Project 2: Who Plays Video Games?

by Michelle Barnes

MTH332

Abstract

Introduction

Faculty and graduate students of the Statistics Department at University of California, Berkeley formed a committee to design a series of computer labs to aid the students of an introductory statistics course. In designing these labs, the committee used information gathered from a survey given to undergraduate students in lower-level stats courses. The survey looked to determine how much and how often students played video games, and what they liked and disliked most about playing them. Students were randomly selected, and of the 314 students in the Stats 2, Section 1 class of Fall 1994, 95 were chosen. Of those, 91 responded to the survey. This paper is to report on the playing habits of the respondents, and from this sample extrapolate to estimates of how these trends might appear in the whole population of 314 students. The following is an examination of the data using the programming language R in the RStudio IDE.

Methods

The data was downloaded into a data frame called "survey" and attached, making it easier to slice and reference variables. Then some variables relating to the population and the sample totals were assigned. The correction factor for the population was also assigned, calculated as $\frac{\sqrt{N-n}}{\sqrt{N-1}}$

```r
survey <- read.table("~/Classes/Spring2020/MTH332/Project_2/mth332_proj2-2020/videogamessurvey.txt", hea
attach(survey)

N <- 314 # Total population
n <- 91 # Sample population
cf <- sqrt(N-n)/sqrt(N-1) # correction factor
```

Fraction of students who played video games in the week before survey   The sample size is n = 91 survey respondents. To find out how many played, the nonzero values from the time column need to be extracted.

```r
timePlay <- time # extracts the time column as a vector
play_yn <- ifelse(timePlay!= 0, 1, 0) # to determine if someone played or not, change nonzero values to
# Now let's see how many of each variable we have
table(play_yn)
```

```
## play_yn
##  0  1
## 57 34
```

With 1 representing those who did play, we see that 34 of the 91 respondents played a video game in the week before the survey. The sample proportion is then calculated as follows:

```
xbar_didPlay = mean(play_yn)
xbar_didPlay
```

```
## [1] 0.3736264
```

And the confidence interval is:

```
sd_didPlay = sd(play_yn)*cf

ciInt = (2*sd_didPlay)/sqrt(n)

confidenceIntPlay <- xbar_didPlay + c(-ciInt, ciInt)
confidenceIntPlay
```

```
## [1] 0.2875420 0.4597108
```

This is a 95% confidence interval, and is calculated as $(\bar{x} - 2s/\sqrt{n}, \bar{x} + 2s/\sqrt{n})$, with s being the sample standard deviation.

Time played in week prior to survey   The mean amount of time all survey respondents spent playing video games is found:

```
xbar_playTime <- mean(timePlay)
xbar_playTime
```

```
## [1] 1.242857
```

And the confidence interval:

```
sd_playTime <- sd(timePlay)*cf
ciInt2 = (2*sd_playTime)/sqrt(n)

confidenceIntPlayTime <- xbar_playTime + c(-ciInt2, ciInt2)
confidenceIntPlayTime
```

```
## [1] 0.5744499 1.9112644
```

Frequency of Play   Now for computing the frequency of play, the variable "freq" must be extracted from survey and examined. Values of 99 indicate blank or incorrect responses, and so are removed from the data.

```
freqPlay <- subset(survey, freq != 99, select = freq)
```

The categories for frequency are: 1=Daily, 2=Weekly, 3=Monthly, 4=Semesterly.

```r
table(freqPlay)
```

```
## freqPlay
##  1  2  3  4
##  9 28 18 23
```

Sorting the categories out for finer analysis:

```r
# Hours played in prior week and general play frequency
freqPlayVal <- subset(survey, freq != 99, select = c(freq,time))

dailyVal <- subset(freqPlayVal, freq == 1, select = time)
weeklyVal <- subset(freqPlayVal, freq == 2, select = time)
monthlyVal <- subset(freqPlayVal, freq == 3, select = time)
semesterVal <- subset(freqPlayVal, freq == 4, select = time)

# Now calculating the mean for each frequency group

dm <- mean(dailyVal$time); dm
```

```
## [1] 4.444444
```

```r
wm <- mean(weeklyVal$time); wm
```

```
## [1] 2.539286
```

```r
mm <- mean(monthlyVal$time); mm
```

```
## [1] 0.05555556
```

```r
sm <- mean(semesterVal$time); sm
```

```
## [1] 0.04347826
```

Calculating the confidence intervals for each frequency group:

```r
# Daily
sd_daily = sd(dailyVal$time)*cf

ciInt3 = (2*sd_daily)/sqrt(n)

confidenceIntDaily <- dm + c(-ciInt3, ciInt3); confidenceIntDaily
```

```
## [1] 3.458699 5.430190
```

```r
# Weekly
sd_weekly = sd(weeklyVal$time)*cf

ciInt4 = (2*sd_weekly)/sqrt(n)

confidenceIntWeekly <- wm + c(-ciInt4, ciInt4); confidenceIntWeekly
```

```
## [1] 1.566142 3.512429
```

```
# Monthly
sd_monthly = sd(monthlyVal$time)*cf

ciInt5 = (2*sd_monthly)/sqrt(n)

confidenceIntMonthly <- mm + c(-ciInt5, ciInt5); confidenceIntMonthly
```

```
## [1] 0.02694187 0.08416924
```

```
# Semesterly
sd_semester = sd(semesterVal$time)*cf

ciInt6 = (2*sd_semester)/sqrt(n)

confidenceIntSemester <- sm + c(-ciInt6, ciInt6); confidenceIntSemester
```

```
## [1] 0.006578322 0.080378200
```

Gender Breakdowns   This section will examine the data of time spent playing in the week prior to the survey, frequency of play, and how much a respondent likes or dislikes gaming relative to gender.
First we divide the survey data into 2 new data frames by gender. Female is coded as 0, male is coded as 1.

```
femaleResp <- subset(survey, sex == 0, select = c(time, like, freq))

maleResp <- subset(survey, sex == 1, select = c(time, like, freq))

# taking a look at the gender proportions of the respondents
nrow(femaleResp)/n
```

```
## [1] 0.4175824
```

```
nrow(maleResp)/n
```

```
## [1] 0.5824176
```

Examining the data on female respondents:

```
# time spent playing in the prior week
play_ynF <- ifelse(femaleResp$time != 0, 1, 0) # to determine if someone played or not, change nonzero

table(play_ynF)
```

```
## play_ynF
##  0  1
## 29  9
```

4

```r
mean(play_ynF) # proportion of female respondents who played
```

```
## [1] 0.2368421
```

```r
fem_timePlay <- femaleResp$time
summary(fem_timePlay)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    0.00    0.00    0.75    0.00   14.00
```

```r
# extracting frequency of play
fem_freqPlay <- subset(femaleResp, freq != 99, select = freq)

# categorizing frequencies
table(fem_freqPlay)
```

```
## fem_freqPlay
##  1  2  3  4
##  3  6  8 14
```

```r
# find hours played in prior week relative to frequency
fem_freqPlayVal <- subset(femaleResp, freq != 99, select = c(freq,time))

dailyValF <- subset(fem_freqPlayVal, freq == 1, select = time)
weeklyValF <- subset(fem_freqPlayVal, freq == 2, select = time)
monthlyValF <- subset(fem_freqPlayVal, freq == 3, select = time)
semesterValF <- subset(fem_freqPlayVal, freq == 4, select = time)
```

The subgroups here are rather small, so the 'summary' function will provide a better picture than the mean alone.

```r
# it is also worth seeing how many respondents in each group did not play (time = 0)
summary(dailyValF); table(dailyValF$time == 0)
```

```
##       time
##  Min.   : 0.000
##  1st Qu.: 0.000
##  Median : 0.000
##  Mean   : 4.667
##  3rd Qu.: 7.000
##  Max.   :14.000
```

```
##
## FALSE  TRUE
##     1     2
```

```r
summary(weeklyValF); table(weeklyValF$time == 0)
```

```
##        time
##  Min.    :0.000
##  1st Qu.:1.125
##  Median :1.750
##  Mean    :2.083
##  3rd Qu.:2.750
##  Max.    :5.000
```

```
##
## FALSE   TRUE
##     5      1
```

```r
summary(monthlyValF); table(monthlyValF$time == 0)
```

```
##        time
##  Min.    :0.000
##  1st Qu.:0.000
##  Median :0.000
##  Mean    :0.125
##  3rd Qu.:0.125
##  Max.    :0.500
```

```
##
## FALSE   TRUE
##     2      6
```

```r
summary(semesterValF); table(semesterValF$time == 0)
```

```
##        time
##  Min.    :0.00000
##  1st Qu.:0.00000
##  Median :0.00000
##  Mean    :0.07143
##  3rd Qu.:0.00000
##  Max.    :1.00000
```

```
##
## FALSE   TRUE
##     1     13
```

The scoring for how much a respondent likes video games is as follows: 1=Never played, 2=Very much, 3=Somewhat, 4=Not really, 5=Not at all

```r
table(femaleResp$like)
```

```
##
##  2  3  4  5
##  5 21  8  4
```

```r
mean(femaleResp$like) # the mean here will give us the average "attitude" regarding video games
```

```
## [1] 3.289474
```

Now examining the data on male respondents:

```r
# time spent playing in the prior week
play_ynM <- ifelse(maleResp$time != 0, 1, 0) # to determine if someone played or not, change nonzero va

table(play_ynM)
```

```
## play_ynM
##  0  1
## 28 25
```

```r
mean(play_ynM) # proportion of male respondents who played
```

```
## [1] 0.4716981
```

```r
male_timePlay <- maleResp$time
summary(male_timePlay)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   0.000   1.596   2.000  30.000
```

```r
# extracting frequency of play
male_freqPlay <- subset(maleResp, freq != 99, select = freq)

# categorizing frequencies
table(male_freqPlay)
```

```
## male_freqPlay
##  1  2  3  4
##  6 22 10  9
```

```r
# find hours played in prior week relative to frequency
male_freqPlayVal <- subset(maleResp, freq != 99, select = c(freq,time))

dailyValM <- subset(male_freqPlayVal, freq == 1, select = time)
weeklyValM <- subset(male_freqPlayVal, freq == 2, select = time)
monthlyValM <- subset(male_freqPlayVal, freq == 3, select = time)
semesterValM <- subset(male_freqPlayVal, freq == 4, select = time)
```

The subgroups here are rather small, so the 'summary' function will provide a better picture than the mean alone.

```r
# it is also worth seeing how many respondents in each group did not play (time = 0)
summary(dailyValM); table(dailyValM$time == 0)
```

```
##       time
##  Min.   : 1.000
##  1st Qu.: 2.000
##  Median : 2.500
##  Mean   : 4.333
##  3rd Qu.: 3.750
##  Max.   :14.000


##
## FALSE
##     6
```

```r
summary(weeklyValM); table(weeklyValM$time == 0)
```

```
##       time
##  Min.   : 0.000
##  1st Qu.: 0.500
##  Median : 2.000
##  Mean   : 2.664
##  3rd Qu.: 2.000
##  Max.   :30.000


##
## FALSE   TRUE
##    19      3
```

```r
# note that all the male respondents indicated that they played weekly or daily, no less than that
```

The scoring for how much a respondent likes video games is as follows: 1=Never played, 2=Very much, 3=Somewhat, 4=Not really, 5=Not at all

```r
# There are some entries indicating blank or incorrect responses, so the data needs a little cleaning
maleRespLike <- subset(maleResp, freq != 99 | like != 99, select = like)

table(maleRespLike)
```

```
## maleRespLike
##  1  2  3  4  5
##  1 18 25  5  3
```

```r
mean(maleRespLike$like) # the mean here will give us the average "attitude" regarding video games
```

```
## [1] 2.826923
```

Results