Who Plays Video Games

May 3, 2020

Cameron Thomas
Chase Oden
Duncan Carlmark

I. Background

The methodology in which the data was acquired was through a random survey in an introductory statistics class at UC Berkeley. Each student selected in the survey was chosen randomly through a pseudo random choice system with each student's responses kept anonymous in order to encourage honest responses. Non-responses were kept to a minimum by acquiring the survey answers during a class discussion.

In various studies, it was found that people enjoyed playing video games for a variety of reasons. When playing video games, people will change their expectations on their own performance and how they're interpreting the game to maximize their enjoyment of the video game[1]. Video games also allow people to expand their feelings of autonomy and desire for fulfillment[2], leading them to enjoy video games even further. In another study, it was also made apparent that people that played video games were likely more capable in academic scopes, possessing more advanced problem solving skills than their peers that didn't play video games[3].

II. Introduction

Every year, thousands of students enroll in statistics courses at UC Berkeley. To aid the instruction of these students, a committee designed a series of computer labs, with the goal of extending traditional teaching methods by providing an interactive learning environment. Because the characteristics of labs have been linked to parallel characteristics in video games, the committee conducted a survey to better understand which aspects of video games students find the most and least fun. The objective of this study is to investigate the responses of the participants in the study with the intention of providing useful information about the students to the designers of the new computer lab.

Scenarios to be addressed in our data analysis:

1. "Begin by providing an estimate for the fraction of students who played a video game in the week prior to the survey. Provide an interval estimate as well as a point estimate for this proportion."
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). How might the fact that there was an exam in the week prior to the survey affect your previous estimates and this comparison?"

3. “Consider making an interval estimate for the average amount of time spent playing video games in the week prior to the survey.”
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?”
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.”
6. “Just for fun, further investigate the grade that students expect in the course. How 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?”

III. Data

The data we worked with is sourced from a survey of a lower division statistics at UC Berkeley. There were 314 available students and 95 were picked randomly to participate in the survey that produced the data, with 91 fully completing the survey. The survey consisted of multiple sections that went over how much people liked to play video games, what they liked about them, and what they disliked about them. Most of the questions were numerical answers (hours played in the past week) while some had ordinal answers (how much do you like to play video games). Data that was missing or unanswered was left as 99 and was later converted to NA values in our R code. The question formats of the survey can be found in the appendix in both Table (2) and Table (3) .

IV. Theory

Survey Sampling

A survey follows scientific methodology to collect data from individuals, typically sampled from a large population, in hopes of describing, exploring, and/or explaining characteristics of the population as a whole. Good survey research is quantitative, careful, replicable, impartial, representative, and theory-based. In general, individuals must have an equal chance of being selected in the sample and the sample must be generalized to the total population of interest.

Confidence Intervals

From the Central Limit Theorem, we know that $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ approximately follows a normal standard distribution, which can be used to construct confidence intervals for the population average, μ, when n is not too small/large compared to the population size, N. The 95% confidence interval is $(\overline{x} - 2\frac{\sigma}{\sqrt{n}}, \overline{x} + 2\frac{\sigma}{\sqrt{n}})$, which means the probability that $\overline{x}$ is within two standard errors of μ is about 95%.

Bootstrap

Bootstrapping is a method used on finite samples to improve accuracy among estimations. This is done by iteratively resampling a dataset with replacement. A sample size, n, is taken from a dataset and resampled until a variety of combinations are created with the original dataset properties. Statistical calculations, including mean, median, and standard deviation, are done on each sample. The “bootstrapper” can choose any k iterations where k samples will be created. Then, a sampling distribution can be used to visualize each sample mean and calculate a new overall mean as well.

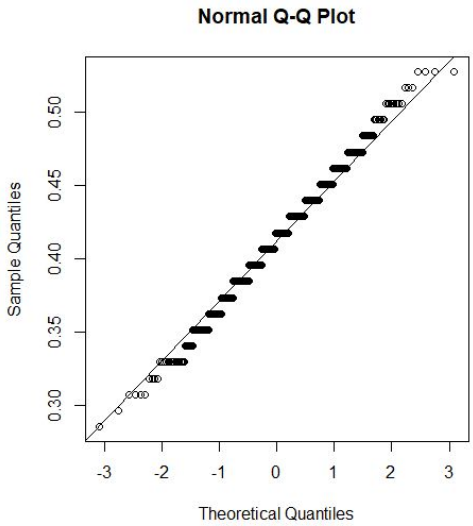
V. Analysis

Scenario 1

```
> print(freq.percentage)
[1] 0.4065934
> print(interval)
[1] 0.3185468 0.4946400
```

Figure (A): Mean time spent playing video games the week before a test between all gaming frequency groups.

To understand how many students had played a video game prior to their exam we first examined the frequency at which students played games. To simplify our analysis we decided that students who stated that they played video games daily or weekly as”successes” meaning they were encoded with a value of one and everyone else a value of 0. The observed value of this proportion from our original sample was .406. In order to determine a confidence interval for our statistic, we used a bootstrap to approximate the actual distribution. After resampling 500 times from our bootstrap, we had created a distribution of sample proportions and used the information from this to create our confidence interval. From running a Kolmogorov-Smirnov test, we determined that the distribution of our bootstrap was approximately normal and applied the standard formula for a confidence interval. In the end we derived an interval (.318, .494) for our observed statistic .046.



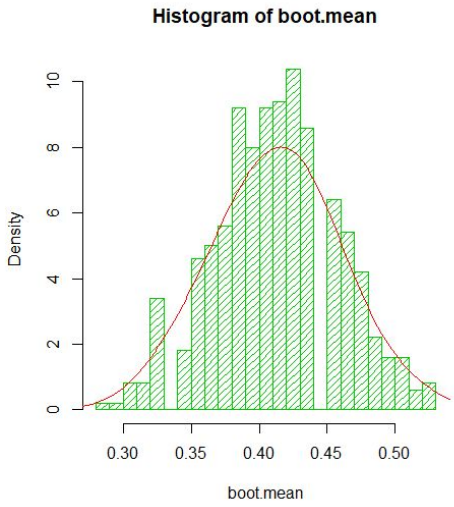


Figure (B): Normal Q-Q Plot and Histogram of Bootstrapped Means

Scenario 2

```
> mean(data$time[data$freq == 1], na.rm=TRUE)
[1] 4.444444
> mean(data$time[data$freq == 2], na.rm=TRUE)
[1] 2.539286
> mean(data$time[data$freq == 3], na.rm=TRUE)
[1] 0.05555556
> mean(data$time[data$freq == 4], na.rm=TRUE)
[1] 0.04347826
```

Figure (C): Mean time spent playing video games the week before a test between all gaming frequency groups.

In looking at the mean time that people played video games the past week it seems that the groups that reported playing video games daily and weekly had a very high amount of time played compared to the other groups. Even though the time played is high, looking at the daily video game players we can see that they only averaged around 4.5 hours for the week prior to the test. Conversely, we can see that the people that reported only playing once a week still managed a solid 2.5 hours, which in the context of an entire week could possibly still be a normal amount of play time. Both the remaining groups of monthly and semesterly video game players had a very minimal amount of hours the previous weeks, which corresponds to their group selection. Overall, the amount of hours played among the groups correspond with their group choices, although both the daily and weekly groups had a significant margin over the monthly and semesterly groups. We believe that the test did have an impact on the answers that were given, as the daily and weekly groups played less than or equal to the same amount of hours they attested to playing in normal circumstances.

Scenario 3

To determine an appropriate interval that gives us insight into the range of time people spent playing video games during the week before an exam, we utilized a process very similar to scenario 1 with some slight deviations. We did not limit the calculation of a mean to those who were daily or weekly gamers since the “Like” column is a direct measure of the hours played during the week prior to the exam. To compute a CI we bootstrapped based on the sample and calculated a distribution of sample means for the “Time” feature. This gave us a confidence interval of (.611, 1.87) hours on average.

In our analysis we also determined that the sample distribution was roughly Poisson shaped. This led us to conclude that calculating a confidence interval using the bootstrap method was viable.

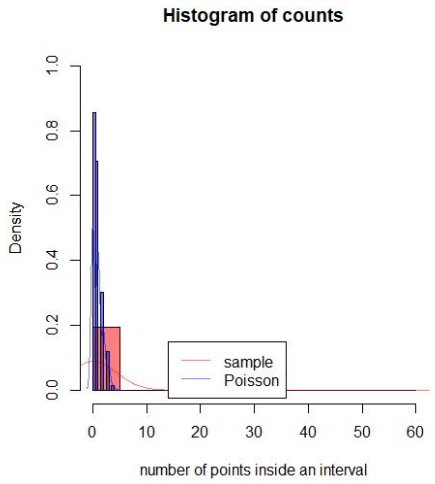


Figure: Counts Histogram

Scenario 4

```
> # Of people that play games, these people dislike them for external reasons
> # A large percentage of people are in this category
> out_of_cont_dislikes
     time      cost   friends
48.780488 39.024390  2.439024
> # of people that play games, these people dislike them for game-related reasons
> # A smaller amount of people are in these categories
> in_cont_dislikes
    frust    lonely     rules    boring     point
25.609756  4.878049 18.292683 17.073171 34.146341
```

Figure (D): Percentages of people who disliked games for various reasons.

In determining if students liked video games based on the given survey, we analyzed the attitude based questions on video games from the survey given to the students. The first piece of evidence that points to people liking video games is actually the information derived from what people most *dislike* about video games. In looking at Figure D, we grouped certain reasons for disliking video games into different groups. To account only for people that played video games, we looked at which types of games people preferred and removed those that didn't select any of the genre options. The first group out_of_cont_dislikes contains the percentages of people that said they didn't like video games for reasons out of their control, such as games taking up too much time, costing too much, or not having a group to play it with. Because the percentage of people in this group is so large, we believe that people do have an affinity for video games, but are restricted because of things that aren't related to the action of playing video games itself. Looking at the lower statistics in Figure D, there are the percentages of people that don't like playing video games for reasons related to video games themselves. All of these percentages are lower than the two dominating out of control reasons, which shows that people are in the majority disliking video games because of outside restrictions and not because of the video game itself.

```
> gamers_data <- data[which((data$like != 1) & !(is.na(data$like))),]
> sum(gamers_data$like <= 3)
[1] 69
> sum(gamers_data$like > 3)
[1] 20
```

Figure (E): Counts of people that like and don't like video games.

In the dataset, there are people that say they don't play video games at all, so by dropping those people from the dataset we can see of those that actually play video games how much they like playing video games. From Figure E, we can see that of the 91 people sampled, 89 people play video games. Of those 89, ~77% of them say they enjoy video games "somewhat" or "very much". This is a large majority of those that play video games, and so leads us to conclude that if a person from this sample is playing video games they most likely also enjoy playing those games.

Scenario 5

We will use visual representations of some of the comparisons we have been discussing throughout this report. This will help us better catalog different groups depending on how the students responded in the survey.

The *Like* variable of *videoData.txt* will be compared against *Sex*, *Work*, and *Own*. Before making these comparisons, we should collapse the range of responses. The *Like* variable is numerically coded 1-5. The *1 = Never Played* data is not useful in this scenario so we remove it and then code 2 and 3 as "Like" and 4 and 5 as "Dislike". Now we are ready to make distinct comparisons.

Comparison 1

Below in Figure F, we have a cross-tabulation between male/female students who disliked/liked video games. As we can see, 77.5% of students from both genders like video games while only 22.5% dislike them. Even though there are more males than females, we can observe that males are more likely to like video games. 68.4% of females like video games while 84.3% of males like video games. This can be due to society stereotyping the typical gamer as male. This stigma makes it difficult for females to adopt this male-dominated activity.

Cell Contents

	N
Chi-square contribution	
	N / Row Total
	N / Col Total
	N / Table Total

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

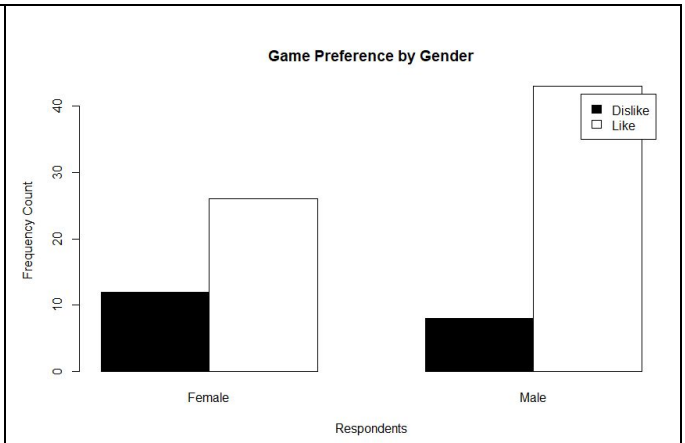


Figure (F): Cross-tabulation and Barplot of Like and Sex

We conducted a chi-squared test to see if there is an association amongst the variables. You can view the results in Figure K in the Appendix. The null hypothesis is that gender is independent of responses in the survey. The alternative hypothesis is otherwise. Since the p-value is 0.1285 > 0.05, we fail to reject the null hypothesis. There is not enough evidence to show *Like* is associated with *Sex*.

Comparison 2

Work can influence a person's perspective on video games. Despite both work proportions being almost equal, we can see a lower percentage of students who work dislike video games in comparison to 31.8% of those who do not work. The boxplot in Figure H also supports Figure G and the claim that those who work more hours are less likely to dislike video games. This may be due to employees viewing video games as stress-relief. Regardless of employment status, there is a general satisfaction toward playing video games.

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

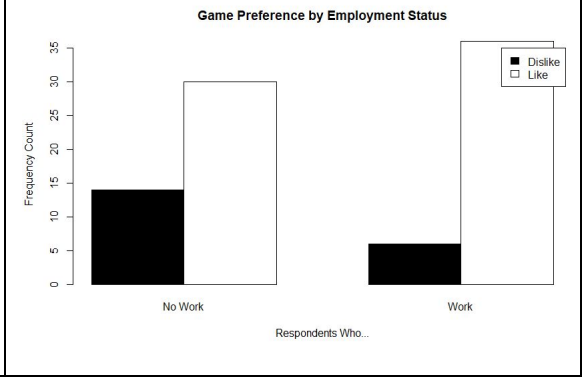


Figure (G): Cross-tabulation and Barplot of Like and Work

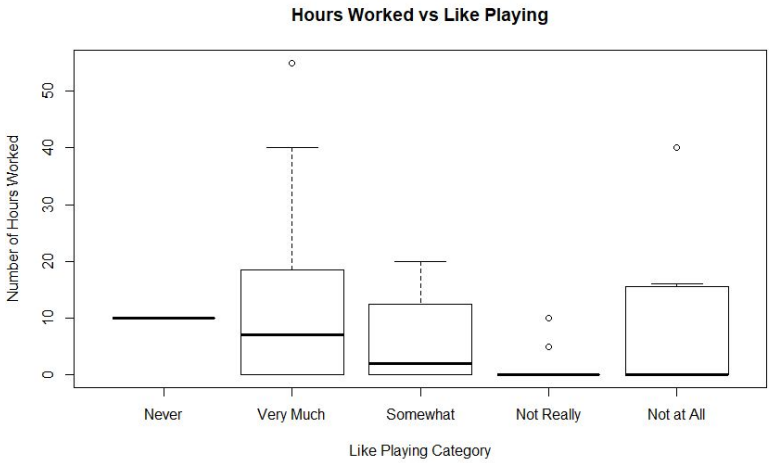


Figure (H): Boxplot of Hours Worked and Enjoyment Level

To see test independence between students who work and do not work, we conducted a chi-squared test. The null hypothesis is employment status is independent of responses in the survey. The alternative hypothesis is otherwise. Since p-value 0.09522 > 0.05, we fail to reject the null hypothesis. There is not enough evidence to show *Like* is associated with *Work*.

<u>Comparison 3</u>

Lastly, we will compare *Like* and *Own*. In Figure I, there are more students who own a computer than those who don't. From the proportions, however, we can see that owning a computer or not will greatly affect if a person likes or dislikes video games. This may be because people can play video games on a multitude of platforms, including in the arcade, on a console, or on a computer.

	data.clean$own		
data.clean$like	No PC	Own PC	Row Total
Dislike	3	17	20
	1.062	0.392	
	0.150	0.850	0.225
	0.125	0.262	
	0.034	0.191	
Like	21	48	69
	0.308	0.114	
	0.304	0.696	0.775
	0.875	0.738	
	0.236	0.539	
Column Total	24	65	89
	0.270	0.730	

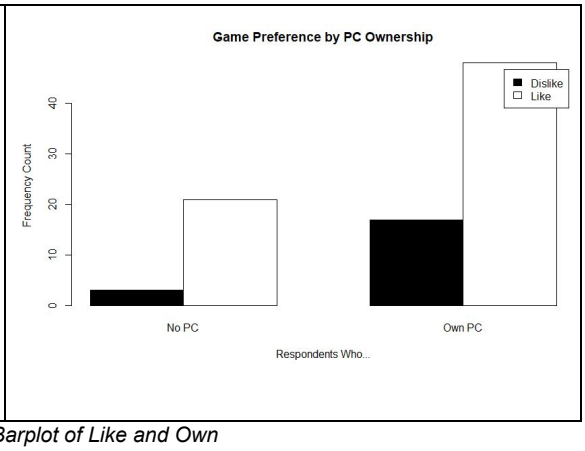


Figure (I): Cross-tabulation and Barplot of Like and Own

The null hypothesis is computer ownership is independent of responses in the survey. The alternative hypothesis is otherwise. Since p-value 0.02786 < 0.05, we fail to reject the null hypothesis. There is not enough evidence to show *Like* is associated with *Own*.

Scenario 6

We will inspect if the percentage of each expected grade from the sample follows the distribution used in grade assignment. Please note that Figure J is based on counts while Table 1 is based on percentages.

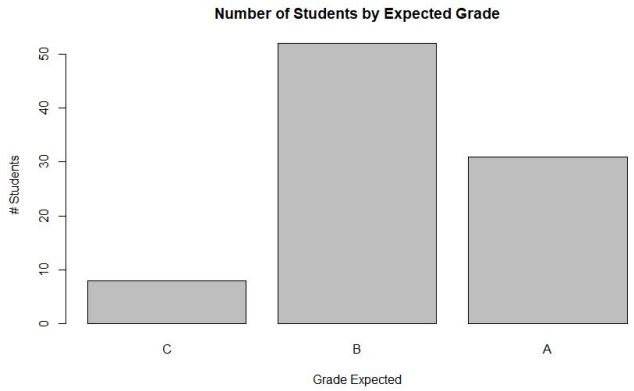


Figure (J): Grade Expectation

From Table 1, the percentage of students who expect to earn A and B are both higher than the target levels, while the percentages of students who expect to earn C and D are both significantly lower than target levels. Thus, students' grade expectations generally do not match the target distribution of 20% A's, 30% B's, 40% C's, and 10% D's or lower. In general, students are expecting a higher grade than they will get. The expected percentage is dependent on the number of students who complete the survey. If nonrespondents were failing students who no longer bothered to come to discussion, the distribution of students' grade expectation would have been more close to the target grade distribution. If these unresponsive failing students are added to the sample space, then the picture will change because the expected percentage of students earning A and B will be lower and the expected percentage of students earning C and D will be higher.

	A	**B**	**C**	**D**
Target	20%	30%	40%	10%
Expected	34.07%	57.14%	8.79%	0%

Table (1): Grade Distribution

VI. Conclusion

Throughout our analysis, we developed many key insights into the role of video games in students' lives and how this information might relate to those looking to integrate games with educational programs. We found that a sizable segment of the students, around 46%, played video games either weekly or daily and spent a decent amount of their time playing video games before an exam. There are many students who

also play video games less frequently, every month or semester, but these students spent significantly less time playing and only averaged around 30 minutes of playtime the week before their exam. Also, we discovered that many of the issues people had with video games were how much the games cost or how much time that it takes to play the games. These are not core to the experience of video games and lead us to conclude that people like video games, but they may not enjoy the conditions of playing video games. Another significant portion of our study was the analysis of correlations to see if the “Like” column was correlated with either “Sex”, “Work”, or “Own[ership]”. However, after conducting our tests we determined that there was not enough evidence to conclude that “Like” was correlated with any of these variables. Our portion of analysis was a slight tangent from our main purpose, but we found that students tended to overestimate the grade that they expected to receive.

Our conclusions support the development of a computer lab that provides a video game as an educational resource. A significant portion of students play video games and spend a significant amount of time playing them, so a game would resonate well with them. Also, even though there are many students that do not play video games frequently these students have no opposition to videogames themselves and instead are opposed to things surrounding video games like purchasing and finding friends to play with. An instructional game would have none of these issues as it would be provided as part of the course material and would not require that students find friends to play with as it is a solo educational experience.

VII. Contributions

Cameron: Scenario 5 and 6. Introduction. Theory.
Chase: Background, Data, Scenarios 2, 4.
Duncan: Scenario 1 and 3. Conclusion

VIII. Works Cited

[1]: Klimmt C., Blake C., Hefner D., Vorderer P., Roth C. (2009) Player Performance, Satisfaction, and Video Game Enjoyment. In: Natkin S., Dupire J. (eds) Entertainment Computing – ICEC 2009. ICEC 2009. Lecture Notes in Computer Science, vol 5709. Springer, Berlin, Heidelberg

[2]: Karyn Riddle, Swee Kiat Tay & Jiaxi Wu (2019) “It Lets Me Fight the Bad Guys”: An Exploration Into the Factors Predicting Enjoyment of Violent Video Games, Communication Studies, 70:1, 36-58, DOI: 10.1080/10510974.2018.1438490

[3]: Adachi, P.J.C., Willoughby, T. More Than Just Fun and Games: The Longitudinal Relationships Between Strategic Video Games, Self-Reported Problem Solving Skills, and Academic Grades. *J Youth Adolescence* 42**,** 1041–1052 (2013). https://doi.org/10.1007/s10964-013-9913-9

IX. **Appendix**

Variable	Description	Type of Data
time	# of hours played in the week prior to survey	numerical
like	How much the respondent likes playing video games 1 = never played, 2 = very much, 3 = somewhat, 4 = not really, 5 = not at all	categorical (ordinal)
where	Where the respondent usually plays video games 1 = arcade, 2 = home system, 3 = home computer, 4 = arcade and either home computer or 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 (ordinal)
busy	Whether the respondent plays video games when busy 1 = yes, 0 = no	categorical
educ	Whether the respondent believes that playing video games is 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 1 = yes, 0 = no	categorical
math	Whether the respondent hates math 1 = yes, 0 = no	categorical
work	# of hours worked the week prior to the survey	numerical
own	Whether the respondent owns a PC	categorical

	1 = yes, 0 = no	
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 (ordinal)

Table (2): Data Dictionary for videodata.txt.

Variable	Description	Type of Data
action	Does the respondent like action-genre games 1 = yes, 0 = no	categorical
adv	Does the respondent like adventure-genre games 1 = yes, 0 = no	categorical
sim	Does the respondent like simulation-genre games 1 = yes, 0 = no	categorical
sport	Does the respondent like sports-genre games 1 = yes, 0 = no	categorical
strategy	Does the respondent like strategy-genre games 1 = yes, 0 = no	categorical
graphics	Does the respondent play video games for the graphics/ realism 1 = yes, 0 = no	categorical
relax	Does the respondent play video games to relax 1 = yes, 0 = no	categorical
coord	Does the respondent play video games for their hand/ eye coordination 1 = yes, 0 = no	categorical
challenge	Does the respondent play video games for the challenge 1 = yes, 0 = no	categorical

master	Does the respondent play video games for a feeling of mastery 1 = yes, 0 = no	categorical
bored	Does the respondent play video games because they're bored 1 = yes, 0 = no	categorical
other	Other reason for liking video games	
time	Does the respondent dislike video games because it takes too much time 1 = yes, 0 = no	categorical
frust	Does the respondent dislike video games because they're frustrating 1 = yes, 0 = no	categorical
lonely	Does the respondent dislike video games because it's lonely 1 = yes, 0 = no	categorical
rules	Does the respondent dislike video games because there's too many rules 1 = yes, 0 = no	categorical
cost	Does the respondent dislike video games because it costs too much money 1 = yes, 0 = no	categorical
boring	Does the respondent dislike video games because it's too boring 1 = yes, 0 = no	categorical
friends	Does the respondent dislike video games because their friends don't play 1 = yes, 0 = no	categorical
point	Does the respondent dislike video games because they don't see a point in it 1 = yes, 0 = no	categorical
other2	Other reason for disliking video games	

Table (3): Data Dictionary for videoMultiple.txt.

Comparison 1 (Like vs Sex)

```
> chisq.test(table(data.clean$like, data.clean$sex))

        Pearson's Chi-squared test with Yates' continuity correction

data:  table(data.clean$like, data.clean$sex)
X-squared = 2.3106, df = 1, p-value = 0.1285
```

Comparison 2 (Like vs Work)

```
> chisq.test(table(data.clean$like, data.clean$work))

        Pearson's Chi-squared test with Yates' continuity correction

data:  table(data.clean$like, data.clean$work)
X-squared = 2.7838, df = 1, p-value = 0.09522
```

Comparison 3 (Like vs Own)

```
> chisq.test(table(data.clean$like, data.clean$own))

        Pearson's Chi-squared test with Yates' continuity correction

data:  table(data.clean$like, data.clean$own)
X-squared = 1.1738, df = 1, p-value = 0.2786
```

Figure (K): Chi-Squared Test Results for Scenario 5