Does ESRB Rating Vary by Genre?

Ethan Salcedo

1 Introduction

Videogames are a big industry, with a report by SuperData stating that in 2018, the videogame industry pulled 119.6 billion dollars in revenue. The rating a game receives by a regulatory board can have the ability to severely limit the number of sales a videogame may receive. Therefore, if there is an association between the genre of a game and the rating it may receive would be beneficial for developers as it can help them make a decision on the types of games they may develop based on who their target audience is. Given the data of many videogames and their respective ratings I will prove that genre can be associated with the rating of a videogame.

2 Data

The data for this analysis was originally compiled together by user Gregory Smith and made publicly available on Kaggle[[1]](#footnote-1). User Rush Kirubi found this set and was inspired by it, choosing to expand upon the data by adding additional variables and republished this expanded data set to his Kaggle[[2]](#footnote-2). The unit of observation for this dataset is the videogame name and Smith initially created the dataset through non-probability sampling, only adding games that sold 100,000 copies or more to his dataset. Kirubi expanded upon this initial collection by conducting a web scrape of the website Metacritic, a popular media review site that has an extensive section for video games. Due to the differences in games listed on each site the limitations of this data show itself in how there are many missing observations in the expanded data set. Of the 16,719 entries Kirubi states that there are about 6,900 entries that have an observation for every variable. Ideally, with this data set we should be able to identify statistically significant relationships to predict the success of games developed in the future.

A screenshot of a cell phone

Description automatically generated

2.1 Variables

For this analysis I will be looking at the genre and rating variable to see if the genre of a game has an influence or can help predict the rating that the game will receive.

2.1.1 ESRB Rating

Ratings are assigned to videogames by a regulatory board, such as the ESRB in the U.S., and serve as recommendations to who should be allowed to play the game. For instance, games with extreme violence or nudity may receive the ‘M’ or ‘Mature’ rating, meaning that only those who are older than 18 should play it. There was one rating, ‘RP’, that was dropped due to how it was used to represent a game that was in the process of receiving a rating, therefore not contributing meaningfully to the dataset. The ratings used in this analysis have been collapsed into 3 categories: E, T, and M. This is due to how the dataset included many ratings that were used in the past by the ESRB but were phased out overtime but still represented three basic groups of games; those that were okay for anyone to play, those that only people who are legal adults should play, and games that fall in between these two categories. Among the categories 54.5% fall into the ‘E’ group, meaning that the majority of videogames released try to make games that can be played by everyone.

2.1.2 Genre

Genre is a variable that is similar to those used by movies and TV shows. For videogames it is used as a way to categorize media based on its gameplay challenges and is not defined by its visual style or narrative. The genre with the largest proportion of videogames is the action genre, making up 22% of the dataset. This genre is only 6.8% higher than 15.2% of the sports genre and 11.8% higher than the shooter genre. The rest of the genres are seen to each make up less than 10% of the games from the dataset.

3 Multivariate Analysis

A screenshot of a video game

Description automatically generatedA picture containing sitting, black, remote

Description automatically generated An initial analysis of how these two categorical variables are distributed via a grouped stack bar chart shows a substantive difference between the genres on ratings. First, the shooter genre has the highest percent of games rated M, with 55% of shooter games within the dataset receiving a M or M equivalent rating. Meanwhile, puzzle, platforming, and sports videogames are almost entirely made up of games that received E or E equivalent ratings. This shows a substantive difference in proportions across genres, with each genre containing proportions that are significantly different from the others. At an alpha of 0.05, I am able to reject the null hypothesis and reach the conclusion that this difference in ratings by genre is statistically significant due to *χ*2(22) = 4256.261*, p* = 0*.*000. This tells me that the rating a game receives is very likely influenced strongly by its genre. Table 2 provides support for this conclusion as the calculated Cramer’s V is seen to be .463, indicating a large substantive effect. This resonates with what is seen on the grouped stack bar chart. In order to have an analysis with 0.8 power the bare minimum number of observations that would need to be present is 31.

4 Conclusion

Overall, the conclusion I have reached by conducting this analysis is that there is both a substantively and statistically significant association between the rating a videogame receives and the genre it is defined as being. Additional studies could aim to illuminate what specifically within each genre has an effect on its ratings, such as levels of violence, complexity of gameplay, and themes.

I believe that the dataset is limited in that it only uses data obtained from Metacritic and VGChartz, meaning that any videogame that isn’t on those two sites would not be represented in this dataset. To fix this, additional web scrapes conducted on other videogame websites.

1. <https://www.kaggle.com/gregorut/videogamesales> [↑](#footnote-ref-1)
2. <https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings> [↑](#footnote-ref-2)