MATH189/289 Homework 2
Brendan Gee 4th Year Math Econ BS
Robert Gougelet 5th Year Cognitive Science PhD
Sidney Browne 4th Statistics BS

Introduction

A lower division introductory statistics class is a requirement for many college students, many of whom are enrolled in the class to fill some general education requirement. At UC Berkeley the statistics department has designed a computer lab component to supplement the traditional curriculum by teaching statistics concepts through an interactive environment. The concept of a computer lab has been compared to an educational video game. This comparison prompted the designers of the computer labs to conduct a survey to understand who plays video games, how much they play, and what they like and dislike about video games in order to determine what aspects of students experiences with video games might improve the effectiveness and enjoyableness of the computer labs. The data set analyzed here was gathered via simple random sampling from a single statistics class of 314 students in fall '94. There were 95 students selected randomly to complete the survey and 91 successfully submitted surveys. The survey consisted of questions about whether or not students like video games, how often they play, what they like about video games, and general demographic background. An important aspect of this survey is that students who indicated that they do not like video games were asked to skip the questions pertaining to why they like video games. Our investigations explore how the data may provide suggestions for designing a more effective and enjoyable statistics computer lab.

[Scenario 1:] Begin by providing an estimate for the fraction of students who played a videogame in the week prior to the survey. Provide an interval estimate as well as a point estimate for this proportion.

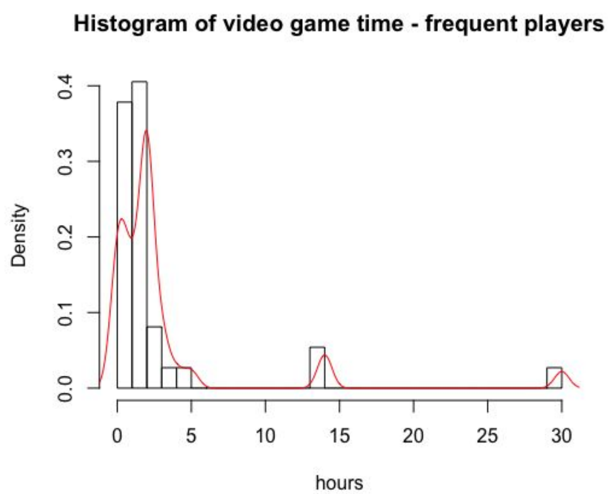
In this scenario, we want to treat time as a Bernoulli random variable, where students who played in the week prior to the survey is equal to one, and otherwise is equal to zero. Looking at the sampled data, we can see that 57/91 of the people did not play video games (time=0) the week prior to the survey. Therefore, since everyone in the sample responded to this question, 34/91 did play video games (time>0) in the week prior. The point estimate, p, is precisely the proportion when the Bernoulli random variable is equal to one. Hence, p=34/91. Next, we proceeded to construct a confidence interval for this point estimate. We did this by calculating the estimator for the standard error,(SE=.04297) using the fact that the point estimate was unbiased, and set up a 95% confidence interval to achieve that p lies in the interval [.2894,.4579]. This means that with 95% confidence, the fraction of students who played video games a week prior to the survey lies between 28.94% and 45.79%. From this estimate, it seems like the majority of students don't play video games on a weekly basis, but this conclusion could be misguided. In scenario 2, we will determine if the exam during this week had an impact on students willingness to play, and if the majority of students actually do play video games on a weekly basis.

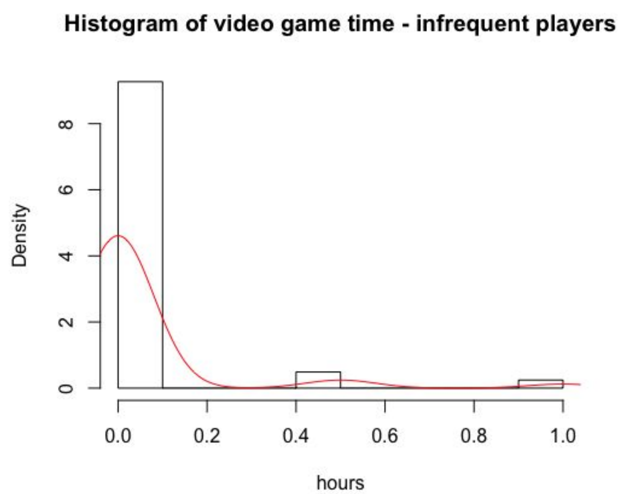
*** [Scenario 2:] Check to see how the amount of time spent playing video games in the week prior to the survey compares to the reported frequency of play (daily, weekly, etc).**

MATH189/289 Homework 2
Brendan Gee 4th Year Math Econ BS
Robert Gougelet 5th Year Cognitive Science PhD
Sidney Browne 4th Statistics BS

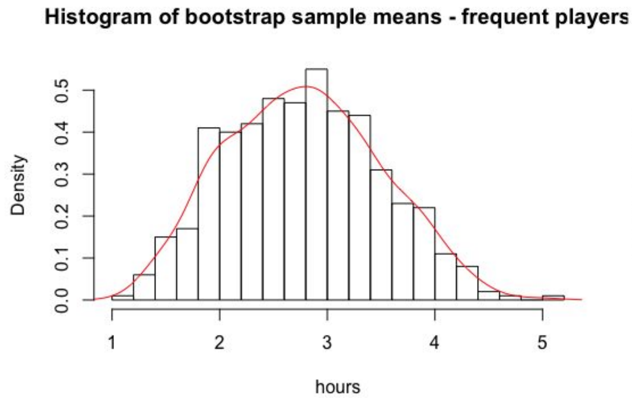
How might the fact that there was an exam in the week prior to the survey affect your previous estimates and this comparison?

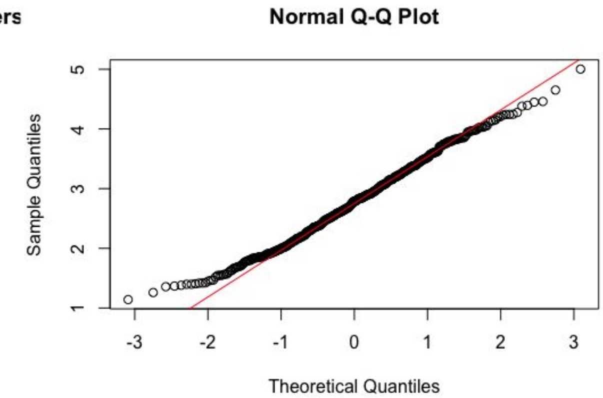
Density histograms, with density curves overlaid in red, were constructed for the amount of time in hours students spent playing video games in the week prior to the survey for each of the four self reported frequencies of play: daily, weekly, monthly, and semesterly. From these four histograms similarities and differences between the self reported frequencies of play are evident. Daily and weekly players have similar histograms in that most of the weight is distributed between 0 and 5 hours. Similarly, monthly and semesterly have similar histograms in that most of the weight is at zero hours, and all of the weight is between 0 and 1 hour. This prompts the choice to divide players into 2 groups; frequent players (daily/weekly), and infrequent players (monthly/semesterly). Density histograms were then constructed for the amount of time frequent or infrequent players spent playing video games in the week prior to the survey. Both histograms are skewed right, and have a lot of their weight distributed near zero which indicates that most students in either frequency group played few or no hours of video games in the week prior to the survey. Additionally, these histograms show that there is a much wider range of times spent playing video games for frequent than infrequent players.

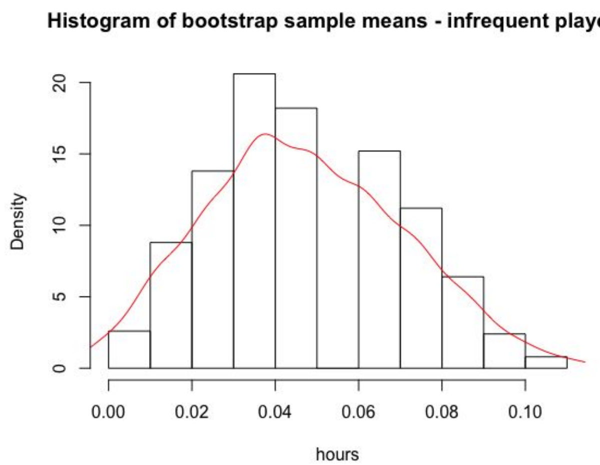


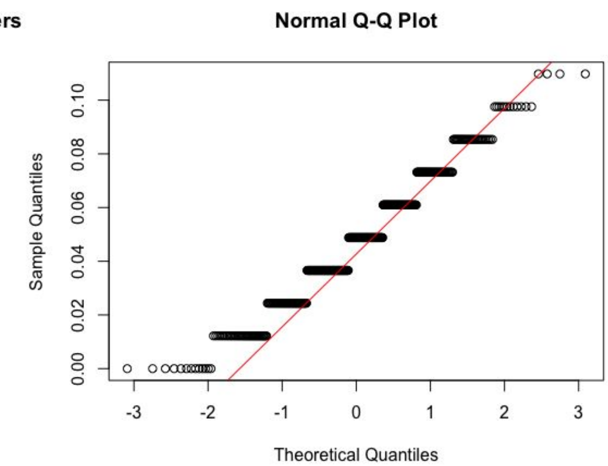


A bootstrap population of hours played was constructed for both frequent and infrequent players with population size determined by the proportion of frequent or infrequent players in the sample data times the total population size of 314 students. Then 500 samples of size equal to the number of frequent or infrequent players in the sample data were collected without replacement, to replicate how the data was originally gathered, and their means were calculated. Those sample means form the bootstrap distribution of sample means for the number of hours spent playing video games in the week prior to the survey for frequent or infrequent players. These distributions can be visualized with their density histograms and normal QQ plots. The bootstrap distribution of sample means for both frequent and infrequent players are both close to a normal distribution. However, there is some deviation in the tails of both distributions. This suggests we use the bootstrap distributions of sample means for frequent or infrequent payers, as opposed to a normal approximation, to construct confidence intervals for the true average number of hours students spent playing video games in the week prior to the survey.







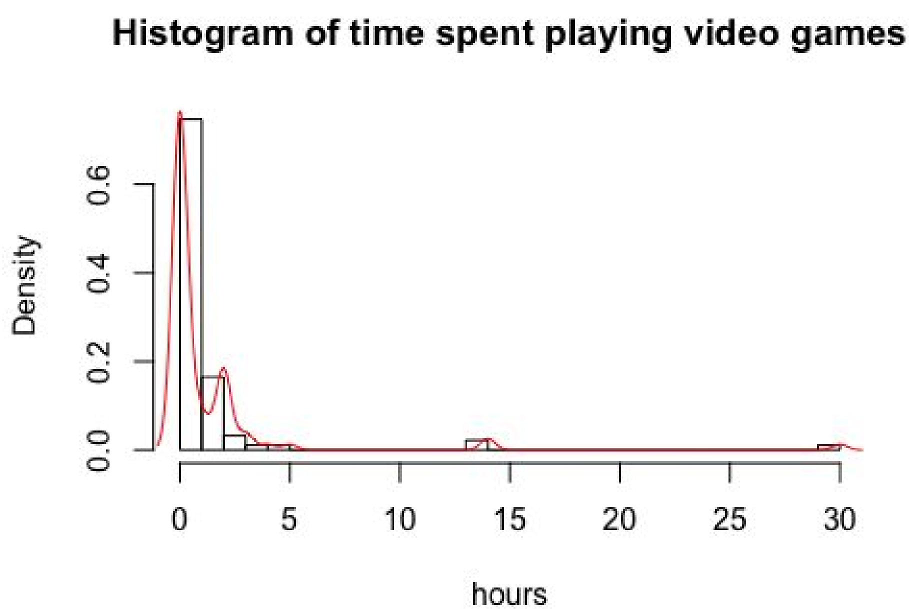


A 95% confidence interval for the true average number of hours frequent or infrequent players played video games in the week prior to the survey was constructed using: the sample average number of hours spent playing video games, the standard deviation, and quantiles from the bootstrap distribution of sample means, and the finite population correction factor where the population size is the proportion of frequent or infrequent players in the sample data times the total population size, and the sample size is the number of frequent or infrequent players in the sample data. A 95% confidence interval for the true average number of hours frequent players spent playing video games in the week prior to the survey, based on the point estimate, the sample mean, 3.00 hours is (2.12, 5.52) hours. This means that the true average amount of time frequent players spent playing video games in the week prior to the survey is almost certainly between 2.12 hours and 5.52 hours. A 95% confidence interval for the true average number of hours infrequent payers spent playing video games in the week prior to the survey, based on the point estimate, the sample mean, .0488 hours is (.0487, .0507) hours. This means that the true average amount of time infrequent players spent playing video games in the week prior to the survey is almost certainly between .0487 hours and .0507 hours, or 2.922 minutes and 3.042 minutes. These confidence intervals show that students who considered themselves daily or weekly players, which we grouped together as frequent players, and students who considered themselves monthly or semesterly players, which we grouped together as infrequent players, behaved differently. The behavioral difference observed in these confidence intervals is that, on average, frequent players spent a lot more time playing video games than infrequent players, and that there is a much broader distribution of time spent playing video games for frequent as opposed to infrequent players. One might expect to observe that students who consider themselves frequent players to have a higher average amount of time spent playing video games than infrequent players as was observed in this data set.

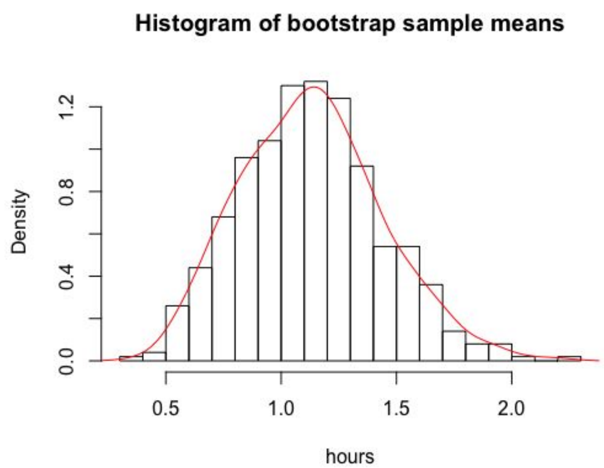
In order to assess whether or not the exam in the week prior to the survey affected whether or not frequent and infrequent players played any video games that week, the proportion of students who played video games at all in the week prior to the survey was calculated for both frequent and infrequent players. Then, 95% confidence intervals were constructed for both frequent and infrequent players using the sample proportion as a point estimator, standard deviation estimated using the sample proportion, the finite population correction factor, and normal quantiles. A 95% confidence interval for the true proportion of frequent players who played some amount of video games in the week prior to the survey based in the point estimate, the sample proportion, .838 is (.736, .939). This means that the true proportion of frequent players who played at least some video games in the week prior to the survey is between 73.6% and 93.9%, so we may conclude that most of the students who are frequent players played some video games in that week. A 95% confidence interval for the true proportion of infrequent players who played some amount of video games in the week prior to the survey based in the point estimate, the sample proportion, .0732 is (.00516, .141). This means that the true proportion of infrequent players who played some video games in the week prior to the survey is between 0.516% and 14.1%, so we may conclude that most of the students who are infrequent players did not play any video games in that week. Viewing these two confidence intervals in the context that there was an exam in the week prior to the survey, we may make assumptions on whether or not the exam had an effect on whether or not students who are either frequent or infrequent players chose to play video games that week. It appears that the exam did not have a strong effect on the frequent players because their confidence interval for the true proportion of students who played any video games in the week prior to the survey indicates that most of the frequent players played video games in the week prior to the survey. However, the exam in the week prior to the exam may have had an effect on the infrequent players, because their confidence interval for the true proportion of students who played any video games in the week prior to the survey is near zero. If it is true that more infrequent players would have played video games in the week prior to the survey if there had not been an exam, then our estimate for the proportion of students in the entire sample who played video games in the week prior to the survey may be low because the proportion of infrequent players who played video games in the week prior to the survey is low.

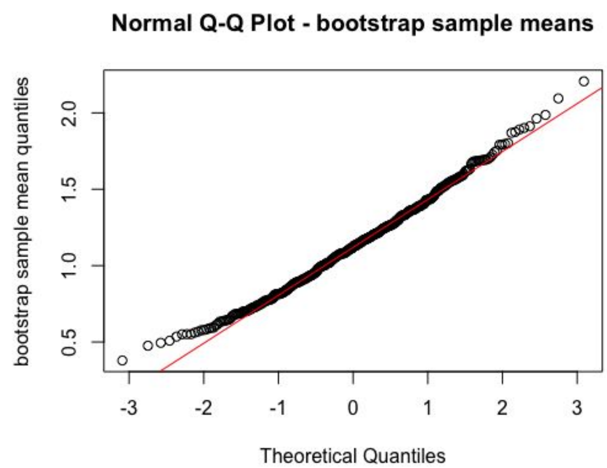
*** [Scenario 3:] Consider making an internal estimate for the average amount of time spent playing video games in the week prior to the survey. Keep in mind the overall shape of the sample distribution. A simulation study may help determine the appropriateness of an interval estimate.**

First a density histogram, with overlaid density curve in red, was constructed for the distribution of time students spent playing videogames in the week preceding the survey. This histogram is skewed right, indicating that most students played few or no hours of video games in the week prior to the survey. Further, this skewness and the bumpiness of the density curve indicate that we cannot use a normal approximation to generate confidence intervals for the true average amount of time spent playing video games in the week prior to the survey. This suggests the use of bootstrap to generate a distribution of the sample average number of hours played.



To derive a distribution for the sample mean for hours spent playing video games a bootstrap population of 314 was constructed from the 91 person sample by repeating each element approximately three times. Then 500 samples of 91 were taken from the population, each done without replacement to mimic the survey's sampling procedures, and their means were recorded. To visualize the distribution of sample means, a density histogram, with density curve overlaid in red, and a normal QQ plot comparing the distribution of sample means to the normal distribution was produced and a theoretical line showing where a true normal distribution would lie was overlaid on the plot in red. The normal QQ plot shows that the bootstrap distribution of sample means is roughly normal. However, we chose to use the bootstrap distribution of sample means, instead of a normal approximation, to generate confidence intervals to produce a more accurate interval estimate, and to account for the deviations from the normal distribution most evident in the tails of the bootstrap distribution of sample means.

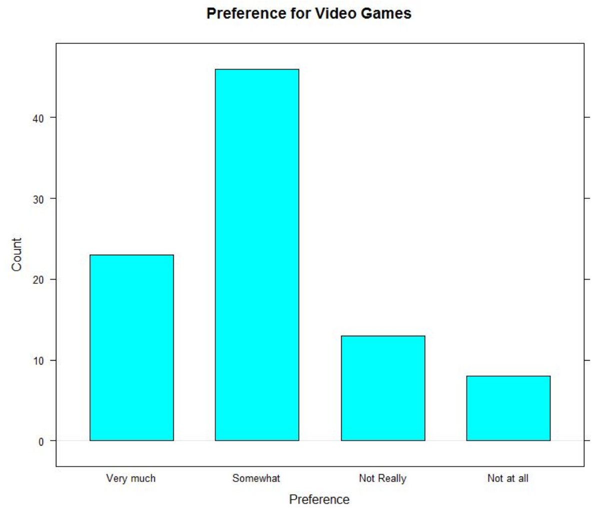


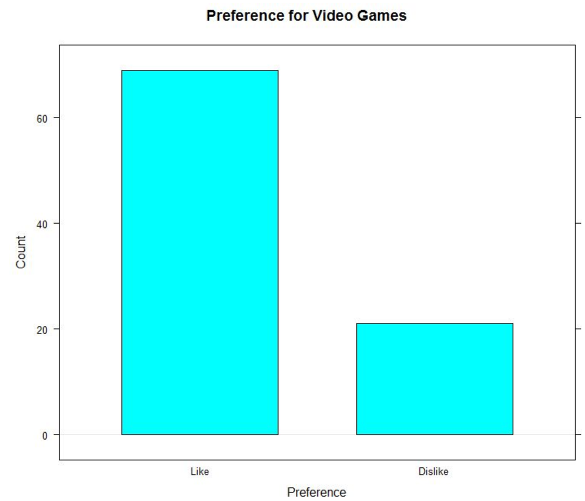


Then a 95% confidence interval for the true average amount of time students spent playing video games in the week prior to the survey was constructed using the sample average as a point estimate for the true average number of hours spent playing video games, the standard deviation and quantiles from the bootstrap distribution of sample means, and the finite population correction factor. A 95% confidence interval for the true average amount of time students spent playing video games based on the point estimate, the sample mean, 1.24 hours is (1.39 , 1.7) hours. This means that the true average amount of time a student spent playing video games is between 1.39 hours and 1.7 hours. Because the purpose of the survey is to design a useful and well liked statistics lab, the 95% confidence interval for the true average amount of time students spent playing videogames in the week prior to the survey can be used to suggest the amount of time each week is allotted to the statistics lab. In particular, we would recommend that the lab meet between about one and a half hours and two hours each week .

*** [Scenario 4:] Next consider the "attitude" questions. In general, do you think the students enjoy playing video games? If you had to make a short list of the most important reasons why students like/dislike video games, what would you put on the list? Don't forget that those students who say that they have never played video games or do not at all like video games are asked to skip over some of these questions. So, there may be many nonrespondents to the questions as to whether they think video games are educational, where they play video games, etc.**

The possible responses were "Very much","Somewhat", "Not Really", "Not at all", and "Never Played." "Never Played" was equated with "Not at All." After relabeling, the distribution across categories is seen below. Clearly the majority of students either Very Much or Somewhat liked video games. To answer the question whether students actually enjoy playing video games, we used a chi-squared test with the null hypothesis that the proportion of students in each response group was equal. In other words, that the count for each category was equal to 15 out of the 90 responses, the outcome if each category were equally likely. The results of the test showed that the proportions are significantly different from chance (X-square = 37.911, df =3, p-value <<0.05).





Refactoring the categories into a binary Liked or Disliked showed similar results. (X-square = 25.6, df =1, p-value <<0.05). In both cases, correcting for sample size did not affect the significance.

We broke down the data by cross-tabulations to investigate whether responses to certain categorical variables were associated with changes in the distributions of preferences. See the like_xtabs.pdf and bin_xtabs.pdf files to see all of the barplots. Based on these plots, it seems that having a home gaming system actually biases preference towards disliking video games. Instead, having a home computer or going to arcade favors liking video games. It also seems that having a home computer does affect liking video games. Students who played daily or weekly were more likely to like video games, whereas students who played semesterly or monthly were less likely to like video games. Students who played video games when busy were more likely to like video games. One of the strongest predictors of liking video games seems to be whether or not one considers video games educational. Having had no computer at home in high school makes it more likely to like video games. Students who liked math were more likely to like video games. Students who owned a PC were less likely to like video games. Students who had a CDROM were more likely to like video games. Having email only slightly affected liking video games. Expected grade distributions were very similar between liking and not liking video games. Males predominantly liked video games. Importantly, looking at the four level barplots of preference, rather than the binary like/dislike barplots, it seems that the B students tend to dislike video games, whereas the A students tend to like video games.

We also aggregated numerical variables, such as age, hours worked for the class within a week of the survey, and hours played video games within a week of the survey, based on preference for video games. See the bin_aggs.pdf and like_aggs.pdf files. From these graphs, we see that older students liked video games, students who had played more in the last week liked video games, and students who spend longer working liked video games. To verify this, two-sample t-tests were run. For work against liking vs. dislike, t = -1.3137, df = 36.517, p-value = 0.1972. For time spent playing in the last week against like vs. dislike, t = -3.1241, df = 68.291, p-value = 0.002617. For age agaisnt like vs. dislike, t = -1.1564, df = 81.385, p-value = 0.2509. Thus, only time spent playing within the past week, as one would expect, was significantly predictive of liking vs. disliking video games. The other two variables do seem like they are trending.

In order to test what other factors contribute to whether or not a student likes or dislikes video games, we conducted a logistic regression with liking vs. disliking as the response variable. Considering that 26 out 91 of the subjects would have been excluded from the regression based on missing values in the predictors, we imputed the missing values based on simple random sampling of the observed values for each variable. Had we not imputed the data, our first order model would have failed to converge.

We excluded interaction variables for our first model and found that with the imputed data, the model converged with a statistically significant fit: X-square = 47.917, df = 21, p = 0.0007. Following up with Wald tests on the coefficients found that time spent playing video games within the past week, whether or not games were considered educational, whether the student had a computer in the home during high school, whether the student owned a computer and had a CDROM, were all predictive of whether the student liked video games.

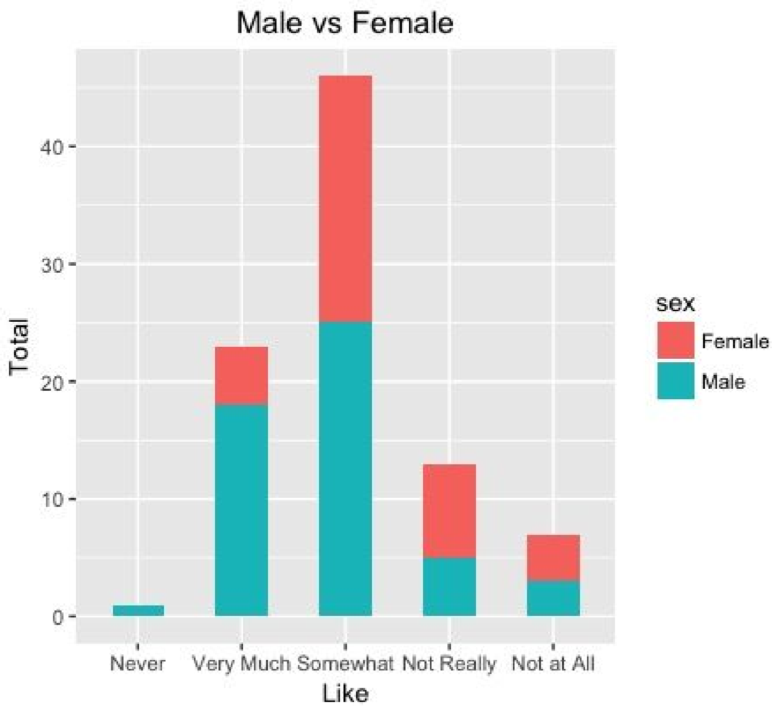
A second model was made that included interactions between expected grade, whether they played video games when busy, and whether they thought that video games were educational, against all other predictors. This model again converged with significant fit: X-square = 98.317, df = 76, p = 0.04. Follow up Wald tests showed that there were interactions between frequency of playing video games and expected grade, age and expected grade, whether video games are educational and sex. Higher order interaction effects could not be modeled due to insufficient sample size.

From the above results, in general it is clear that students like playing video games. The main factors determining whether or not students like playing video games are time spent playing, whether they are considered educational, whether the student had a computer in the home during high school, and whether the student owned a computer and had a CDROM. The interaction effects seem to be responsive to variables identified visually in the barplots as showing differences in preference for video games, i.e. frequency of playing video games, expected grade, age, sex, and whether video games are considered educational. Clarifying these interactions through visual inspection are necessary.

MATH189/289 Homework 2
Brendan Gee 4th Year Math Econ BS
Robert Gougelet 5th Year Cognitive Science PhD
Sidney Browne 4th Statistics BS

*** [Scenario 5:] Look for the differences between those who like to play video games and those who don't. To do this, use the questions in the last part of the survey, and make comparisons between male and female students, those who work for pay and those who don't, those who own a computer and those who don't. Graphical display and cross-tabulations are particularly helpful in making these kinds of comparisons. Also, you may want to collapse the range of responses to a question down to two or three possibilities before making these comparisons.**

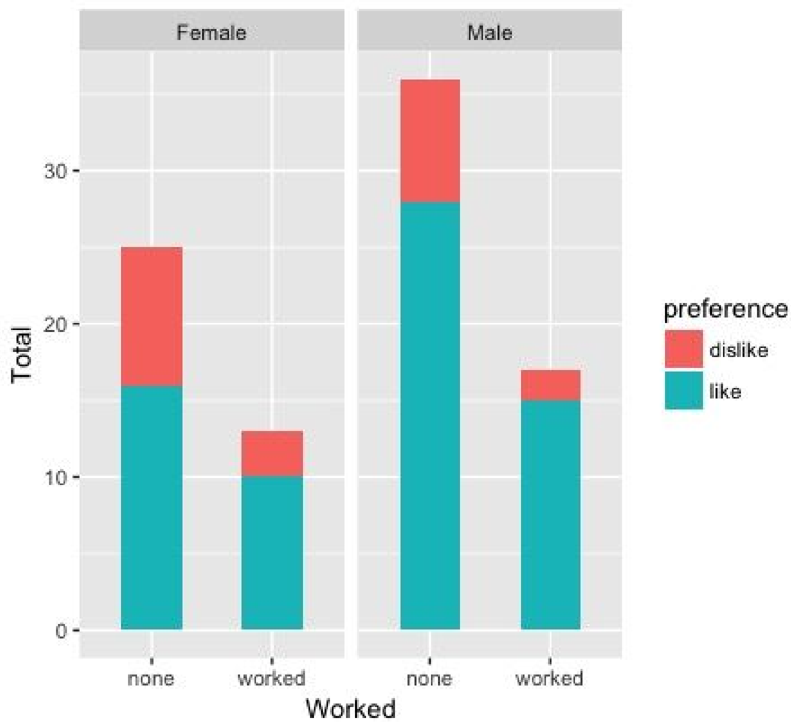
First, we compared male and female students, and whether there were clear results on one gender liking to play more than the other. The graph below displays a comparison between men and women and their preference on liking video games:



It is important to note that the vast majority of men either somewhat like to play or very much like to play video games. There is also a noticeable disparity between men and women who very much like to play. In addition, we can see that women match up equally against men in not at all liking to play and women have a majority in not really liking to play, which is significant since there is a smaller proportion of women in the sample data(38 women, 53 men). Therefore, doing a chi-squared test, we can reject the null hypothesis that women and men enjoy playing video games equally, meaning that in general, female students don't like playing video games as much as male students.

Then, we created a new variable "preference" by making a function of the "like" variable and combining it into the data frame. We assigned "never played", "not really" and "not at all" to **dislike**, and "somewhat" and "very much" to **like**. We are assuming if you never played video games that you have stayed away from them for a reason, and hence dislike them. We made this variable in order to obtain a better understanding of the data and to classify the data into two categories instead of 5.

Next, we looked at whether work affected if a student liked to play games. In doing this, we created a new variable, to extract the students who worked for pay versus the students who didn't. Then, using our new preference variable, we compared to see whether working for pay had an effect on gaming. In the graph below, we can see a significant pattern that tells us a lot about the data:



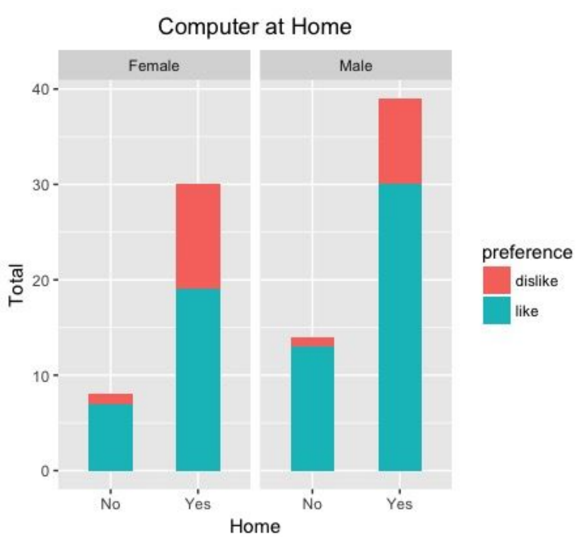
It is evident to see that students, whether male or female, seemed to enjoy playing video games more if they were working for pay. There could be several reasons for this; and we can look at the follow up surveys to help us better understand their motivations.

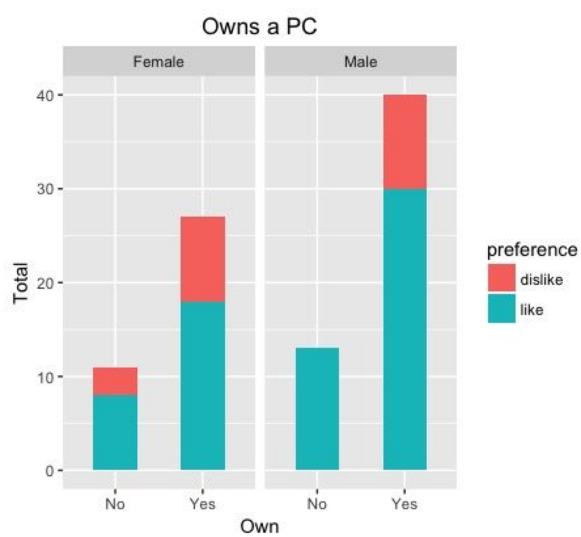
Possibility 1: If we look at the follow up survey data table 2, it describes reasons for why students play video games. One of the most common answers among this, at 66%, was relaxation. After a stressful day of school and work, it can easily be seen that students want to find a way to relax and take their minds off of a long day. We can interpret the data in the graph above as exactly that; explaining why the majority of students who worked also liked to play video games.

Possibility 2: Looking at the follow up survey data table 3, it summarizes what students dislike about video games. A significant proportion(40%) answered that video games cost too much, which relates to the above graph as well. In other words, students who liked playing video games needed a way to keep up with their spending, and hence resorted to getting a job to help fund their addiction. We can also argue that students uninterested in video games wouldn't have the problem of affording them and would therefore not need a job in order to keep up with the spending required to fund gaming.

From both of these possibilities and the follow up surveys, we can see that students who worked for pay generally enjoyed video games more than those who didn't work for pay. This is because of the complaint of costs that video games provided a student, and the benefit of relaxation after a day of work and school.

Finally, we are asked to make comparisons between students who own a computer and those who don't. In the data, we are given two variables related to this: home(computer at home) and own(own PC). We do this in a similar way; by comparing those who don't have a computer to those who do and whether not there is any pattern in the graphs. We see an interesting result from the two graphs below:





From these graphs, we can see that if you don't have a computer at home or own a PC, you are more likely to enjoy video games, regardless of gender. This can be interpreted in several different ways, as we look at the follow up surveys to proceed:

Possibility 1:In table 3 of the follow up survey, it is important to note that 48% believed that video games take up too much time. If you have regular access to a computer or own a PC, you also have regular accessibility to playing video games, since you don't have to go to a lab in order to play. This is important, and students who are well aware of how time consuming video games realize how much they can interfere with school and other activities. So while having a computer may mean a student has easier access to gaming, it can also be realized that it is distracting and make a student dislike wasting time on something significant. In table 3, another big category result was that video games were pointless(33%) and boring(17%). This also relates to how students with convenient computer accessibility might feel. On the flip side, students with no PC or access to a computer won't have this problem, and most likely play games at their convenience in moderation. This is due to the fact that they have to go out of there way to a computer lab or friends house in order to Therefore, they won't complain about how much time they're wasting and are instead just enjoying the games. Hence, it can easily be seen why students without instant computer accessibility might enjoy playing games more, since they don't have to worry about wasting all of their time on them.

Possibility 2: Again, we can argue that video games costing too much may result in students with computer accessibility to enjoy video games less. This is because to play video games on your electronic device, you have to buy them first, which is something students with no easy access to a computer don't have to deal with. Buying several games starts to add up, especially as a college student, which indicates that students with computers may not enjoy gaming as much if they keep spending all their cash. Therefore, we can understand why many students with computer accessibility want to refrain from playing, as it is due to cost.

With both possibilities and information from the follow up surveys, we conclude that students without PC's or computers at home are more likely to enjoy video games since they don't have to deal with the costs associated with video games. In addition, they are less likely to worry about videogames taking up too much time or becoming bored, since they aren't always accessible.

In total, we have made comparisons between gender,work for pay, and owning a computer to determine if any of these factors have an influence on students enjoyment of video games. We have concluded that males typically enjoy games more than women, students who work have a greater tendency to enjoy video games than those who don't, and finally, students who don't own a computer at home or a PC tend to enjoy video games more than those who do. Hence, it is sufficient to say we have concluded this scenario.

*** [Scenario 6:] Just for fun, further investigate the grade that students expect in the course. How will does it match the target distribution used in grade assignment of 20% A's, 30%B's,40% c's and 10%D's or lower? If the nonrespondents were failing students who no longer bothered to come to the discussion section, would this change the picture ?**

Brendan Gee 4th Year Math Econ BS
Robert Gougelet 5th Year Cognitive Science PhD
Sidney Browne 4th Statistics BS

We can compare the distributions of the expected grades vs. actual grades of the students using another chi-squared test with target proportions: A = 0.2, B = 0.3, C = 0.4, D or less = 0.1. The proportions of expected grades for students was A = 0.34, B = 0.57, C = 0.09, D or less = 0. The chi-squared test showed that the students had significantly different predictions of their grades (X-square = 62.61, df =3, p-value <<<0.05). The total number of sampled students for the survey was 95, meaning that 4 students did not respond to the survey at all. Assuming that they were failing, we can change the expected grade distribution to A = 0.33, B = 0.55, C = 0.08, D or less = 0.04. Repeating the chi-squared test again, we see that this does not change the significance of the results. (X-square = 53.83, df =3, p-value <<<0.05). So much fun!

Conclusion

Our analysis indicates that the majority of students enjoy playing video games. This means that designing a statistics computer lab using favored aspects of video games is likely to appeal to the majority of students who will participate in those labs in the future. In addition, students who don't own a computer would be more likely to participate in the computer lab if video games are present, as well as those that work for pay. This appeal for video games would further encourage these students to go to lab regardless of circumstance. Our specific suggestions for improving the computer lab based on the results are as follows: We suggest a lab that meets between 1.5 and 2 hours per week, based on the average amount of time students spent playing video games in the week prior to the survey. Further, the exam given in the week prior to the survey likely had an effect on whether or not students played video games that week, so a lab duration longer than our suggestion may be appropriate.

Gathering information about the students themselves before deciding year-to-year whether or not to add a lab component to the class would be useful. For instance, if students tended to play more video games during the week, they would be more likely to enjoy lab. Older students also prefer video games, and likely labs, as well. Students who have experience with computers at home would also benefit from the labs. Because students are more likely to enjoy video games when they are educational, this suggests that students would be amenable to educational labs on the computer, especially if the labs are more like video games. We also suggest that females receive more attention in the labs, as they are less likely to like the labs based on their preference for video games.

Regarding expected grades, students were overconfident in their performance in the class. Based on a logistic regression predicting expected grades, the model converged with a statistically significant fit: X-square = 54.178, df = 75, p = 0.004. Students expected better performance in the class when they were male, older, had a computer at home during high school, and played video games more often. This confounds the relationship that these variables have with liking video games. At the least, it would be pertinent to manage the expectations of the students who fit this profile when running the lab, but it is also likely that these students would benefit most from having a lab, as well.