Case Study 2: Who Plays Video Games

Yu-Chun Chen, A13356506
Yanyu Tao, A13961185
Bolin Yang, A92111272
Shuibenyang Yuan, A14031016
Haoming Zhang, A14012520
Mingxuan Zhang, A13796895

Introduction:

In UC Berkeley, a committee of faculty and students has determined to design a series of computer labs for undergraduate students who enrolled in introductory statistics courses to satisfy quantitative reasoning requirement, which is around half of 3,000 to 4,000 statistics courses students every year. These labs aim to extend traditional courses by providing an interactive learning environment where students can learn statistics and probability in an alternative way. The process of designing computer labs is then compared to that of designing video games. Thus, to improve their design of the labs, the committee at UC Berkeley conducts a video game-related survey on undergraduate students enrolled in a lower-division statistics course. The survey collects data on the extent students play video games and also what aspects of video games they like and dislike.

Before we go into our own investigation, some literatures are reviewed. Past researchers have proven the causal relationship between playing video games and these aspects. Shirley Matile Ogletree and Ryan Drake investigate the relationship between gender and gaming, which shows that college men spend more time on video games. Also, men and women have different preference towards video games. They show that playing video games interfere with sleeping and class preparation for college students. Male college students are more likely to be negatively affected than female college students since playing video games interfere with men's sleeping and other college activities more, such as interpersonal relationship and academic performance (Shirley Matile Ogletree & Ryan Drake). Playing video games also connects to the players' personality. In "Video gameplay, Personality and Academic Performance," it is shown that college students who are more open and extroversive prefer violent video games while college students who are neurotic like playing non-violent video games. Matthew Ventura, Valerie Shute, and Yoon Jeon Kim also find that playing video games affect GPA negatively since many video games can be reached on mobile devices and encroach on study time. However, in "Digital Games in Education," the author concludes that user-centered digital games are considered to provide the skills used in the education field, such as the ability of reading images and getting visual attention, through developing co-operation, promoting challenges, and problem-solving strategies (Begoña Gros). This research shows the possibility of applying some aspects of video games to designing computer labs for students to learn.

In this study, **we will analyze the survey data and interpret students' answers to questions regarding various aspects of video games, so we can propose some advice for the committee to better construct computer labs**. Before we are able to provide any suggestion on designing labs, we have to understand the methodology used in the survey. A survey is a scientific methodology that collects data from individuals. It is usually conducted on a sample group that represents a population for the purpose of description, exploration, and explanation. A good survey should be quantitative, self-monitoring, contemporary, replicable, systematic, impartial, representative, theory-based. Moreover, all individuals in the population must have equal chance of being selected so the sample can generalize to total population. If sampled appropriately, a survey conducted on this sample can be more accurate than a census. However, if the population is identical, sampling would be unnecessary.

Data:

95 students were selected out of 314 students enrolled in Statistics 2, Section1, during Fall 1994 to participate in the survey. Our data contains responses to survey questionnaire from 91 students who completed the survey out of 95 selected students. Variables of survey results are summarized in the table below.

Variable	Description	Variable Type
`time`	# of hours played in the week prior to survey	numerical
`like`	how much one likes playing video games 1=never played, 2=very much, 3=somewhat, 4=not really, 5=not at all	categorical
`where`	where the respondent usually plays video games 1=arcade, 2=home system, 3=home computer, 4=arcade and either home computer of system, 5= home computer and system, 6=all three	categorical
`freq`	how often the respondent plays 1=daily, 2=weekly, 3=monthly, 4=semesterly	categorical
`busy`	whether the respondent plays video games when busy 1=yes, 0=no	categorical
`educ`	whether the respondent believes that playing video games if educational 1=yes, 0=no	categorical
`sex`	gender of the respondent 1=male, 0=female	categorical
`age`	respondent’s age in years	numerical
`home`	whether the respondent has computer at home	categorical

	1=yes, 0=no	
`math`	whether the respondent hate math 1=yes, 0=no	categorical
`work`	# of hours worked the week prior to the survey	numerical
`own`	whether the respondent owns a PC 1=yes, 0=no	categorical
`cdrom`	whether the respondent's PC has CD-Rom 1=yes, 0=no	categorical
`email`	whether the respondent has email 1=yes, 0=no	categorical
`grade`	the grade expected by the respondent 4=A,3=B,2=C,1=D,0=F	categorical

Table 1: Detailed Information of Variables of Data in Survey Results

In the survey results, if there were any questions that were not answered or if any was answered improperly, then the answers were coded as 99. Several questions were skipped for those who had never played video games or who did not at all like playing video games.

There was a second part of the survey that covers whether the respondent likes or dislikes playing video games. Questions in this follow up survey were different in that more than one response may be given.

First, students were asked about what types of games they played and they were allowed to check up to three types that they played. This question was skipped for those who had never played video games or did not at all like playing video games. The results to this question are summarized in the table below. In the table, the percentage represents the proportion of respondents who play this certain type of video games.

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

Table 2: Stats to Question "What types of games do you play?" (at most three answers)

Students who did answer the first question were then also asked to provide reasons why they play the games they do. Again, they were asked to select up to three such reasons. Their responses are presented in the table below.

Why	Percentage
Graphics/Realism	26%
Relaxation	66%
Eye/hand coordination	5%
Mental Challenge	24%
Felling of mastery	28%
Bored	27%

Table 3: Stats to Question "Why do you play the games you checked above?

(at most three answers)

Finally, all students were asked to answer the third question, which asked for the aspects about video games that the students did not like. Like both questions above, they were asked to select up to three reasons.

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

Table 4: Stats to Question "What don't you like about video game playing?"

(at most three answers)

Background:

In this study, all of the population studied were undergraduates enrolled in Introductory Probability and Statistics, Section 1, during Fall 1994. This class is the low-division prerequisite for students majoring in Business. This class is composed of the lecture from 1 to 2 pm on MWF for maximum four hundred students, and the one-hour discussion on either Tuesday or Thursday. There were ten discussion sections for the class, with approximately 30 students individually. The list of all students who had taken the second exam of the semester was used to select the students to be surveyed and the exam was a week before the survey. Each student was assigned a number from 1 to 314 and then a pseudo random number generator selected 95 numbers between 1 to 314. There was a three-stage system of data collection to limit the number of non-respondents. The collector must visit both Tuesday and Thursday discussion sections. One week before the survey, students took the exam, and the graded papers were returned to them during the discussion section in the week of the survey, which ensured that all the students attending the discussion session. If there were still student not showing up during the discussion, they would be reached in lecture on Friday. Students were informed the purpose of the survey and all of them were anonymity. Finally, 91 students completed the survey. All the above qualification ensures the random sample of the survey.

For the video games, they are classified according to the device on which they are played and skills needed to play the game. Arcade games is fast and need eye and hand coordination. Console games are more likely to be action, adventure or strategy games. PC games are simulation and role-play exclusively.

Investigation:

1. Scenario One

We begin by estimating the fraction of students who played a video game in the week prior to the survey. We estimate by both point estimate, in this case, the percentage of students in the sample who played a video game in the week prior to the survey, and interval estimate, which is a confidence interval that might contain the true population fraction of students who played a video game in the week prior to the survey.

<u>Point Estimate</u>

Sample mean, $\bar{X} = 0.3736$

<u>Simple Confidence Interval by Central Limit Theorem</u>

Since our sample size, 91 students who completed the survey, is larger than 30, it is large enough for us to apply Central Limit Theorem. According to CLT, if the sample size is large, the probability distribution of the sample average is often well approximated by the normal curve. Although in our case, the sample students are not independent and identically distributed, the normal approximation can still hold as in addition to the sample size being large, it is also not too large (91) relative to the population size (314). Thus, normal distribution can be used here to provide the confidence interval for the population parameter, in our case, the fraction of students who played a video game in the week prior to the survey.

The formula for a 95% confidence interval is $(\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}})$, and the 95% confidence interval for the fraction of students who played a video game the week prior to the survey is (0.289, 0.494).

Confidence Interval with Finite Sample Population Correction Factor

We usually ignore the finite population correction factor when $Var(\bar{X}) \sim \sigma^2/n$. Without the correction factor, we are assuming that our sample is drawn with replacement or that we are sampling from an infinite population. Since both are not true for our case, finite population correction factor should not be ignored. The fact that we should not ignore the correction factor is also shown as $\sqrt{314 - 91}/\sqrt{314 - 1} = 0.84$.

The 95% confidence interval with finite population correction factor is $(\bar{X} - 2\frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n-1}}\sqrt{\frac{N-n}{N}}, \bar{X} + 2\frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n-1}}\sqrt{\frac{N-n}{N}})$, and the 95% confidence interval with finite population correction factor for the fraction of students who played a video game the week prior to the survey is (0.288, 0.460).

Point and Interval Estimation with Bootstrap Resampling Technique

Bootstrap applies to finite samples, which is true for our case, and provides numerical solutions for non-standard situations so that it is particularly appealing when dealing with finite populations and complex sampling designs. Thus, we will use bootstrap resampling method here to obtain a mean estimate and a confidence interval. We take 1000 bootstrap samples from the bootstrap population. The distribution of these 1000 bootstrapped mean in shown in Figure 1 below.

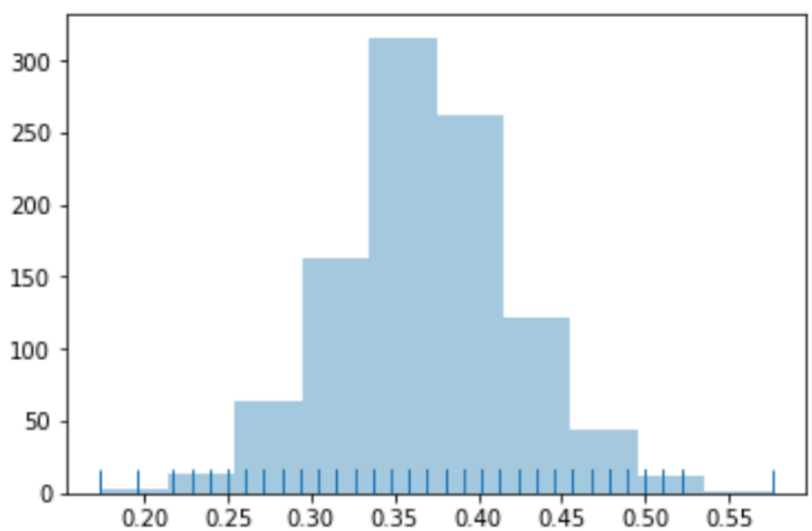


Figure 1: Distribution of 1000 Bootstrapped Means

We then compute the expected mean value from the bootstrapped population, which is $\bar{X} = 0.369$, and construct a 95% confidence interval of (0.271, 0.478). This means that our estimated fraction of students who played a video game the week prior to the survey to be 36.9%, and that the probability that fraction of students who played a video game the week prior to the survey is between 27.1% to 47.8%.

Comparison

By comparing all three confidence intervals, we find that the confidence interval computed with finite population correction factor has the smallest range, followed by the confidence interval by Central Limit Theorem, then the confidence interval computed by bootstrap resampling. Thus, we conclude that the confidence interval computed with finite population correction factor is the most precise interval estimation among all three.

2. Scenario Two

The box plots shown below in *Figure 2* display the five distributions of the total amount of time students spend on playing video games under the corresponding game-playing frequency categories of daily, weekly, monthly, per semester and never. First, we notice that the medians of the categories of daily and weekly are similar, and they are higher than those of the other three categories. This indicates that students who play games more frequently have the tendency to spend more time playing.

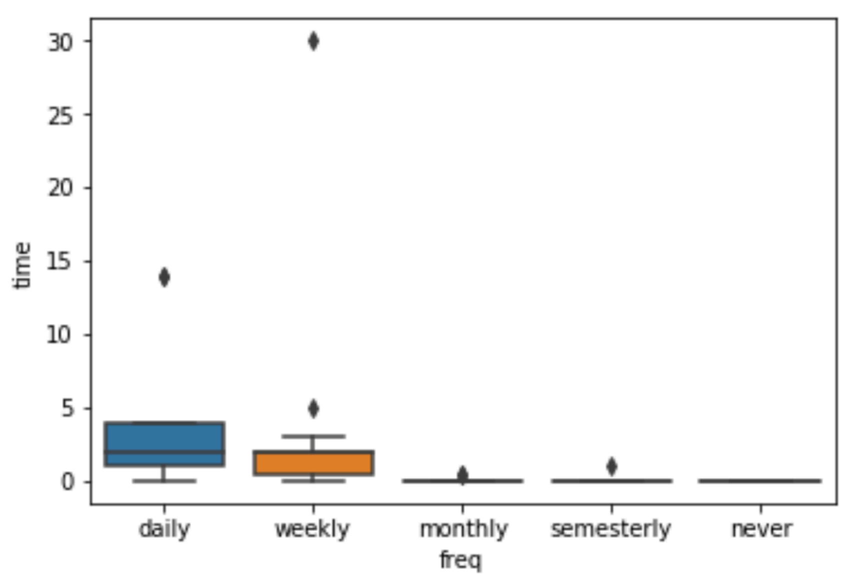
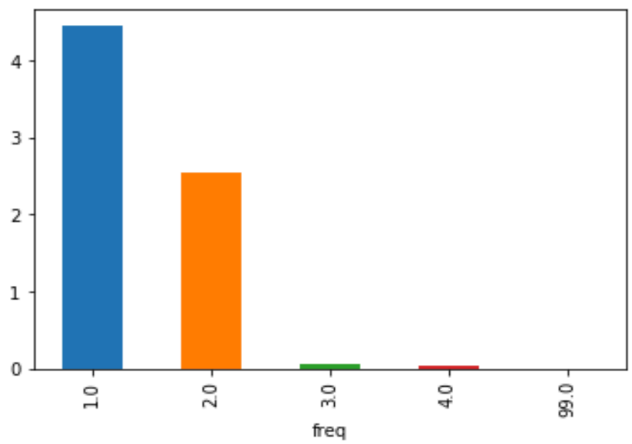


Figure 2: Boxplot of Game-Playing Frequency v.s. Game-Playing Time

Taking a closer look at the box plots of the first two categories, though their medians are at approximately the same level, the interquartile range of the first is greater the that of the second, which indicates that students playing daily have more variety in playing time than students playing weekly. In addition, the first box plot lies comparatively higher than the second one in which the first quantile and third quantile of the first are higher than those of the second, suggesting that students playing daily tend to play for longer time. Moreover, the first box plot is positively skewed as the second is negative skewed. The skewness also proves that students playing daily are concentrated at the zone with more time spent while students playing weekly are concentrated at the zone with less time spent. Thus, we can conclude that students play video games more frequently not only have more various time range of playing but also are more likely to spend greater amount of time on playing.

In the bar chart shown in *Figure 3*, categories of 1.0, 2.0, 3.0, 4.0, and 99.0 correspond to playing daily, weekly, monthly, per semester, and never respectively. Both the bar chart and the table shows that students playing daily has the greatest amount of average playing time among all groups, which, again, proves the above conclusion we draw from the box plots.



Frequency	Average Time (hr)
Daily	4.44
Weekly	2.54
Monthly	0.06
Per Semester	0.04
Never	0.00

Figure 3: Bar Chart of Average Time Playing According to Frequency of Playing
Table 4: Average Time Playing Amount According to Frequency of Playing

From *Table 5* below, we can see that the average time spent on playing is 5.14 hours for students who play when busy as the average time spent on playing is 0.985 hour for those who play when not busy. There are 5 students playing video game when busy with an average playing time of 7.2 hours under the daily category while 2 students playing video game when not busy with an average playing time of 1 hour under the same frequency category. For students who play daily, more people still choose to play even when they are busy, and they have a tendency to play for a longer amount of time. Under the weekly category, though less people choose to play when busy, those who still play also have a tendency to play longer. For students who play monthly or once a semester, they are not sensitive to how busy they are with respect to their playing time, as none of them play when busy. In fact, they rarely play even when they are not busy. We thus claim that those who play daily and weekly are more sensitive to how busy they are with respect to playing frequency and their playing time, and the average time they spent has a positive correlation with stress level. The more stress they are, or the busier they are, the longer and more frequently they play. Therefore, having an exam in the week prior to the survey might have skewed the data, and the preview estimates might be higher than the true value.

Frequency	number of students who played video games when not busy	Mean playing time for students who played when not busy	number of students who played video games when busy	Mean playing time for students who played when busy
Daily	2	1	5	7.2

Weekly	15	1.594	9	4
Monthly	2	0.059	0	0
Per Semester	0	0	0	0

Table 5: Number of Students and Mean Playing Time for Students Played When Busy and not Busy Based on Frequency

3. Scenario Three

The histogram of frequency of students' time spent playing video game in the week prior to the survey is shown in *Figure 4* below. X-axis represents the amount of time students spent on playing video games, while the y-axis represents the number of students who played such amount of time. We found that most students spent less than three hours playing video games in the week prior to the survey.

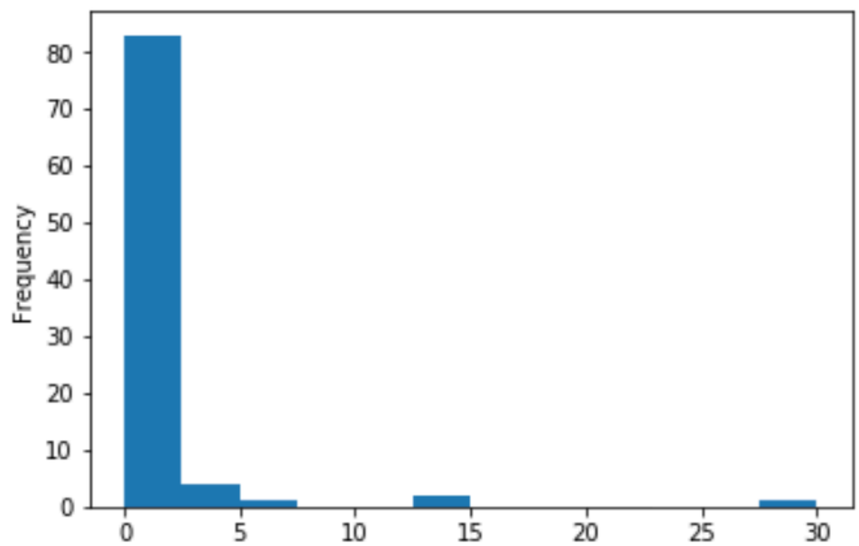


Figure 4: Histogram of Frequency of Students' Time Spent on Video Games in the Week Prior to the Survey

With these data, we will make an interval estimate for the average amount of time spent playing video games in the week prior to the survey. Our interval estimate will include three confidence intervals: by Central Limit Theorem, with finite sample correction, and by bootstrap resampling technique. Confidence intervals along with point estimates are concluded in the *Table 6* below. The point estimate represents the average time students spent playing video games in the week prior to the survey, and is used to construct our confidence intervals.

Type	Point Estimate	95% Confidence Interval
CLT	1.243	(0.451, 2.035)

w/ Finite Sample Correction	1.243	(0.576, 1.910)
Bootstrap	1.235	(0.573, 2.088)

Table 6: Point and Interval Estimate of Students' Time Spent on Video Games in the Week Prior to the Survey

By applying Central Limit Theorem, we get an estimate of average hour students spent playing video games in the week prior to the survey, which is 1.243 hours. The 95% confidence interval is (0.451, 2.035), indicating that there is 95% chance that the true average hour spent on video games of student population lies between 0.451 and 2.035.

We then also compute estimates with finite sample correction factor, and get an estimate of average hour students spent playing video games in the week prior to the survey same as that of by applying Central Limit Theorem, which is 1.243 hours. The 95% confidence interval with finite sample correction is (0.576, 1.910), indicating that there is 95% chance that the true average hour spent on video games of student population lies between 0.576 and 1.910.

Finally, we apply bootstrap resampling technique to estimate the average hour students spend on video games in the week prior to the survey. The distribution of 1000 bootstrapped sample means is shown below in *Figure 5*. The kurtosis value for this distribution is 2.97, and the skewness value is 0.53, which means that the distribution of bootstrapped means is approximately normal.

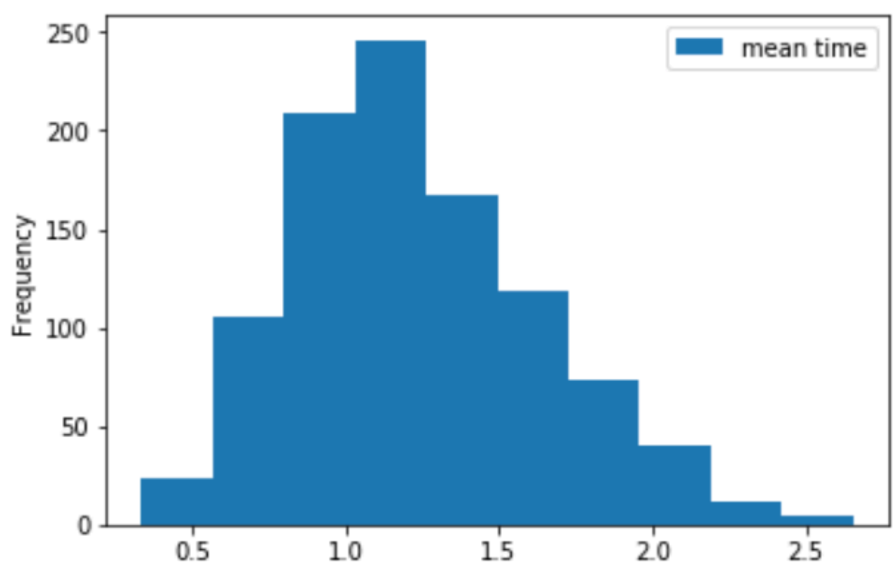


Figure 5: Distribution of 100 Bootstrapped Sample Means

We then compute the average hour students spent on playing video games in the week prior to the survey to be 1.235, and the 95% confidence interval with finite sample correction is (0.573, 2.088), indicating that there is 95% chance that the true average hour spent on video games of student population lies between 0.573 and 2.088.

4. Scenario Four

From the tables shown below in *Table 7*, we would like to show that college students enjoy playing video games overall. In the table of students' reasons for enjoying video games playing, relaxation is the most frequently checked answer with a significantly high percentage of 66%, 38% higher than the second frequent answer. This indicates that video game playing is a strategy for a lot of college students to relax themselves and sit back from their busy student lives. This reason with the highest frequency is realistic because college students, especially those from top universities like UC Berkeley, are commonly known as busy and stressed. It is understandable that they take playing video games as decompression thus enjoy it.

Why	Percentage
Graphics/Realism	26%
Relaxation	66%
Eye/hand coordination	5%
Mental Challenge	24%
Felling of mastery	28%
Bored	27%

Table 7: Stats to Question "Why do you play the games you checked above? (at most three answers)

In the table, shown below in *Table 8*, of students' reasons for disliking video game, gaming taking up too much time is the choice with the highest percentage of 48% checked by students. On the one hand, we understand that college student from top universities are busy with their school lives, so they don't have enough time for game playing. Certain types of games may require practicing time or longer playing time per round, which are not practical for busy students. However, not having enough time does not necessarily mean that students don't like video games. On the other hand, we can even interpret it as many students do not have sufficient self-control ability to stop them from keeping playing the games. They tend to spend more and more time on gaming such that they have less time on school stuff. This interpretation even shows that students actually enjoy video game playing so much that they spend too much time on it. The second frequently chosen reason is that playing video game costs too much money. Similar to how we interpret above, one possible interpretation of students complaining game playing costing too much can be that students, mostly with low income or no income, are not affordable of both living and costly games. Another interpretation can be that students tend to spend more and more money as they get more and more into game playing. Both interpretations show that claiming games being costly does not necessarily conflict with enjoying playing video game. Although there are some other reasons not listing in the tables in which students provide separately, the percentage of those reasons are so small that we decide to omit them.

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

Table 8: Students' answer: Why don't you like about video game playing?
(at most three answers)

List of reasons why student like video games:

1. Playing video games relaxes college students from their busy and stressed school lives.
2. Playing video games provides students with a sense of achievement and mastery.
3. Playing video games is one way of getting away from boredoms.

List of reasons why students dislike video games:

1. Playing video games takes up too much time that students do not have enough time with their student lives.
2. Playing video games cost too much money that students cannot afford it.
3. Playing video games is a pointless action that don't improve lives.

5. Scenario Five

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.

<u>Difference in liking playing video games between genders:</u>

Like to play video games:

0: students responded 1, 4, 5, 99 in the survey 1: students responded 2, 3 in the survey
Sex: 0: Female students 1: Male students

pref	like	notlike
sex		
female	26	12
male	43	10

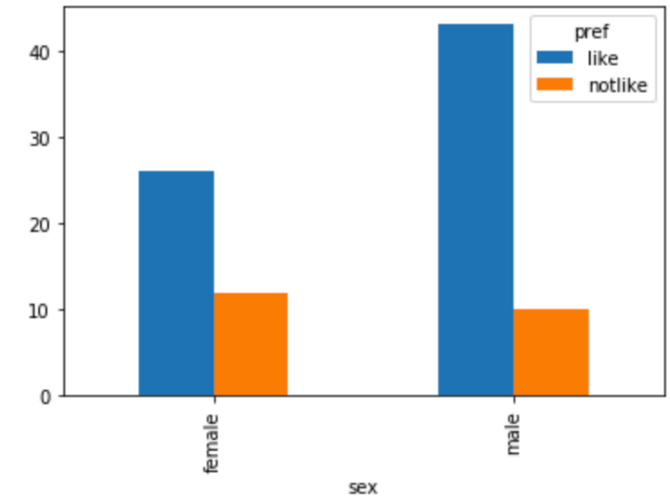


Figure 6: Bar plot of frequency of female and male students liking playing video games

We conducted Pearson's Chi-squared test for test of independence between two genders.

X-Squared = 1.319, df =1, p-value = 0.251. The null hypothesis is that gender is independent of whether liking to play video games, and the alternative hypothesis is that gender is not independent of whether liking to play video games. Since we have p-value equals to 0.251, we can reject the null hypothesis at 0.05 significance level and conclude that there is no difference between gender in liking playing video games.

<u>Difference in liking playing video games between work status:</u>

Like to play video games: 0: students responded 1, 4, 5, 99 in the survey 1: students responded 2, 3 in the survey
Work: 0: Students responded with 0 hours in the survey 1: Students responded with more than 0 hours in the survey

pref	like	notlike
work_or_not		
not_work	30	14
work	39	8

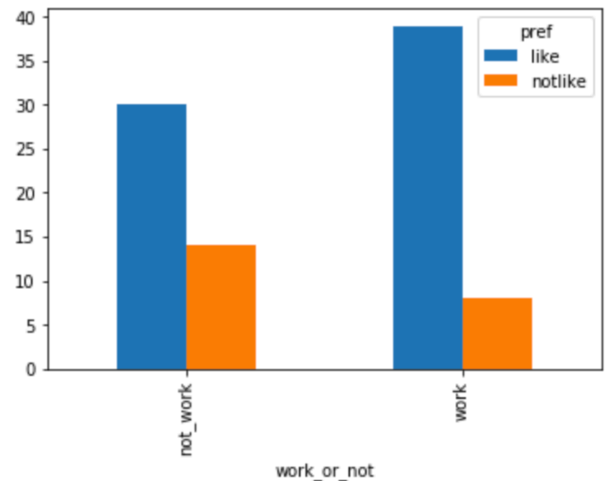


Figure 7: Bar plot of frequency of students who work for pay or not liking playing video games

We conducted Pearson's Chi-squared test for test of independence between students work for pay or students does not work for pay.

X-Squared = 1.967, df =1, p-value = 0.161. The null hypothesis is that work for pay status is independent of whether liking to play video games, and the alternative hypothesis is that work for pay status is not independent of whether liking to play video games. Since we have p-value equals to 0.161, we can reject the null hypothesis at 0.05 significance level and conclude that there is no difference between work for pay status in liking playing video games.

<u>Difference in liking playing video games between owning a PC or not:</u>

Like to play video games: 0: students responded 1, 4, 5, 99 in the survey 1: students responded 2, 3 in the survey
Own: 0: Students who do not own a computer (responded 0 in the survey) 1: Students who do own a computer (responded 1 in the survey)

pref	like	notlike
own		
not_own	21	3
own	48	19

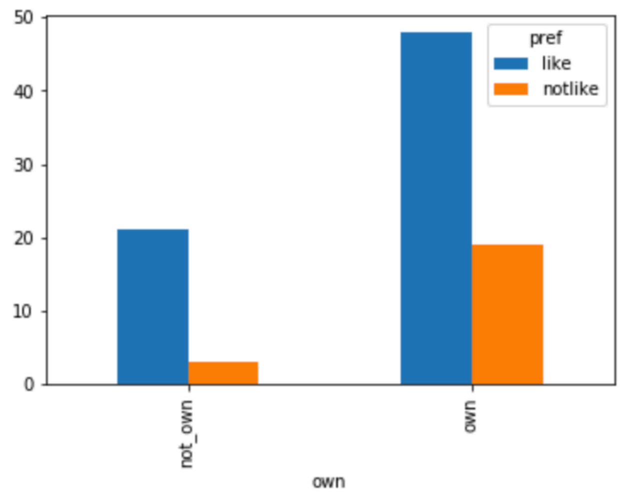


Figure 8: Bar plot of frequncey of students who own a PC or not liking playing video games

We conducted Pearson's Chi-squared test for test of independence between students own a PC or students does own a PC.

X-Squared = 1.636, df = 1, p-value = 0.201. The null hypothesis is that owning a PC or not is independent of whether liking to play video games, and the alternative hypothesis is that owning a PC or not is not independent of whether liking to play video games. Since we have p-value equals to 0.201, we can reject the null hypothesis at 0.05 significance level and conclude that there is no difference between owning a PC or not in liking playing video games.

<u>Conclusion</u>

From all the comparisons above, we can see that there is no association between student's gender and liking playing video games, no association between student's working status and liking playing video games, no association between student's whether owning a PC and the liking playing video games.

6. Scenario Six

Just for the purpose of fun, we will also further investigate the grade that students expect in the course and decide how well 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.

Comparison of sample students' expectation of grades versus target distribution of letter grades is shown below in *Table 9*. Bar chart of the comparison is also displayed below in *Figure 9*. From both the table and the chart, we see that expectation of grade A or B is much larger than the target distribution intends, while the expectation of grade C and D or lower is much lower than the target distribution of C and D or lower.

Letter Grade	A	B	C	D or lower

Expectation	0.3407	0.5714	0.0879	0.000
Target	0.2000	0.3000	0.4000	0.1000

Table 9: Expected Grades of Students v.s. Target Distribution of Grades

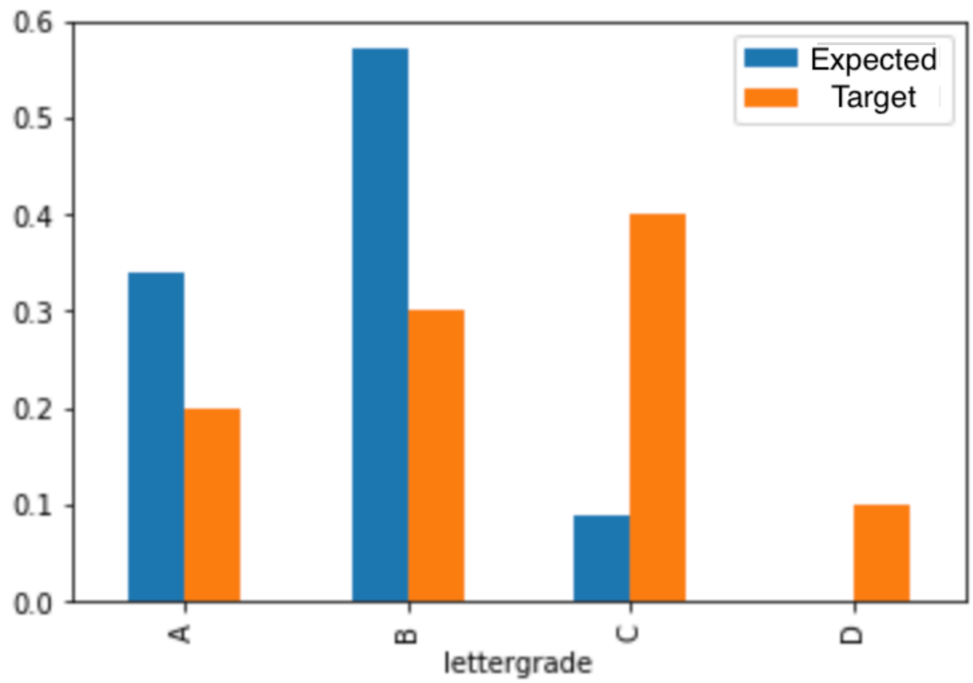


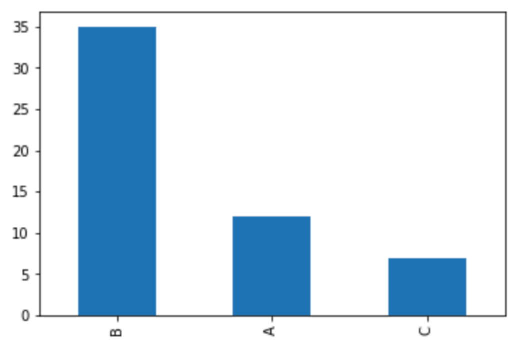
Figure 9: Expected Grades of Students v.s. Target Distribution of Grades

We then perform chi-square test goodness of fit, in which the null hypothesis is that both distributions are the same and the alternative hypothesis is that they are different distribution, to compare both distributions. Our test result shows a p-value of 1.63e-13, which is smaller than the significance level 0.05. Thus, we reject the null hypothesis and conclude that students' expected grades do not match the target grade distribution well. We also perform chi-square test goodness of fit test on bootstrapped data and obtain p-value of 1.34e-13, again confirming our conclusion that students' expected grades do not match the target grade distribution well.

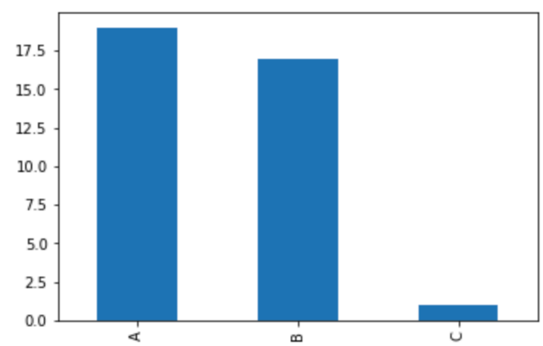
If the nonrespondents were failing students who no longer bothered to come to the discussion section, the distribution of students' expected grades will match the target grade distribution more. The fact that nonrespondents were failing students means that the number of students who expected letter grade of A, B, and C does not change, while the number of students who expected letter grade D or lower to increase and that the total students used as denominator to calculate percentage also increases. As a result, the percentages of students' expected grades for each of A, B, and C will all decrease, but the percentage of students' expected grades for D or lower will increase. This changes the picture in a way that make distributions of students' expectation and target distribution for A, B, and D closer, but further differentiate the distributions for C.

Additional Hypothesis:

To further investigate on the relationship between playing video games and students' academic situation, we decide to look into the association between the frequency students play video games and their expected grades. To do this, we classify those playing video games daily and weekly as those who play video games more frequently, and those playing video games monthly, semesterly, or not at all as those who play video games less frequently. From the bar charts given below in *Figure* 10, we see that most students who play video games less frequently tended to expect a letter grade of B, followed by others who expected an A or C. On the other hand, half of the students who play video games more frequently expect an A, with a little fewer people in this group expecting B, and a small proportion of people expecting C.



Students with less frequent video game time



Students with more frequent video game time

Figure 10: Bar plots for expected grades for students with different video games play frequency

We then perform a t-test on the mean of expected grades for the two groups by replacing A with 4.0 GPA, B with 3.0 GPA, etc. The null hypothesis is that there is no difference between the expected grades for students with less frequent video game time and students with more frequent video game time. The alternative hypothesis is that there is difference between the expected grades for students with less frequent video game time and students with more frequent video game time. The t statistic is -3.225, with p-value equals to 0.0018. Thus, we reject the null hypothesis under 0.05 significance level and claim that there is a difference between the means, and from the plots we can see that students with more frequent video game time tend to expect a higher grade.

Theory:

- **Goals**:
  - To determine the exact amount of average time spent on playing video games for the entire class by investigating on the survey results from sample students
- **Probability Method**:
  - Probability method provides us a scientific way to do select sample with knowing the relation between the sample and the population. Meanwhile, we can get the chance of each possible example. In our study, we applied simple random sample.
- **The Probability Model**:

- By the rule of simple random sample, each unit in population has the same chance (n/N) of being in the sample. However, there is dependence between selections. $\mathrm{P(unit\ \#1\ in\ the\ sample)} = \frac{n}{N}$

  $$\mathrm{P(\ unit\ \#1\ and\ unit\ \#2\ are\ in\ the\ sample)} = \frac{n(n-1)}{N(N-1)}$$
- In general, let I(1), I(2), … represent the first, second, … number drawn from the list 1, 2, …, N. Then, $\mathrm{P}(\mathrm{I}(1) = j_1, \mathrm{I}(2) = j_2, \dots, \mathrm{I}(n) = j_n) = \frac{1}{N(N-1)\dots(N-n+1)}$.

* **Sample Statistics**:
  - Let X_i be the value of characteristic of the ith unit in the population. Then, the population average is $\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$.
  - Let $X_{I(j)}$ represents the value of characteristic of the jth unit in the sample. Then $\mathrm{E}[X_{I(j)} = \sum_{i=1}^{N} X_i P(I(j) = i) = \sum_{i=1}^{N} X_i \frac{1}{N} = \mu$. $\mathrm{E}[\bar{x}] = \frac{1}{n}\sum_{j=1}^{n} E[X_{I(j)}] = \mu$.
  - $\mathrm{Var}(X_{I(j)}) = E\left[\left(X_{I(j)} - \mu\right)^2\right] = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2 = \sigma^2$

    $$\mathrm{Var}(\bar{x}) = \frac{1}{n^2} Var\left(\sum_{j=1}^{n} X_{I(j)}\right) = \frac{1}{n^2}\sum_{j=1}^{n} Var(X_{I(j)}) + \frac{1}{n^2}\sum_{j=1, j\neq k}^{n} Cov(X_{I(j)}, X_{I(k)})$$

    $$= \frac{1}{n}\sigma^2 + \frac{n-1}{n}\left(-\frac{\sigma^2}{N-1}\right) = \frac{1}{n}\sigma^2\frac{N-n}{N-1}$$

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

* **Estimators for Standard Error**:
  - when the variance of the random variable σ^2 is unknown, a common estimator for it is $s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_{I(j)} - \bar{x})^2$. In our study, since our sample was chosen by the simple random sample method and we assume our sample should roughly be in the same distribution as our population, we can then plug in s^2 for σ^2 and use $\frac{s^2}{n}\frac{N-n}{N-1}$ to estimate $\mathrm{Var}(\bar{x})$.
  - The reason for using s^2 is that the sample, when chosen by the simple random sample method, should look roughly like a small-scaled version of the population, so we plug in s^2 for σ^2 in the variance of $\bar{X}$
  - In fact, there is a slightly better estimate for $\mathrm{Var}(\bar{x})$. To estimate $\mathrm{Var}(\bar{x})$, we can use $s^2 = \frac{N}{N-1}\sigma^2 = \frac{N}{N-1}\sum_{j=1}^{n}(x_{I(j)} - \bar{x})^2$. Hence, an unbiased estimator of σ^2 is then $s^2\frac{N-1}{N}$, and the unbiased estimator of $\mathrm{Var}(\bar{x})$ is $\frac{s^2}{n}\frac{N-n}{N}$.

* **Population Total and Percentage**:
  - For some of the categorical data, we usually use the proportion as parameter, where the characteristic value x_i is 0 or 1 to denote absence or presence of the characteristic. For example, for I = 1, 2, …, 314, x_i=1 represents the ith student owns a PC while x_i=0 means he doesn't have one. Then, the total number of the students who have PCs is denoted as τ, where $\tau = \sum x_i$. The proportion of students

in the population is then denoted as $\pi = \frac{1}{N}\sum x_i$. In this case, we can use $\bar{x}$ as an unbiased estimate of π, while N$\bar{x}$can be used to estimate τ.

- The population variance and be unbiasedly estimated by $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \pi)^2 = \pi(1-\pi)$. Also, the estimator for the standard error is $\widehat{SE}(\bar{x}) = \frac{\sqrt{\bar{x}(1-\bar{x})}}{\sqrt{n-1}}\frac{\sqrt{N-n}}{\sqrt{N}}$.

- **Central Limit Theorem:**
  - If $X_1, \ldots, X_n$ are independent, identically distributed with mean μ and variance σ^2 then, for large n, the probability distribution of $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is approximately standard normal, no matter if X_i is normal distributed or not.
- **Confidence Intervals**:
  - From the Central Limit Theorem, we know that $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ approximately follows normal standard distribution, which can be used to conduct the confidence intervals for the population average, μ, when n is not too small or too large comparing to the population size.
  - One interval estimate of μ, called 68% confidence interval is $(\bar{x} - \frac{\sigma}{\sqrt{n}}, \bar{x} + \frac{\sigma}{\sqrt{n}})$, which means approximately 68% of chance that the parameter μ will fall in the interval. The other popular interval estimate is the 95% confidence interval, which is $(\bar{x} - 2\frac{\sigma}{\sqrt{n}}, \bar{x} + 2\frac{\sigma}{\sqrt{n}})$.
- **Bootstrap**:
  - For most of time, it's hard to tell whether the sample average follows the normal probability distribution without knowledge of the population. Bootstrap algorithms can then be applied to finite sample and provide numerical solutions for non-standard conditions, which can help us assess the estimators' accuracy and produce confidence interval for the population parameters.
  - Since the distribution of the sample generated by simple random sample should look similar to that of the population, a new population with the sample size, defined as bootstrap population, can be used to find the probability distribution of sampling average. When creating the bootstrap population, each unit in the sample represent N/n units in the new population with the same value, where N and n represent the sizes of the population and the sample, and we round off the number to the nearest integer.
  - Once we create the bootstrap population, we repeat selecting a simple random sample of size n from our bootstrap population, defined as bootstrap sample and calculating its average for k times. Then, by plotting a histogram of the k sample averages from the k iteration, we can evaluate the probability distribution of the sampling average.

Discussion & Analysis:

Since the process of designing computer labs is comparable to designing video games, we can conclude a lot from analysis of the data that generated from the survey that ask students about their video games status.

From Scenario 1, we can see that the sample mean is around 37%, and the confidence interval with finite population correction factor is (0.288, 0.460). We can observe that around 37% of students playing video games prior to the survey, thus if the statistic labs are similar to play video game, we can anticipate that there will be less than half of the students who are not in favor of such labs, thus faculties need to pay much attention to the lab attendance.

From Scenario 2, we can see that for students who play video game more frequently than others, their average playing time are also much longer, which indicates that if a student like to play video games, they will highly likely to go to the statistics lab and stay a longer time in the lab as well. We also discover that students who play daily and weekly are also the ones who will play a lot during the busy time, such as having exams afterwards compared to the students who do not play or play monthly. There is a positive linear relationship between students' playing frequency and students' busyness. In this sense, since there are only half of students like to go to labs, there would be much less students who would like to go to labs before having exams. Thus, in order to attract students to go to labs, faculties should place labs time aside from the exam times.

From Scenario 3, at first, we recognize that a lot of students only play video games for less than three hours. Thus, in order to design the length for the statistic labs, we make a confidence interval on the students' playing time. From three confidence intervals, by Central Limit Theorem, with finite sample correction, and by bootstrap resampling technique, we choose the confidence interval with finite sample correction, the interval is (0.576, 1.910). The peek among the bootstrapped sample means is around 1.0 hours to 1.5 hours. Therefore, we conclude that the best length for the statistic lab should not be more than 2 hours and should be no less than 1 hour. Students will focus more if the lab length is between 1 hour and 2 hours. Thus, faculty should accommodate the lab time according to this.

From Scenario 4, we observe that 66% of students like to play video games for relaxation, much higher than any other reasons, which indicating that faculties should design the lab to be relatively easy in order to attract students. The most frequent reason for disliking video games is that it could take a lot of time, thus similarly, if the statistic labs are going to take a very long time, students have a high chance to complain about it. Thus, the faculties should make sure the labs are doable and not lasting for too long.

From Scenario 5, we want to see if there is a difference with liking playing video games between genders, work status, and ownership of a PC. After doing cross-tabulation and chi-square test, we find that there is no difference between genders, work status, and ownership of a PC in liking playing video games. We can eliminate the worry about individual difference when designing the statistic labs.

From Scenario 6, we observe that if nonrepondents were failing students who no longer go to discussion section, the current proportion of expected grade for A, B, C will decrease, and the current proportion of expected grade for C and F will rise, the actual distribution of the grades will be similar to the target grades

From the additional hypothesis, we want to see whether there is relationship between the frequency students play with their expected grade. After comparing with the plot and using t-test to test the mean grade for two groups, we observe that there is a difference between the expected grades from students who play frequently and students who do not play frequently. From the plot, we can see that students who play frequently tend to expect a higher grade than the others. This might due to the fact that students who play a lot are the ones that knows and understand a lot about the course materials and thus have time to play.

To conclude, there is a data limitation in our survey data: our survey data is too small (only have 95 out of 314 students), if we can gather more data, we can eliminate the process of bootstrapping, and thus get a more precise analysis from the true population. However, based on our current data, we can conclude that there would be less than half students would like to go to labs, and would be a little bit more during the busy time. The lab length should be around 1.5 hours and should be set aside form other exam times. The lab should be relatively easy for students in order to attract their attendance. And if faculties could develop labs based on above standard, we can assume that it could be generalized to all students since there is no individual difference that need to be considered.

Work Cited:

Begoña Gros (2007) Digital Games in Education, Journal of Research on Technology in Education, 40:1, 23-38, DOI: 10.1080/15391523.2007.10782494

Ogletree, S. M., & Drake, R. (2007). College Students' Video Game Participation and Perceptions: Gender Differences and Implications. *Sex Roles,56*(7-8), 537-542. doi:10.1007/s11199-007-9193-5

Ventura, M., Shute, V., & Kim, Y. J. (2012). Video gameplay, personality and academic performance. *Computers & Education,58*(4), 1260-1266. doi:10.1016/j.compedu.2011.11.022