Megan Cusey

Quantitative Methods Course Project

Spring 2019

Exploratory Analysis of Video Game Data in Central Illinois

The objective of this study is to answer some important questions regarding Video Gaming in Central Illinois. The following information was compiled based on data extracted from the Illinois Gaming Board’s website for years 2012-2019 for Auburn, Springfield, and Sangamon County Municipalities.

The below summarizes important information needed to understand the information.

* The Illinois Gaming Board separates Auburn, Springfield, and Sangamon County into their own individual Municipalities for data reporting purposes even though Auburn and Springfield are both geographically located in Sangamon County.
* For this study’s purpose, I focused on the Net Wager data as it relates to other variables such as season, number of video game terminals, and fiscal years.
  + The Net Wagering amount is the difference between Amount Played and Amount Won per location.
  + Amount Played is the amount of money bet in the video game terminals per location.
  + Amount Won is the amount of money won per location.

The goal of this analysis is to address the following questions:

* How does the number of video game terminals effect net wagers per location? Is there a significant difference in the mean of net wagers for establishments with different numbers of terminals?
* Since video gaming has become legal in the State of Illinois in 2012, there’s no doubt that the number of establishments that provide video game terminals have increased. Does the net wager amount have a significant difference between 2012 and 2019?
* A friend has offered an alternative hypothesis that the mean of net wagers decreases in winter months. Is there a significant difference between the net wagers recorded in winter as opposed to other seasons?

# Data Clean Up

The data recorded is relatively clean and consistent throughout the years. There are a few problems I addressed:

* 12 rows returned the number of video game terminals in a location reported was 0. This does not make sense with the data, so I removed the 12 rows.
* 5 rows returned number of video game terminals in a location reported at 1. 4/5 did not have any Net Wager dollars. I decided to delete all 5 rows since one observation with a single video game terminal does not add value to my analysis.
* 5 rows with a VGTCount but no Net Wager data.
* Removed year 2012 and 2019 due to impartial data.

# Assumptions

Per Central Limit Theorem, as the scores are measured for the entire data set, the sampling distribution of those scores approach a normal distribution.

The data have homogeneity variances.

# Data Source

Information for this analysis came from the Illinois Gaming Board’s website. This data was extract for Auburn, Springfield, and Sangamon County for all available years on a per month basis. You can access the data here: <http://www.igb.illinois.gov/VideoReports.aspx>

# Methodology

## How does the number of video game terminals effect net wagers per location? Is there a significant difference in the mean of net wagers for establishments with different numbers of terminals?

After completing my data corrections, I created a new column in Excel – “NetWagerPerMachine”. For each row, I took the amount of Net Wager divided by the VGTCount column. As a result, I can determine the average Net Wager for each location for each video game terminal, regardless of how many terminals a location may have.

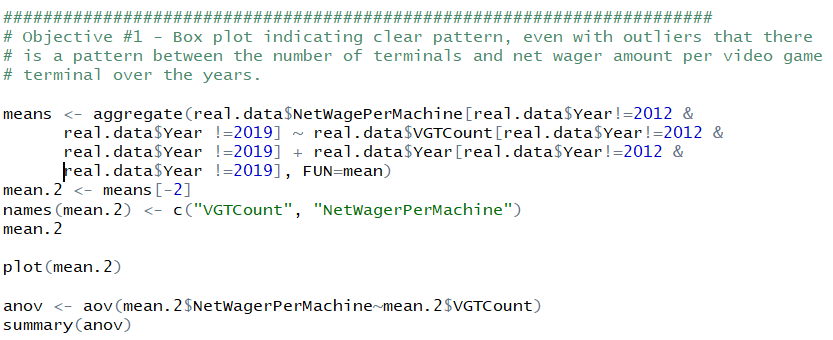
**NOTE**: The higher the Net Wager per video game terminal is, the more the net difference is between Amount Played and Amount Won per video game terminal. A lower Net Wager signifies better results for the one playing the machines.

I imported the data into RStudio.

I created a box plot displaying the raw data to identify any patterns, see [Figure 1](#_Figure_1) in the Appendix. The box plot shows a clear pattern for each year that the net wagers per video gaming terminal is lower when the location has fewer video gaming terminals.

Knowing there was an interesting pattern displayed from my box plot, I decided to take the mean of the NetWagersPerMachine which removed my outliers and helped move my data to a more normal distribution.

I executed the following R code:

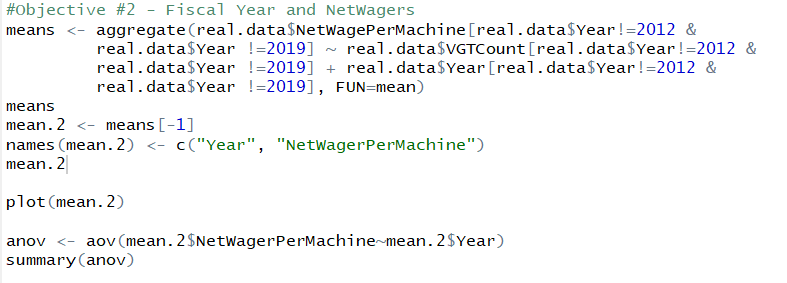


The results of the R code indicated a P value less than .05. Therefor, I rejected the null hypothesis that the means of the Net Wager Per Machine are the same for establishments with different numbers of video game terminals.

## Since video gaming has become legal in the State of Illinois in 2012, there’s no doubt that the number of establishments that provide video game terminals have increased. Does the net wager amount have a significant difference between 2012 and 2019?

To answer this question, my null hypothesis is that there are no significant differences in the mean of NetWager between the fiscal years. My alternative hypothesis is that there is a significant difference between the means of NetWager throughout the years.

To answer this question, I executed the following R code:



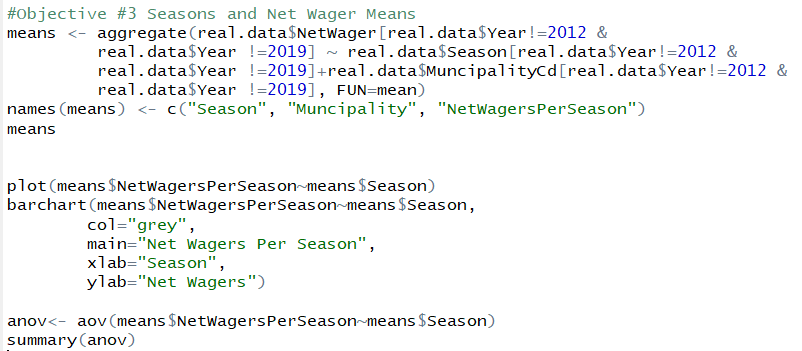
The results of the summary of the ANOVA analysis concluded that the p-value is greater than 0.05 and therefore, I cannot reject the null hypothesis.

## A friend has offered an alternative hypothesis that the mean of net wagers decreases in winter months. Is there a significant difference between the Net Wagers recorded in winter as opposed to other seasons?

I added the column “Season” to my excel spreadsheet and assigned each row of data to a season based on the following:

* + - If the row recorded was for December-February, the season is Winter.
    - If the row recorded was for March-May, the season is Spring
    - If the row recorded was for June-August, the season is Summer
    - If the row recorded was for September-November, the season is Fall.

I executed the following R code to get the results of my analysis:



The results of the summary of the ANOVA analysis concluded that the p-value is greater than 0.05 and therefore, I cannot reject the null hypothesis.

# Conclusions

To conclude my results, the only objective that found a statistically significant effect that had an outcome to reject the null hypothesis was Objective #1. Objective #1 indicates that there is an effect on the number of video game terminals and the Amount Played to Amount Won ratio.

While the results of Objective #2 and Objective #3 were not statistically significant, the returns are still intriguing. Many video gamers are pretty skeptical of the honesty of how money is distributed across the games and what has happened since the law passed in 2012. In the very least, we can assure the population of gamers that there are no significant difference in the mean of the Amount Played to Amount Won ratio between years. In addition, we can conclude that there that the results were inconclusive when attempting to determine if the mean of the Net Wagers were different between the seasons of a year.

# Suggestions for Future Studies

I found it interesting that the means of Amount Played and Amount Won ratio was found to be significantly different across the different groups of establishments with a variety of number of video game terminals. I wonder what other factors might play into that finding. While I constructed the study to consider the number of video game terminals, amount played, and amount won to try to eliminate those extra factors, I am skeptical that there isn’t any other factors that play into the increase in the Amount Played and Amount Won ratios. This is where I would focus if I were to continue the study on the Video Game Data for Central IL.

Another interesting question I have always had with video gaming data is if the day of the week has any bearing on the “luck” for the video gamer. While this data set lacks the information necessary to calculate those results, I suspect that the results could be interesting to those who like to play on the video game terminals.

# Appendix

### Figure 1

