Does Xbox 360 and PS3 platform and genre influence video game critic score?

Arthur Wu

Introduction

Playing video games is a hobby that has grown exponentially in popularity and technology within the past few decades. Two of the largest competing consoles in the video game platform market are Microsoft's Xbox 360 released in 2005 and Sony's PlayStation 3 (PS3) released in 2006[1]. These consoles brought forth new technology into gaming including motion capture and online play, as well as a large selection of new video games of various genres and gameplay. With the PS3 selling 87.41 million units and the Xbox 360 selling 85.80 million units globally, the two gaming platforms constantly compete for dominance in the video game market[2].

In this analysis, I have data on video games varying in platform, genre and publisher that have been released since 1980. This data includes critic scores and user scores drawn from Metacritic, game details such as genre and video game platform, and game sales by region. I will use this sample of video game data to represent the total population of released video games, including games that will be released in the future. Due to the dominance and competitiveness of the Xbox 360 and PS3 consoles, I want to see the distribution of video game critic scores by these two game platforms. There may be a chance that critic scores of video games are dependent on the platform the game is released on. I also want to see if the genre of video games affect critic score, notably the action, racing, role-playing, shooter and sports genres. I also believe that genre has a significant impact on critic scores. The main question that I ask in this study is: Does the difference in game platforms between Xbox 360 and PS3 and video game genre influence video game critic scores? I hypothesize that the choice between Xbox 360 and PS3 platforms and genre have no influence on game critic score. Using data on video game critic scores and video game platforms, I will show that both game platform and genre have little to no influence on the critic score a video game receives.

[1] https://www.did.ie/content/blog/history-of-video-game-consoles
[2] https://www.vgchartz.com/analysis/platform_totals/Hardware/Global/

Data

The dataset used in this study is a video game sales with ratings dataset that was compiled by Rush Kirubi and made available on his Kaggle profile[3]. The columns that this dataset includes are basic game descriptions including game platform, genre and publisher, global and regional sales statistics, and critic and user ratings. Kirubi obtained data on global and regional video game sales from Gregory Smith's Video Games Sales dataset, which was compiled through a web scraping of VGChartz, a popular website used to track video game sales [4]. Kirubi extended Smith's video game sales data by adding critic and user score data from a web scraping of Metacritic.com, which is a review and rating website that gives ratings to video games.

To make the video game sales and ratings dataset more suitable for my study, I performed data cleaning to remove missing critic scores. A majority of video games released before 2001 had missing critic scores in the dataset. This is because Metacritic did not begin rating video games until it was founded in 2001[5]. Therefore, I needed to remove all games that did not have critic scores to prevent skewed results. I also filtered the Platform column so that the dataset would only contain the platforms that are important to the study which are the PS3 and X360 platforms. Furthermore, I filtered the Genre column of the dataset to only include action, racing, role-playing, shooter and sports genres, which are the five most populated genres for PS3 and Xbox 360 which I am interested in observing. Lastly, I removed all games that had a critic score under 30.0 since there were only 7 entries within that interval and could skew the results of this study. Before cleaning, the dataset contained 16,719 data entries. After cleaning, the dataset contained 1,327 entries.

Variables

To analyze the influence of platform and genre on video game critic score, I will use critic score observations, the platform column containing only PS3 and Xbox 360 video games, and the genre column containing only action, racing, role-playing, shooter and sports genres.

---

[3] https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings
[4] https://www.kaggle.com/gregorut/videogamesales
[5] https://www.metacritic.com/about-metacritic

Critic Score

The critic score data used in this study was taken from the "Metascore" on Metacritic's website. To determine the Metascore of a video game, Metacritic takes the weighted average of all critic scores and creates a final numeric score on a 100-point scale[6].

Table 1: Summary Statistics for Video Game Critic Score

Variable	Frequency	Min	Max	Mean	Median	Std. Dev.	Std. Error
Critic_Score	1327	30	98	69.89	72	14.15	0.3884

Examining the summary statistics for the critic score variable shown in table 1, we can see that the sample mean of critic score is 69.89 and the median is 72. This shows us that on average, critics rate Xbox 360 and PS3 games in the genres of interest around 70 points on the 100 point scale, which is 20 points above the halfway mark of 50 on the variable's 100-point scale. We can also see that the standard deviation is 14.15, which tells us that the spread of data is relatively concentrated around the average. The standard error is 0.39 which is small and indicates that our sample of critic scores represents the general population well.

Platform and Genre

The two categorical variables that are used in this study are video game platform and genre. The platforms that I am interested in are PS3 and Xbox 360 (X360). Examining the univariate summaries of the two platforms featured in the dataset shown in table 2, we can see that PS3 and Xbox 360 have very similar frequencies and percentages. PS3 has a frequency of 637 and a percentage of 48%, while Xbox 360 has a frequency of 690 and a percentage of 52%.

Table 2: Univariate Summaries for Video Game Platform and Genre

Category	Frequency	Percentage
Platform		
PS3	637	48 %
X360	690	52 %
Genre		
Action	485	36.5 %
Racing	142	10.7 %
Role-Playing	136	10.2 %
Shooter	281	21.2 %
Sports	283	21.3 %

[6] https://www.metacritic.com/about-metascores

This similarity in frequency and percentage reflects that our data sample is balanced in size between the two platforms we are studying.

For the genre category, I specifically want to analyze the five most populated genres for PS3 and Xbox 360 video games which include action, racing, role-playing, shooter and sports genres. Examining the summaries for genre in table 2, we can see that the frequencies and percentages of each genre is more varied compared to the platform category. Racing and role-playing genres have percentages of around 10%, shooter and sports genres have percentages around 21% and action genre has a percentage of 36.5%. The difference in frequency between the post populated genre (action) and least populated genre (role-playing) is 349 observations.

Analysis and Interpretation

To examine the relationship between platform, genre and critic score, I will use a two-way ANOVA test. For my specific study, platform and genre are used for the categorical predictor variables and critic score is used as the numerical outcome variable. I will also use 0.05 as my alpha value.

Distribution Visualization

Before performing the two-way ANOVA test, I visualized and examined the distribution of my critic score outcome variable by my categorical predictor variables using violin graphs.

Examining the graph from figure 1, we can see that both Xbox 360 and PS3 platform distributions are negatively skewed and unimodal. We can also see that both distributions look very similar and have a critic score mean of around 70 points. From the results of this plot, I conclude that there seems to be very little difference in the

Figure 1: Distribution of Video Game Critic Score by Platform

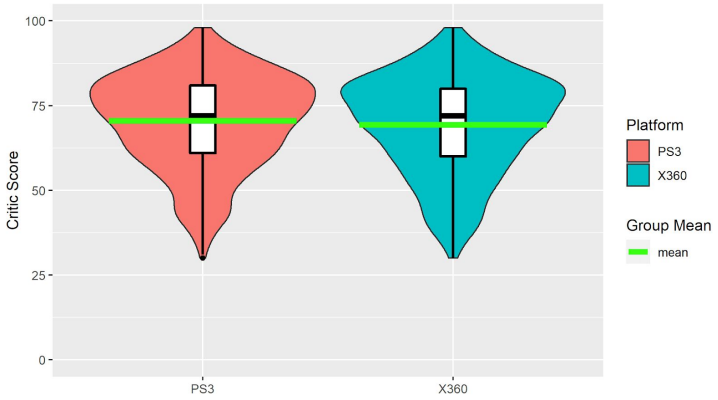
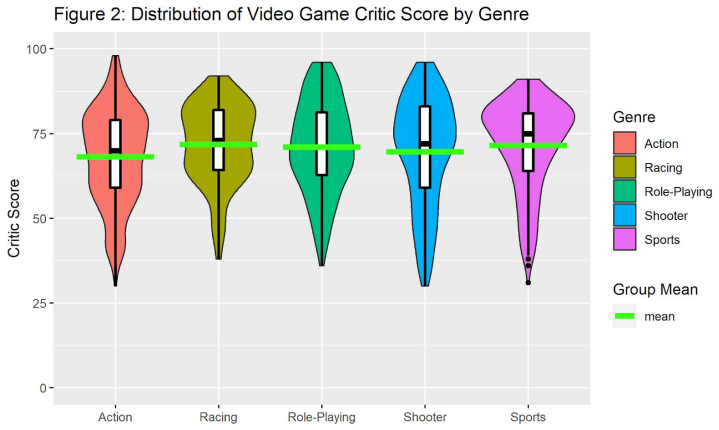


Figure 2: Distribution of Video Game Critic Score by Genre



variance between the two distributions and the means also do not vary by group as well.

Looking at figure 2, we can see that the distributions of critic scores by video game genre are more varied than compared to platform. All distributions by genre are negatively skewed which suggests a majority of critic scores are concentrated in the upper end of the critic score scale. The action and racing genre distributions are bimodal while the role-playing, shooter and sports distributions are unimodal. Inspecting the variance of the distributions, there seems to be variance for the sports and racing genres, which at a glance have stronger negative skewness. The mean of all distributions fall between a 65 and 75 which suggests little difference in mean. Overall, the distributions grouped by genre have slight hints of variance, but have little difference in group means.

Verifying Assumptions

Before performing the ANOVA test, we also must verify the six assumptions of the ANOVA analysis. The first assumption is that the dependent variable is numeric. Our critic score is a numeric variable on a 100 point scale which confirms the first assumption. The second assumption is that group sample sizes are approximately equal. For our study, the groups by platform are very close in magnitude and groups by genre are not as close, but are close enough in magnitude to ignore the difference. The third assumption is that observations must be independent from one another. Each observation in our dataset is a different video game which suggests independence. The fourth assumption is that there are no extreme outliers which we were able to satisfy by removing outliers in our original dataset during the data cleaning step of the study.

The ANOVA test also assumes homogeneity of variance which we check using Levene's test. When performing Levene's test on the platform variable, we obtain a p-value of 0.195, which is larger than our alpha value of 0.05. Therefore, we can fail to reject the null hypothesis and do not violate homogeneity of variance. Performing Levene's test on the genre variable gives us a p-value of 0.0005, which is less than our alpha value and results in violating the homogeneity of variance. Lastly, we perform Levene's test on the interaction between platform and genre variables and get a result of 0.004, which is smaller than our alpha value causing us to violate the homogeneity of variance again. Even though we violate homogeneity of variance in

two of the three effects, we will continue to perform the two-way ANOVA test and assume that the variances are equal while acknowledging that we are violating the homogeneity of variance assumption.

ANOVA Test

The last ANOVA analysis assumption is the normality of residuals which will be checked after performing the ANOVA test. Now that we have checking assumptions, I am able to perform the two-way ANOVA test.

ANOVA Results: The Impact of Video Game Platform and Genre on Critic Score

	df	Sum of Squares	Mean Squares	F-ratio	p-value
Platform	1	465.27	465.27	2.345	0.126
Genre	4	3,017.58	754.39	3.803	**0.004**
Interaction	4	693.65	173.41	0.874	0.479
Residuals	1317	261,268.22	198.38		

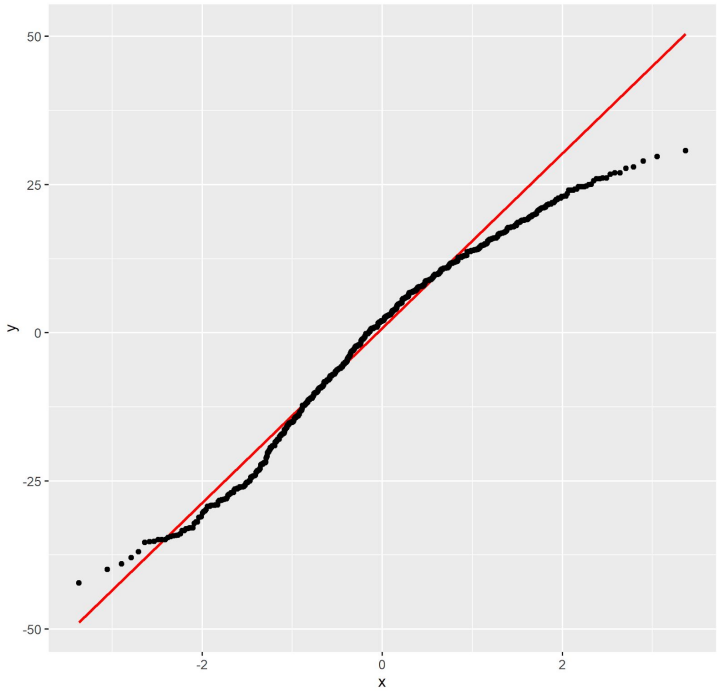
Examining the results of the ANOVA test, we can see that the p-value for the platform effect is 0.126, which is larger than our alpha value of 0.05. Therefore, we fail to reject null and state that platform is not a significant predictor of video game critic score. We can also see that the p-value of the genre effect is 0.004, which is less than our alpha value and allows us to reject null and conclude that genre is a significant predictor of video game critic score. Lastly, the p-value for the intersection between the platform and genre effects is 0.479, which is larger than our alpha value. Therefore, we fail to reject null and claim that the intersection between video game platform and genre is not significant. In other words,

Normality of Residuals

For our final assumption check, I will use a QQ plot to examine the normality of residuals. Looking at Figure 3, we can see that the plot for residuals of the ANOVA test has slight deviation at the lower tail and greater deviation at the upper tail. The upper tail dips below the reference line of normality and the lower tail rises above the reference line which suggests that the distribution of residuals is a light tailed distribution. This may be related to the removal

of outliers less than a critic score of 30 during data cleaning and the absence of a video game in our dataset with critic scores of 99 or 100, effectively causing the tails of the distribution to lack data. Since a large majority of residual plots lie close to or on the reference line, I would still claim that the distribution of residuals is still normally distributed despite the presence of light tails and successfully address the final assumption for ANOVA tests.

Figure 3: Plot of Residuals



Pairwise HOC Comparison

Now, we will analyze the pairwise differences between each grouping of main effects in our model using Tukey's Honest Significant Difference to see which groups differ. Examining the TukeyHSD results for the platform main effect, we see that there is only one interaction pair between the PS3 and Xbox 360 platforms. With a p-value of 0.126, we fail to reject null and determine that the means between the two platforms are not significantly different from one another. This finding is further supported by the distribution of critic score by platform plot created during our visualization step as the means in the plot look extremely close to one another.

Analyzing the genre main effect, we see that a majority of the genre pairs have p-values above our alpha value of 0.05, which implies that a majority of genre pairs are not significantly different from one another in regards to their means. However, the Racing-Action and Sports-Action pairs have p-values of 0.042 and 0.01 respectively. This allows us to reject null for the two genre pairs and reveals that there is a significant difference in means between racing and action genres, and between the sports and action genres.

Lastly, we will examine the pairwise differences of the interaction between platform and genre. Looking at the p-values of the interaction pairs, almost all interaction pairs have p-values

greater than our alpha value which indicates that almost all interaction pairs have no significant difference in means. Another observation is that a majority of interaction pairs above our alpha value have p-values of around 1.0, which further reinforces no significant difference in means between platform and genre interaction pairs. The only interaction pair that has significant results is PS3:Sports - X360: Action with a p-value of 0.005 and a difference of 5.91, which means that the critic score averages between PS3 sports games and Xbox 360 action games are significantly different.

Effect Size

The final step of two-way ANOVA analysis is a study of effect size to determine if the effect of the study is large enough to matter substantively. I will analyze effect size using Cohen's F, R-squared, and partial eta-squared.

For the platform main effect, the difference in means between PS3 and Xbox 360 is 1.185 critic score points, which is an extremely small difference on the critic score variable's 100-point scale, Leading me to state that the difference of means in critic score between Xbox 360 and PS3 platforms is not substantially significant. For the genre main effect, the largest difference in means is 3.7 critic score points between the racing and action genres which is also a very small difference in critic score. Due to this, I would state that the difference of means in critic score by video game genre is also not substantially significant.

Looking at the Cohen's F for the Anova analysis, we obtain a result of 0.126. This result suggests a very small effect size. For our r-squared test, we get a result of 0.0157, which suggests that the overall combination of platform, genre and the interaction between platform and genre explains 1.57% of the variance in video game critic score. This percentage is extremely small and conveys little to no significant impact. The Partial-Eta-squared test reveals that platform explains approximately 0.2% of the variance in critic score, genre explains around 1% of variance and platform-genre interaction explains around 0.3% of variance. Therefore, we can conclude that genre is the strongest predictor of critic scores within the 1.57% r-squared value and within the entire model. Overall, the results of all three standardized effect tests support my unstandardized effect size claim that the difference of means by video game platform and genre are not substantially significant.

Conclusion

In all, I conclude that both game platform (between Xbox 360 and PS3) and genre have little to no influence on the critic score a video game receives, supporting my hypothesis that platform and genre have no influence on video game critic score. Directly generalizing the population of current and future games that our sample represents, the choice of platform between PS3 and Xbox 360 and video game genres do not influence video game critic score. Applying the results of the ANOVA test, we can draw the practical conclusion that how well a video game is enjoyed and rated by critics is not affected by whether it is a game on PS3 or Xbox 360, and is also not affected by the type of genre it is. Instead, those who seek the factors that influence video game critic score should look elsewhere. Our test is also substantively insignificant which means that our observed effect is not large enough to be meaningful. A bias that may exist in our dataset is that we only look at the Xbox 360 and PS3 platforms. Therefore, our results cannot be applied to other video game platforms.

By cleaning and filtering the dataset of this study to only include video games released on the PS3 and Xbox 360 within the five most populated video game genre categories, and the removal of all video games with critic scores less than 30, my dataset was very comprehensive and allows me to apply the results of this study to the larger population. One small error that may have skewed my ANOVA test may be that the group sample sizes were not as close as they could have been. If this error was addressed and the group sample sizes were balanced to be closer, our results would be more precise.