who expected an ”A” grade, and university policy.

- Procedure:

Among all students, the total proportion of expected ”A” grades was found to be $\hat{p_A} = 0.341$. Since only 20% of A’s may be distributed, the observed difference in proportion of students who expected an ”A” and university policy was $\hat{p_A} - 0.20 = 0.141$. The population bootstrapping algorithm was applied, and $B = 1,000,000$ samples were drawn from the shifted (to the null hypothesis) pseudo-population. The proportion difference was computed, and the achieved significance level was computed to give an approximate p-value.

- Results:

The approximate p-value was found to be $p = 0.00331$. Since $p < \alpha$, there is sufficient evidence against the null hypothesis. Hence, the null hypothesis is rejected, and there is a statistically significant difference between the proportion of students who expected an ”A” when compared to university policy.

The proportion test indicates that a statistically large number of students from the survey expected to do better in the course than they actually did. Since university policy on grades is binding, it can be assured that only 20% of students actually did receive a grade of an ”A”. Why is there a difference expected grades and actual grades? It stands to reason that students are aware of the university policy, and if accurate judges of their own ability, would estimate properly where they stand relative to the class in terms of ability. One possible explanation is the Dunning-Kruger effect. A now infamous study by Dunning and Kruger concluded that high-achievers tended to underestimate their abilities, while lower achieving students tended to over-estimate. It is hypothesized that students who are relatively untrained in a task simply do not know how difficult the task is - and since this survey was conducted while the class was ongoing, and before an exam, it’s likely that students were simply unaware of the difficulty of the class. (Dunning & Kruger, 1999)

What follows are a pair of proportion tests on those who enjoy video games very much and those who are daily players.

- Procedure:

Due to the overlap of those who play video games every day and those who identify as lovers of video games, it is sufficient to perform one analysis. In particular, on those who identified as individuals who enjoyed video games "very much". The data was partitioned into just those who identified in the affirmative of "Do you enjoy video games very much" and the proportion of those who expected a grade of an "A" was calculated – $\hat{p_A} = 0.826$, thus giving an observed difference of $\hat{p_A} - 0.20 = 0.626$ between expected distribution of "A"s among gamers and the university policy. Note that this proportion is quite a bit larger than the proportion of expected "A" grades when considering the entire class.

It should be noticed that the gamers represent a relatively small part of the survey population, that is, they only made up 25.3% of the survey population. While this closely aligns with university policy of 20% "A"s, our intent is to better understand gamers as a group and their estimation of their own abilities. The pseudo-population bootstrapping techniques that have been in use throughout the analyses is applied here.

- Null Hypothesis ($\alpha = 0.05$)

There is no statistically significant difference between the proportion of expected "A" grades of gamers and the university's policy (20% of the class to receive A's).

- Results:

After fashioning $B = 1,000,000$ bootstrap replicates of the test statistic $t(\mathbf{x}) = \hat{p_1} - \hat{p_2}$, the achieved significance level was computed and a p-value estimate obtained, $p < 0.0001$. Such an incredibly small p-value indicates overwhelming evidence against the null hypothesis. The null hypothesis is subsequently rejected, and the observed differences may not be attributed to chance alone.

It is beyond the scope of this paper to analyze why gamers as a group appear to be so confident in their abilities. However, many gamers responded to the survey follow up question "Why do you play video games?" with "to achieve a feeling of mastery". It is possible that self-identified gamers enjoy the feeling of success and winning, and tend to place themselves into situations where they feel that can be achieved. Such considerations should be taken into account upon designing an educational lab.

V. DISCUSSION

The population from which the data was collected consisted of students at the University of California, Berkeley who were enrolled in Statistics 2, a lower-division statistics course, during Fall 1994. Although the data was collected from a random sampling of the target population– the 1500-2000 students who enroll in introductory statistics courses at UC Berkeley every year– there are still variables involved that could have a confounding effect when studying the extent to which students enjoy video games.

For example, this limited dataset does not provide information on socioeconomic standing of the participants. UC Berkeley, a public university that offers financial aid and scholarships, can be assumed to have a diverse student body representing all social classes, even during the time the survey was conducted in 1994. That being said, certain students might "very much" enjoy playing video games but lack the means to play as frequently as they would like (indeed, in the follow up survey, 40% of participants reported disliking video games because they are too costly). Socioeconomic standing can also influence the number of hours a student works for pay and consequently, the number of hours they played video games the week prior to the survey.

Although there exists information on whether the students like or dislike math, the dataset also fails to report on the respective majors and class loads of the students. The target population, while consisting of students enrolled in introductory statistics courses, is not necessarily made up of all statistics majors because the students took these classes for the purpose of satisfying UC Berkeley's quantitative reasoning requirement. Thus, it can be assumed that they are studying a variety of subjects, some of which require more time consuming core classes which would have a confounding effect on the frequency of play as well.

While we do control for some confounding variables, like sex, age, and whether there is a computer at home, selection bias stems from lack of information. Out of 314 students in Section 1 of Statistics 2, 95 were randomly selected to participate in the survey and out of those 95 participants, 4 submitted incomplete surveys which were then excluded from the data set. Additionally, 1.1% of participants failed to provide information as to what extent they like video games and 3.3% did not provide the number of hours worked the week prior to the survey. It is important to note that students who reported that they had never played video games before or did not like video games at all were instructed to skip certain questions so variables like where they play, whether games are educational or not, and frequency of play had many nonrespondants.

After investigating the results of the Decision Tree Classifier (Appendix, figure 10), one of the most defining features of the dataset is the reported frequency of play. Whether or not the student has a computer at home is also deterministic. If students choose to leave home to play video games at an arcade, this is a decent indicator that they really do enjoy video games, thus its importance in the tree. Whether or not the participant considered games to be educational is also divisive, having a Gini impurity of 0.499, which is a measure of how a random sample would be misclassified.

For future analyses, there are improvements that can be made to the survey in order to design a better, more effective statistics lab for undergraduate students. One such change would be to increase the size of the random sample of students from which the data was collected and to include some of the confounding variables listed earlier to reduce uncertainty (shrink confidence intervals) and the effect that missing responses have on the outcome. It would be possible to conduct a more thorough analysis of who likes video games and why or why not by controlling for noise that could possibly affect students' ability to play and accessibility to video games. This analysis could also be further developed by studying the relationship, if any, between variables provided by the dataset such as hours worked and time played or frequency of play and whether the student considers games to be educational.

VI. CONCLUSIONS

The statistical analysis indicated that approximately 37% of students played video games during the week prior to the survey and 31% reported "very much" liking video games while 20% reported only "somewhat" liking video games. When comparing the responses of female and male participants, analysis showed that men and women have fairly similar attitudes towards video games, the only difference being that men were significantly more likely to be "video game fanatics" or someone who enjoys video games "very much". Analysis of the attitudes towards video games of employed versus unemployed students demonstrated that the students who worked during the week before the survey were the participants who enjoyed video games the most. Even so, the majority of both groups, 68.2% of non-working students and 85.7% of working students, reported liking video games to some extent. The analysis also revealed a relationship between playing time and frequency of play. Participants who played daily, played on average 3.1 hours during the week prior to the survey while those who played weekly recorded an average of 1.5 hours. The hours logged by participants who only played video games monthly or semesterly were insignificant.

In terms of deeply learning and retaining material, there is a relationship between positive results and the amount of time students spend thinking and interacting around the material (McKeachle, *et al.*, 1986). In the follow up survey, 48% of students responded that they disliked playing video games because it is too time consuming. Also, the average time spent playing video games during the week prior to the survey was approximately 53 minutes, and students, on average, played video games biweekly. Thus, the committee should design a 50 minute lab that takes place every other week. This allows students to get in enough time engaging with the material in a fun, interactive way (which is known to solidify learning more than by traditional textbook or lecture) without making it seem too time consuming. The follow up survey also revealed that students enjoyed playing strategy and action games the most so the statistics labs should mirror specific attributes of those types of games, like hand-eye coordination and the ability to develop one's own strategies based on a set of simple rules. These video game inspired labs should be especially utilized when the class consists of mostly male or working students as they are more likely to enjoy and benefit from such a lab. On the other hand, if the class is largely female or unemployed, the labs may not be as effective because these demographics tend to be less interested in video games.

Overall, we highly recommend that based on these conclusions, the UC Berkeley Statistics committee should consider designing labs that incorporate specific aspects of video games to improve undergraduate students' educational experience.

METHODS AND THEORY

What follows is a collection of results and accompanying methodology used throughout this report. The list is too extensive to be given in full detail here. The most important results are stated, and citations are given to where one might find the relevant proof or full description of methodology.

Dependence of Sample Data. When sampling from a finite population, the sampled data points are going to be dependent upon each other due to conditional probability. In a finite population, the probability of choosing one data point increases given that other data points have already been selected. The process behaves similarly for our data: the probability for one student to be chosen to participate in the study is dependent upon the others have already been chosen. The mathematical property is observed:

$$P(A|B) \neq P(A) \tag{1}$$

where A and B are independent events.

Probability Distribution of Units Chosen in Sample. Let $I(j), j = 1, ..., n$ be the jth unit drawn from from the set $\{1, ..., N\}$ where N is an integer. Then,

$$P(I(1) = 1) = \frac{1}{N}$$

$$P(I(1) \text{ and } I(2) = N) = \frac{1}{N(N-1)}$$

and it is easy to see that if

$$1 \leq j_1 \neq j_2 \neq ... \neq j_n \leq N$$

then

$$P(I(1) = j_1, ..., I(n) = j_n) \\ = \frac{1}{N(N-1)\dots(N-n+1)}$$

which is the probability structure of the simple random sample of size $n < N$, where n and N are both integers.

Population Correction Factors L. et $x_{I(j)}$ denote the jth data point sampled from the population distribution F. Suppose that distribution F has a true mean of μ. Then the expectation of $x_{I(j)}$ is

$$\begin{aligned} E(x_{I(j)}) &= \sum_i^N x_i P(I(j) = i) \\ &= \sum_i^N E(x_i)\frac{1}{N} \\ &= \mu. \end{aligned}$$

Likewise, the expected variance is given by

$$\begin{aligned} Var(x_{I(j)} &= E[(x_{I(j)} - \mu)^2] \\ &= \frac{1}{N}\sum_i^N (x_i - \mu)^2 \\ &= \sigma^2 \end{aligned}$$

where σ^2 is the true population variance. This implies that variance of the sample average $\bar{x}$ is as follows

$$\begin{aligned} var(\bar{x}) &= \frac{1}{n^2} Var\left(\sum_j^n x_{I(j)}\right) \\ &= \frac{1}{n^2}\sum_j^n Var(x_{I(j)}) + \frac{1}{n^2}\sum_{j\neq k}^n Cov(x_{I(k)}, x_{I(j)}) \\ &= \frac{1}{n}\sigma^2 + \frac{n-1}{n} Cov(x_{I(j)}, x_{I(k)}) \end{aligned}$$

where the last equality follows from the fact that all pairs $(x_{I(j)}, x_{I(k)})$ are identically distributed. Note that the covariance is non-zero, because the sampling procedure necessarily makes the pairs $(x_{I(j)}, x_{I(k)})$ dependent. And since $Cov(x_{I(j)}, x_{I(k)}) = -\frac{\sigma^2}{N-1}$ it follows easily that

$$Var(\bar{x}) = \frac{1}{n}\sigma^2\frac{N-n}{N-1}$$

and

$$SD(\bar{x}) = \frac{1}{\sqrt{n}}\sigma\frac{\sqrt{N-n}}{N-1}.$$

The ratio $\frac{n}{N}$ is called the ”sampling fraction”. If the sampling fraction is sufficiently large, say $\frac{n}{N} > 0.10$, then the estimates $Var(\bar{x})$ and $SD(\bar{x})$ inherently biased. In order to correct the bias, we apply what we call the ”finite population correction factor”, given below:

$$\frac{N-n}{N-1} = 1 - \frac{n-1}{N-1}\ 1 - \frac{n}{N}$$

In order to correct for the bias, note that an unbiased estimator for σ^2 is given by

$$s^2\frac{N-1}{N}$$

which then implies an unbiased estimator for $Var(\bar{x})$ is

$$\frac{s^2}{n}\frac{N-n}{N}.$$

For a more rigorous discussion of population correction factors as it relates to bootstrap and confidence intervals, see (Hinkley & Davison, 2006).

Central Limit Theorem. Let $X_1, X_2, ...$ be a sequence of independent random variables each having the same distribution, $f_X(x)$. Suppose that the mean μ and the variance σ^2 of $f_X(x)$ are both finite. If $\sum_{i=1}^n X_i = S_n$, then for any numbers a and b,

$$\lim_{n\to\infty} P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx \tag{2}$$

Normal Approximation Confidence Intervals. Suppose that $X_1, ..., X_n$ is a sequence of n independent random variables. Let $\bar{X} = \frac{1}{n}\sum_i^n X_i$ denote the sample mean. When $n \geq 30$, and lack of heavy skew is present in the sampling distribution of $X_1, ..., X_n$, then under the central limit theorem, the following holds:
If μ_0 is the true mean, σ the standard deviation, and z^*_α so that $P(Z \geq z^*_\alpha) = \alpha$, where $Z \sim N(0,1)$, then the $100\%(\alpha)$ confidence interval is a range of values with the form

$$\left[\bar{X} - z^*_{(\alpha/2)}\cdot\frac{\sigma}{\sqrt{n}}, \bar{X} + z^*_{(\alpha/2)}\cdot\frac{\sigma}{\sqrt{n}}\right]$$

where if $z^*_{(\alpha/2)} = 1.96$ when $(\alpha/2) = .975)$, then we have the 95% confidence interval.

Error Corrected Confidence Intervals. Confidence intervals suffice for the interval in which the true mean is located about the sample statistic, under the assumptions of the Central Limit Theorem. These criteria are as follows: the data points are independent identically distributed, the mean and the variance of the data are finite values, and the population size, n, is sufficiently large so that the population size is considered infinite in comparison to sample size. For sampling without replacement in finite populations, the regular assumptions of the confidence intervals under the CLT fail. As a rule of thumb, when the sample size greater than 5% taken from a finite population, the standard errors about the estimators no longer hold as the standard error represents the observed error of the sample. To correct for this over estimation, the finite population correction factor is applied to the standard error to decrease the value to a proportion that represents the error when drawing a sample large enough to the total population size.

The adjusted confidence interval more closely represents the actual interval for the sample mean drawn without replacement from a finite population:

$$\left[\bar{X} - z^*_{(\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}} \cdot \frac{\sqrt{N-n}}{\sqrt{N}}, \bar{X} + z^*_{(\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}} \cdot \frac{\sqrt{N-n}}{\sqrt{N}}\right]$$

Bootstrap Confidence Intervals. Let $\hat{\theta}(\mathbf{x})$ be an estimator function for an unknown parameter θ. By applying the bootstrapping algorithm described below, draw B samples of size n, where n is the size of the original data, and pass the B samples through the estimator $\hat{\theta}(\mathbf{x})$ in order to produce B bootstrap replicates of the estimator. That is, produce $\theta^*_b, b = 1, ..., B$ bootstrap replicates via passing the simulates through $\hat{\theta}$. Then, the approximate $100\%(1-\alpha)$ confidence interval for θ is given by

$$\left[\theta^*_{\alpha/2}, \theta^*_{1-\alpha/2}\right]$$

where $\theta^*_{\alpha/2}$ is the $\alpha/2$ percentile of the B bootstrap replicates θ^*_b. This is equivalent to taking the $\alpha/2$ quantiles of the B bootstrap replicates if they were ordered from least to greatest. This procedure is fully described in (Efron, 2000).

Population-Bootstrap Algorithm. Used throughout this paper was an algorithm referred to as a "Population Bootstrap" in order to fashion non-parametric confidence intervals for various point estimates, as well as to conduct a simple hypothesis test using a the test statistic $T(\mathbf{x}) = \hat{p_1} - \hat{p_2}$ to test a difference in proportions. What follows is a brief description of the algorithm and the procedure used to correct for the bias inherent in a finite-population survey sample. For a truly rigorous discussion the techniques used, we refer the interested reader to "Bootstrap Methods and their Application" by A. C. Davison and D. V. Hinkley.

Let $\mathbf{x} = (x_1, ..., x_n)$ be a random vector obtained from the distribution function F. Suppose further that the population from which $\mathbf{x}$ is finite and that each $x_i, 1 \leq i \leq n$ was drawn without replacement so that no two data points are the same, and that each drawing is necessarily dependent on the preceding drawings before it. In such cases, the normal bootstrap algorithm (to draw with replacement) does not adequately mirror that of a simple random sample without replacement from a finite population. Confidence intervals constructed via the normal bootstrap algorithm are systematically too wide. Moreover, hypothesis tests used with the normal algorithm will give p-values that do not converge to the true p-value as $B \to \infty$, as intended from a bootstrap. In order to account for the dependencies between data points and simulate the simple random sample (without replacement), a population bootstrap must be used.

Suppose that $\mathbf{x}$ consists of n data points from the population F of N size. Also suppose that $n < N$ to a large enough degree that the sampling proportion $\frac{n}{N} \geq \epsilon$, where $\epsilon > 0$ is large enough to disrupt the normal approximation by the Central Limit Theorem. No hard and fast rule for how large ϵ must be before this occurs, and so the practitioner must use their best judgment. As a rule of thumb, $\epsilon \geq 0.10$ is sufficiently large to cause such a disruption.

In this case, a pseudo-population is constructed using the random sample $(x_1, ..., x_n)$. The process is done in the following way: $(x_1, ..., x_n)$ is concatenated with itself k times. Then a simple random drawing of size l (without replacement) is taken from the original sample and concatenated with the kn data points, where k and l are integers chosen so that $kn + l = N$, the size of the original population. We now have a pseudo-population of size N, from which B bootstrap replicates of size n are drawn **without** replacement, in order to form confidence intervals and hypothesis tests. For a more rigorous discussion on why this method works, please see Hinkley's book, "Bootstrap Methods and Its Application".

In order to conduct a hypothesis test, one first notes their observed test statistic $s(\mathbf{x})_{obs}$, then manipulates $\hat{F}$ so that it is the empirical distribution *under the null hypothesis* H_0. Then B replicates of the bootstrapped test statistic $s(\mathbf{x}^*)$ are computed and an approximate p-value is calculated by

$$p \approx ASL = \frac{1 + \sum_{i=1}^{B} I(s_i(\mathbf{x}^*) \geq s(\mathbf{x})_{obs})}{(B+1)} \quad (3)$$

Where I is the indicator function, and ASL denotes the Achieved Significance Level.

That is, we sum the number of simulated observations as extreme or more extreme than our actual observation, $s(\mathbf{x})_{obs}$, and the $+1$ is added to the numerator and denominator to include the initial observation and avoid numerical errors. While this p-value is approximate, it may be made arbitrarily accurate to the true p-value by taking B arbitrarily large. In particular, throughout this paper, the test statistic used was the proportion, $\hat{p} = \frac{1}{n}\sum_i^n x_i$, where $x_i, i = 1, ..., n$ is a series of Bernoulli trials. (Efron, 2000)

Shifting to simulate the Null

Let $\mathbf{x} = (x_1, ..., x_n)$ be the random vector from distribution F, and $\mathbf{y} = (y_1, ..., y_m)$ be the random vector from empirical distribution G. Take $\bar{z}$ to the pooled arithmetic mean of $(x_i, y_j), i = 1, ..., n, j = 1, ..., m$. Also, put $\bar{x}$ and $\bar{y}$ be the arithmetic mean of $\mathbf{x}$ and $\mathbf{y}$ respectively. The test statistic used to check a difference in means between two groups is the difference $s(\mathbf{x}, \mathbf{y}) = \bar{x} - \bar{y}$. Then H_0 is simulated by taking the empirical distributions $\hat{F}$ and $\hat{G}$ to put equal probability on the shifted data points,

$$\tilde{x_i} = x_i - \bar{x} + \bar{z}, \quad i = 1, ..., n$$
$$\tilde{y_j} = y_j - \bar{y} + \bar{z}, \quad j = 1, ..., m$$

Samples are drawn with replacement from $\hat{F}$ and $\hat{G}$ to form the B bootstrap replicates $(\mathbf{x}^{*\mathbf{b}}, \mathbf{y}^{*\mathbf{b}})$, $b = 1, ..., B$, which are then passed through our test statistic $s(\cdot)$. An approximate p-value is calculated using the ASL defined above.(Efron, 2000)

VII. Acknowledgments

We would like to thank Professor Jelena Bradic for her suggestions. A.L. contributed to the introduction, literature review, the decision tree model, and the interpretation. All authors drafted the manuscript. All authors approved the final version of the manuscript for submission.

References

[1] Annetta, L. (2008). Video Games in Education: Why They Should Be Used and How They Are Being Used. *Theory Into Practice, 47*(3), 229-239.

[2] Coller, B. D., & Scott, M. J. (2009). Effectiveness of using a video game to teach a course in mechanical engineering. *Computers & Education, 53*(3), 900-912.

[3] Criswell, C. (2009). Can video games be educational? *Teaching Music, 16*(6), 24-24,26. Retrieved from https://search.proquest.com/docview/227183662?accountid=14524

[4] Dickey, M. D. (2005). Engaging by design: How engagement strategies in popular computer and video games can inform instructional design. *Educational Technology Research and Development, 53*(2), 67-83.

[5] Dominguez, A., Saenz-De-Navarrete, J., De-Marcos, L., Fernandez-Sanz, L., Pages, C., & Martinez-Herraiz, J. J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education, 63*, 380-392.

[6] Fishman, B., & Dede, C. (2016). Teaching and Technology: New Tools for New Times. In Gitomer D. & Bell C. (Eds.), *Handbook of Research on Teaching* (pp. 1269-1334). Washington, DC: American Educational Research Association.

[7] McCarthy, J. P., & Anderson, L. (2000). Active learning techniques versus traditional teaching styles: Two experiments from history and political science. *Innovative higher education, 24*(4), 279-294.

[8] McKeachle, W. J., Pintrich, P. R., Lin, Yi-Guang, & Smith, D. A. F. (1986). Teaching and Learning in the College Classroom: A Review of the Research Literature. *National Center for Research to Improve Postseconotoxy Teaching and Learning*.

[9] Richard, G. (2017). Video Games, Gender, Diversity, and Learning as Cultural Practice: Implications for Equitable Learning and Computing Participation Through Games. *Educational Technology, 57*(2), 36-43.

[10] Schiller, N. (2008). A portal to student learning: what instruction librarians can learn from video game design. *Reference Services Review, 36*(4), 351-365.

[11] Kruger, J, and D Dunning. Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. Journal of Personality and Social Psychology., U.S. National Library of Medicine, Dec. 1999, www.ncbi.nlm.nih.gov/pubmed/10626367.

[12] Davison, A. C., and D. V. Hinkley. Bootstrap Methods and Their Application. Cambridge University Press, 2006.

[13] Efron, B. An Introduction to the Bootstrap. Boca Raton, Florida, 2000.

[14] Sherry, John & Greenberg, Bradley & Lucas, Kristen & Lachlan, Kenneth. Video game uses and gratifications as predictors of use and game preference. International Journal of Sports Marketing and Sponsorship, 8ed, 213-224, 2006.

VIII. Appendix

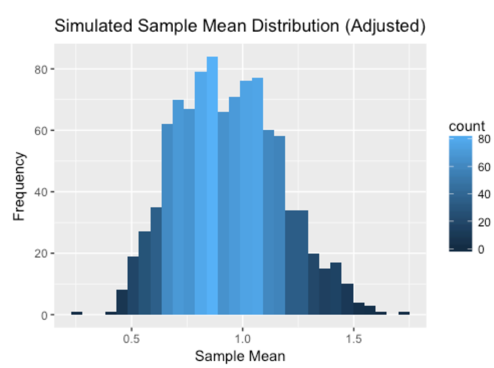


Fig. 9

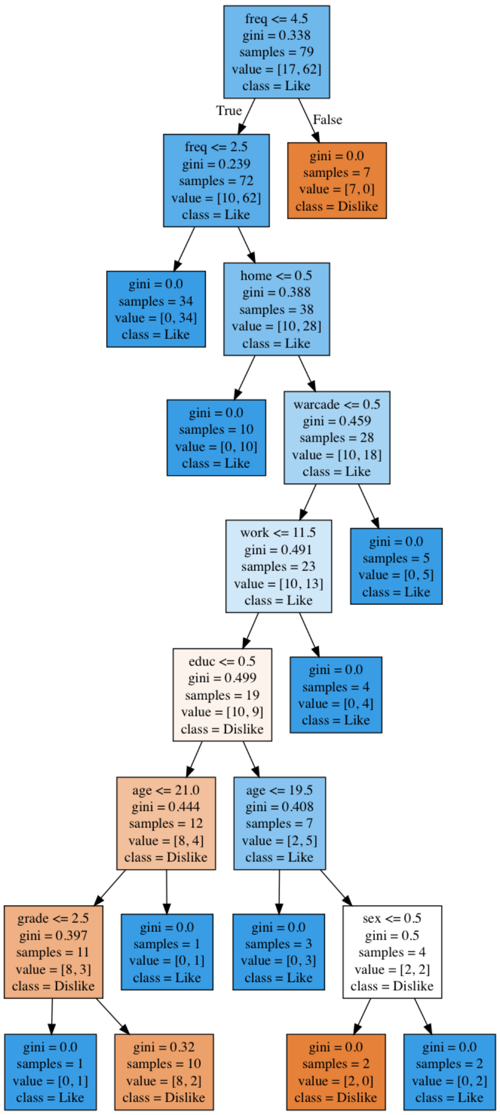


Fig. 10: The results of the Decision Tree Classifier, where branches visualize pathways to traits of students who like video games and those who do not.