Stat 461 Final Project Group 3: Video Game Sales

Ben Gutierrez, Zining Yin, L Hansen, Luciana Lai, Yurui Niu

4/18/2022

Executive Summary

Problem statement

In recent decade, video games have played a more and more important role in worldwide entertainment. Due to the giant market of both now and future of video games sales, video games' makers become popular. They have produced video games with different genres and platforms. To understand if genres and platforms affects the number of video game sales.

Proposed solution

We construct an observational study and use two-way ANOVA to analyze a dataset generated by a scrape of vgchartz.com which found in kaggle. Together with the use of R studio, we accomplish our goal by the following steps to understand if genres and platforms affects the number of video game sales:

- Clean data of the selected dataset and explore the dataset
- Use knowledge of analysis of variance to
  - Check the appropriateness of ANOVA
  - Check interactions
  - Fit the model
- Check the assumptions of ANOVA which includes
  - Gaussian residuals
  - Homoscedasticity
  - Independence of observations
- Get result by using
  - Omnibus
  - Point estimates
  - Post Hoc-Pairwise comparison of marginal means
  - Pairwise comparison
  - Effect sizes
  - Post Hoc-Contrasts

Value

By using two-way ANOVA to do further exploration of the observational study, we know how genres and platforms affect the sale of video games. In addition, the data we are using are all games exceed 10,000 copies. Thus, our information can lead to more sales on video games, which help future video game makers to decide the combination of genres and platforms with the most profits. The specific genre and the platform can be a good start point for future makers who do not know much about the game area.

Final thoughts & next steps

Genres and platforms can be two critical points of the beginner makers. The two piece of information shows what players mostly interested in. For the next stage, we will further explore the other possible characteristics on video games sales such as the price on game sales, producers, etc.

Introduction and Background

Different characteristics of games often result in number changes of game sales. Among all the characteristics listed in the dataset of video game sales which generated by a scrape of vgchartz.com, two specific characteristics, genres and platforms, might have effect on sales of video games which were sold greater than 10,000 copies. In this context, our study focuses on genres and platforms to understand the impact toward sale of video games.

Literature Review

The video game industry, an industry that is equal parts entertainment and software development, has been the subject of ever-mounting academic interest. Though many factors drive a particular title's sales, the platforms it can be played on drive a great deal of a consumer's decision making. Hardware manufacturers such as Microsoft (Xbox), Sony (PlayStation), and Nintendo create a market where consumers of a title all interact on a similar platform (Bellovich 2020; Davidovici-Nora, Bourreau 2012). This creates an incentive for a consumer to stick with a platform, even as some titles cross platforms. Multi-homing, when developers create a game for multiple platforms, still gives the platform owner control over the title's appearance on the platform. For a developer to make a game compatible across multiple consoles is a difficult and expensive process that is only worthwhile in some cases (Landsman & Stremersh 2011).

A game's platform also has a more indirect impact as it pertains to multiplayer titles. In these games, a consumer typically derives enjoyment from the presence of other players within the game. This creates a bandwagon effect where consumers gravitate toward interacting with other players on a similar platform (Buxmann et al., 2007). Even for multi-homed games, multiplayer games do not typically connect a player on one platform with a player on another.

While platform is an important expression of the way in which consumers interact with developers, it does not tell the whole story. A consumer does not choose a game by randomly selecting a title that is compatible with their hardware, but a consumer may gravitate toward a type or style of game that they enjoy. Our team seeks to expand upon the discussion of platform effects on video game sales by additionally considering a range of general genres.

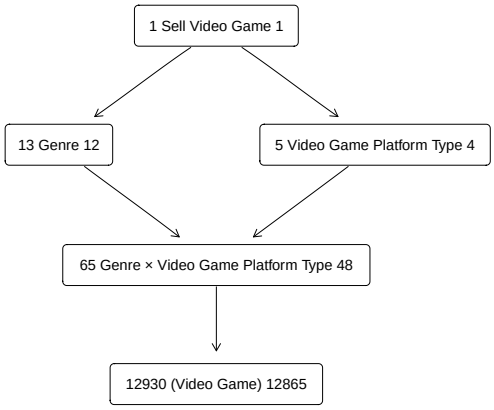
Study Design and Methods

To investigate our research question, we did an observational study. We found a dataset generated from a scraping of vgchartz.com, which has 55,792 records. The script to scrape data is available at https://github.com/GregorUT/vgchartzScrape. The script is mainly based on Beautiful Soup using Python. The datasets contains variables rank, name, genre, platform, global_sales, and so on.

The dataset is publicly available at:

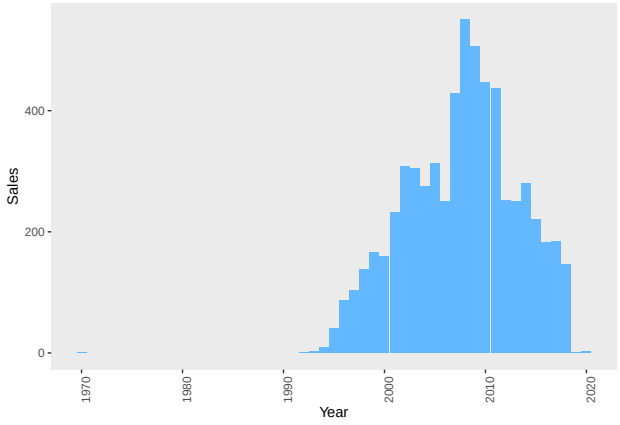
https://www.kaggle.com/datasets/ashaheedq/video-games-sales-2019

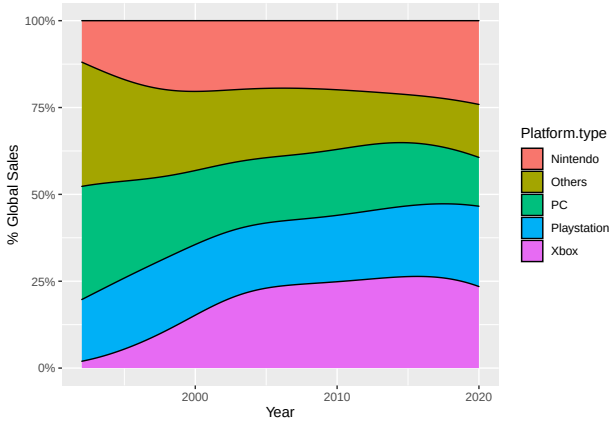
Our primary response is which factors affecting game sales. We are interested in the genre and platform of the top selling video games. So two-way ANOVA method is appropriate in our study. The following figure shows the Hasse diagram for our video games sale study. From the Hasse diagram, we have two categorical variables: genre and platform. We will estimate the main effects, interaction, and residuals in the following work.

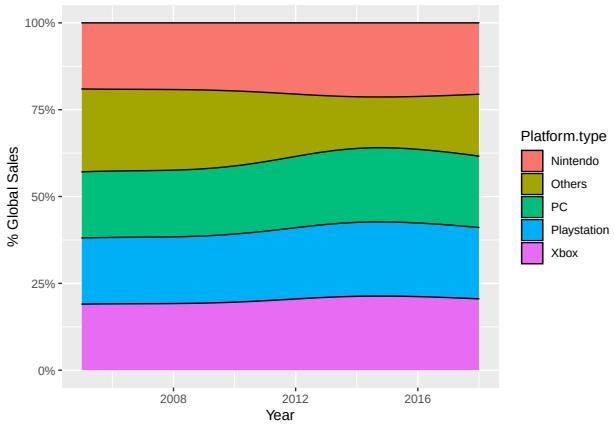


Thus, we will use the following Null hypotheses: There is no significant impact on how many units a video game will sell due to its genre, There is no significant impact on how many units a video game will sell due to its platform, There is no significant interaction effect on how many units a video game will sell between genre and platform.
Out Alternative hypothese are: There is a significant impact on how many units a video game will sell due to its genre, There is a significant impact on how many units a video game will sell due to its platform, There is a significant interaction effect on how many units a video game will sell between genre and platform.
For multiple comparisons, we will use Tukey's HSD.

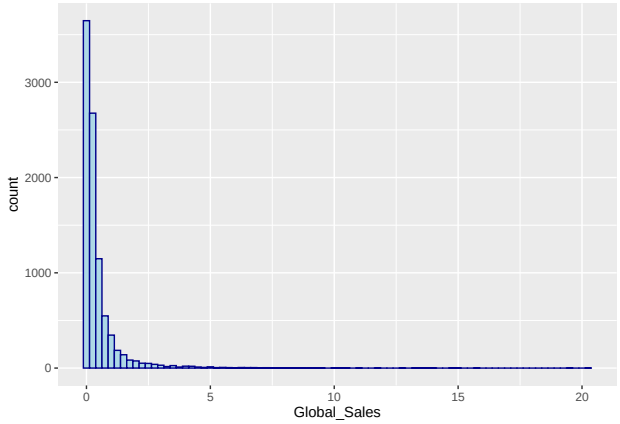
Exploration of the Data



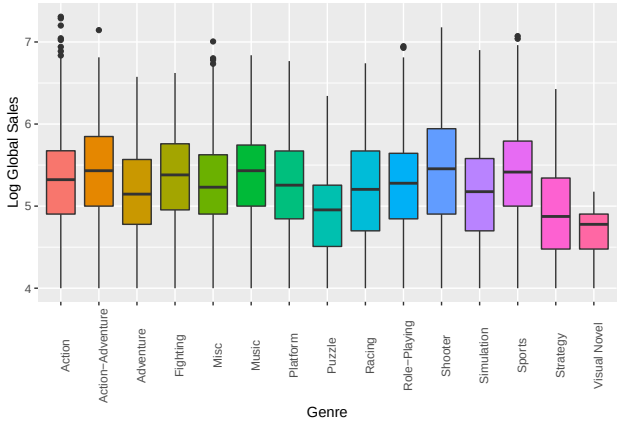


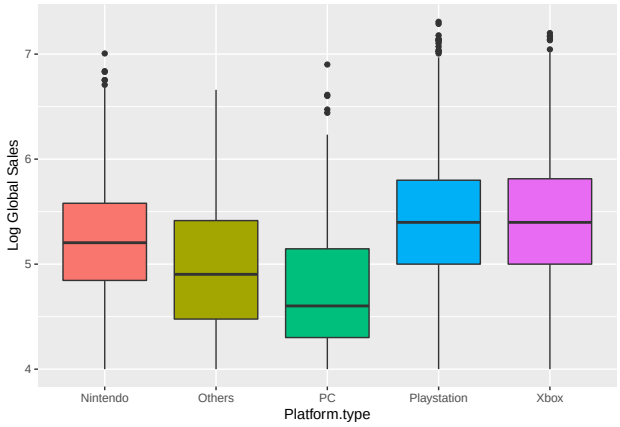


Before doing any kind of visualization of the data, we went through the initial dataset and removed all rows without either Global_Sales or Total_Shipped values. We saw this as the potential real response variable, so any row with blanks for those columns were removed. Additionally, we removed all rows that listed Global_Sales or Total_Shipped as zero because these games likely had sales, but showed up as zero due to rounding.It is important to note that these variables are in terms of millions of units and go up to two decimal places. Originally, there was reason to suspect that Global_Sales and Total_Shipped referred to the same value upon cross-examining the data with other sources on video games sales, despite the names of the columns indicating otherwise. Therefore, we played around with combining the two columns into a new one called Total_Sales because we noticed that every row either had a value in Global_Sales or in Total_Shipped, but not in both. However, upon investigating further, we chose to only use Global_Sales as it seemed possible that the reason why the two columns were separate was because the data came from different sources. Additionally, it created some extra issues with using ANOVA methods. While exploring the data, using Global_Sales as the response, we looked at a few different potential factors that could explain the response of sales in terms of units sold. The main three we focused on were genre, platform, and ESRB rating. However, upon inspecting ESRB rating further, we noted that there was some problems with it because it included RP which stood for rating pending. This doesn't make much sense because every game has to have a rating when it is sold, so it made rows invalid. Thus, we decided to only focus on genre and platform as factors. Since platform has many different values because many platforms were released at different times, we decided to combine platform into groups based on the platform series it came from. Therefore, we have Nintendo, Playstation, Xbox, pc, and other. Since not all platforms existed at the same time, we decided to check view the percentage of each group throughout the years. Initially, we see the other and pc group take up the majority of the percentage. However, when we limit the data to just platforms released after 2004 until 2019 (which is when the data was collected), we see that the percentages are more or less equal among each group.

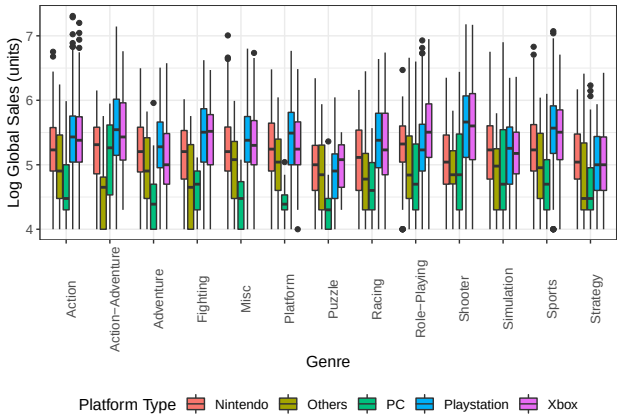


Based on the histogram above, it is very clear that we have an extremely right skewed response variable. Therefore, we need to take this into account.





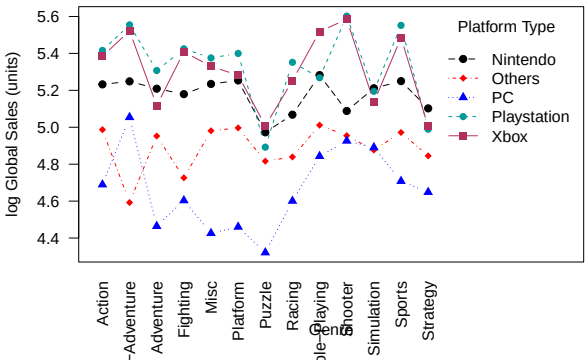
Since Global Sales appears to follow an exponential distribution, we will apply a log function after multiplying Global_Sales by 1 million so that it is in terms of units instead of millions of units. Now looking at the box plots for each factor individually, we see that there is a notable amount of variation between the levels of the individual factors. While this is good to look at, we also need to look at the boxlots for the combinations of the levels of the factors. We will be removing the Visual Novel and Music levels due to low observation counts.



Looking at these boxplots, we see that every combination is represented in the visualization. As we can see, the sample medians of the boxplots vary between platform type more noticeably. Most of the outliers shown appear at the top, indicating that few games sold extremely well compared to what is typical.

In the table above, we have the summary statistics for each combination of our two factors (genre and platform type). One important thing to note is that we do not have a balanced design. This can be seen by looking at the sample size. Therefore, when we review the ANOVA model we will choose the Type 2 Sums of Squares. Every possible combination is represented in this table, so we do not have to worry about what to do when one or more is missing. Looking at the sample arithmetic means there seems to be difference in performance for each combination. Additionally, there seems to be different amounts of variation as evident by the sample arithmetic deviations.

The Model



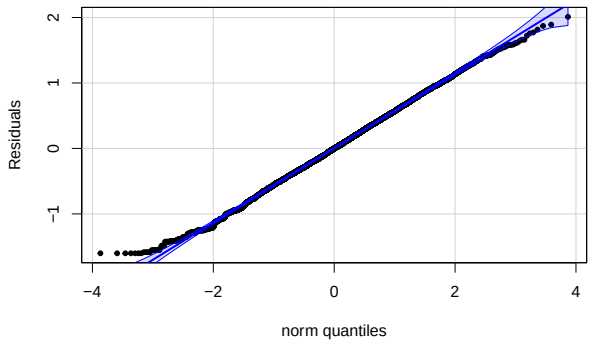
Here we see that there is noticeable amount of interaction between the two factors. Especially when considering Nintendo, Playstation, and Xbox (the three biggest gaming consoles). This indicates to use that we need to use the interaction term in our ANOVA model.

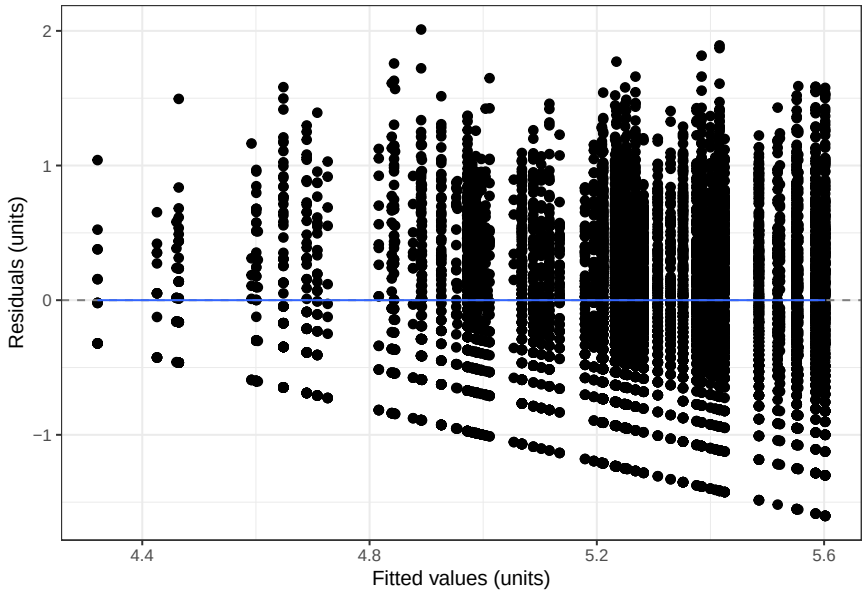
Table 1: Summary Statistics for Video Games Log(Global Sales)

	n	Min	Q1	Median	Q3	Max	MAD	SAM	SASD	Sample Skew	Sample Ex. Kurtosis
Action-Adventure x Nintendo	62	4.000	4.860	5.312	5.582	6.152	0.446	5.248	0.483	-0.456	-0.522
Action-Adventure x Others	8	4.000	4.000	4.650	4.809	5.756	0.669	4.592	0.603	0.563	-0.914
Action-Adventure x PC	10	4.000	4.533	5.263	5.617	5.949	0.816	5.054	0.719	-0.302	-1.606
Action-Adventure x Playstation	84	4.000	5.146	5.544	6.017	7.144	0.663	5.554	0.656	-0.052	-0.418
Action-Adventure x Xbox	63	4.301	5.073	5.431	5.961	6.761	0.649	5.522	0.570	-0.025	-0.666
Action x Nintendo	466	4.000	4.903	5.230	5.577	6.753	0.485	5.232	0.498	-0.122	0.047
Action x Others	83	4.000	4.477	4.903	5.462	6.245	0.680	4.987	0.574	0.048	-1.053
Action x PC	69	4.000	4.301	4.477	5.000	5.987	0.546	4.689	0.553	0.706	-0.478
Action x Playstation	626	4.000	5.041	5.431	5.756	7.308	0.551	5.416	0.579	0.124	0.124
Action x Xbox	364	4.000	5.041	5.380	5.744	7.200	0.510	5.384	0.583	0.084	0.141
Adventure x Nintendo	316	4.000	4.889	5.204	5.583	6.497	0.532	5.209	0.527	-0.038	-0.196
Adventure x Others	27	4.000	4.477	4.903	5.419	5.826	0.631	4.953	0.571	-0.068	-1.198
Adventure x PC	48	4.000	4.000	4.389	4.699	5.959	0.518	4.464	0.427	0.974	1.232
Adventure x Playstation	185	4.000	4.954	5.279	5.663	6.506	0.555	5.307	0.511	0.047	-0.468
Adventure x Xbox	96	4.000	4.699	5.000	5.484	6.575	0.590	5.117	0.585	0.431	-0.396
Fighting x Nintendo	77	4.000	4.778	5.204	5.532	6.017	0.532	5.179	0.457	-0.193	-0.405
Fighting x Others	12	4.000	4.000	4.650	5.313	5.756	0.965	4.726	0.666	0.228	-1.619
Fighting x PC	5	4.000	4.301	4.699	4.903	5.114	0.590	4.603	0.451	-0.188	-1.955
Fighting x Playstation	184	4.000	5.041	5.505	5.869	6.622	0.618	5.425	0.582	-0.442	-0.204
Fighting x Xbox	89	4.000	5.000	5.519	5.778	6.470	0.600	5.409	0.575	-0.460	-0.475
Misc x Nintendo	491	4.000	4.903	5.204	5.585	7.006	0.532	5.234	0.535	0.122	0.115
Misc x Others	49	4.000	4.477	5.079	5.362	5.991	0.522	4.981	0.562	-0.240	-0.891
Misc x PC	15	4.000	4.000	4.477	4.739	5.079	0.446	4.426	0.364	0.100	-1.433
Misc x Playstation	230	4.000	5.041	5.380	5.748	6.801	0.534	5.375	0.549	0.008	-0.137
Misc x Xbox	127	4.000	5.000	5.301	5.686	6.735	0.514	5.330	0.532	0.157	-0.052
Platform x Nintendo	192	4.000	4.903	5.243	5.646	6.480	0.590	5.253	0.509	-0.185	-0.494
Platform x Others	63	4.000	4.602	5.041	5.398	6.045	0.578	4.997	0.571	-0.119	-0.912
Platform x PC	12	4.000	4.301	4.389	4.533	5.041	0.130	4.460	0.283	0.548	-0.573
Platform x Playstation	150	4.000	5.000	5.491	5.813	6.767	0.573	5.400	0.598	-0.244	-0.345
Platform x Xbox	60	4.000	5.000	5.243	5.665	6.484	0.434	5.282	0.513	-0.095	-0.302
Puzzle x Nintendo	273	4.000	4.602	5.000	5.301	6.342	0.536	4.972	0.533	0.065	-0.356
Puzzle x Others	19	4.000	4.301	4.845	5.305	5.939	0.807	4.816	0.672	0.240	-1.457
Puzzle x PC	17	4.000	4.000	4.301	4.477	5.362	0.446	4.321	0.397	1.055	0.264
Puzzle x Playstation	66	4.000	4.477	4.903	5.169	6.045	0.466	4.892	0.470	0.360	-0.454
Puzzle x Xbox	15	4.301	4.650	5.079	5.310	5.505	0.546	5.008	0.393	-0.287	-1.385
Racing x Nintendo	183	4.000	4.602	5.114	5.538	6.161	0.656	5.068	0.578	-0.214	-0.871
Racing x Others	29	4.000	4.301	4.778	5.176	6.450	0.707	4.839	0.667	0.598	-0.517
Racing x PC	47	4.000	4.301	4.602	5.034	5.568	0.522	4.601	0.500	0.368	-1.184
Racing x Playstation	266	4.000	4.954	5.380	5.799	6.641	0.631	5.352	0.623	-0.218	-0.554
Racing x Xbox	173	4.000	4.845	5.230	5.799	6.740	0.682	5.252	0.658	-0.048	-0.684
Role-Playing x Nintendo	177	4.000	5.041	5.322	5.602	6.470	0.416	5.282	0.452	-0.631	0.594
Role-Playing x Others	38	4.000	4.477	4.841	5.447	6.660	0.800	5.011	0.673	0.559	-0.541
Role-Playing x PC	68	4.000	4.301	4.699	5.322	6.601	0.639	4.843	0.695	0.619	-0.657
Role-Playing x Playstation	354	4.000	4.903	5.230	5.628	6.928	0.526	5.268	0.531	0.306	0.254
Role-Playing x Xbox	97	4.000	5.114	5.505	5.944	6.948	0.631	5.518	0.635	-0.059	-0.508
Shooter x Nintendo	130	4.000	4.699	5.041	5.462	6.350	0.508	5.088	0.543	0.356	-0.538
Shooter x Others	12	4.000	4.703	4.845	5.218	5.839	0.446	4.955	0.569	0.128	-1.149
Shooter x PC	112	4.000	4.301	4.845	5.477	6.441	0.807	4.926	0.663	0.255	-1.036
Shooter x Playstation	315	4.000	5.114	5.663	6.068	7.179	0.679	5.602	0.682	-0.135	-0.317
Shooter x Xbox	289	4.000	5.079	5.602	6.104	7.171	0.775	5.585	0.694	-0.042	-0.474
Simulation x Nintendo	352	4.000	4.778	5.230	5.605	6.753	0.582	5.211	0.558	-0.115	-0.339
Simulation x Others	16	4.000	4.301	4.979	5.248	5.799	0.535	4.877	0.560	-0.165	-1.331
Simulation x PC	82	4.000	4.301	4.699	5.544	6.901	0.590	4.891	0.715	0.667	-0.502
Simulation x Playstation	92	4.000	4.699	5.255	5.585	6.348	0.608	5.195	0.570	-0.131	-0.559
Simulation x Xbox	58	4.000	4.860	5.176	5.505	6.364	0.489	5.134	0.503	-0.241	0.030
Sports x Nintendo	388	4.000	4.903	5.230	5.623	6.830	0.518	5.250	0.542	-0.100	-0.042
Sports x Others	23	4.000	4.477	4.954	5.488	6.041	0.707	4.972	0.635	-0.053	-1.389
Sports x PC	43	4.000	4.301	4.699	5.079	6.100	0.590	4.708	0.549	0.396	-0.679
Sports x Playstation	489	4.000	5.176	5.568	5.914	7.072	0.528	5.552	0.565	-0.084	0.026
Sports x Xbox	306	4.000	5.079	5.505	5.851	6.708	0.580	5.485	0.571	-0.160	-0.443
Strategy x Nintendo	75	4.000	4.778	5.041	5.477	6.170	0.508	5.102	0.503	0.016	-0.566
Strategy x Others	11	4.000	4.301	4.477	5.338	6.413	0.707	4.845	0.805	0.666	-1.076
Strategy x PC	107	4.000	4.301	4.477	4.954	6.230	0.446	4.648	0.571	0.900	-0.052
Strategy x Playstation	74	4.000	4.602	5.000	5.438	5.939	0.590	4.989	0.541	-0.154	-0.823
Strategy x Xbox	43	4.000	4.602	5.000	5.431	6.426	0.639	5.004	0.581	0.086	-0.695

Assumptions

Gaussian assumption and Homoscedasticity





Given the data's strict adherence to the diagonal of the Quantile plot, we will assume the normality of our sample, especially when given its size. This is despite data points falling outside of the envelope toward the lower end of the QQ-plot.

The plot mapping our residuals demonstrates some fanning, as there appears to be a straight, decreasing line, beyond which no observation's residual will drop. This is likely due to the effect of rounding that we found in the data. While the data demonstrates some imperfection, we will proceed with caution, keeping in mind the potential for inaccuracy in some point estimates.

Independence of observations

This assumption is difficult to assess given that there is no data on the measurement order. Therefore, we need to be cautious about the independence of observations because the data was likely obtained by gathering data on whatever number the developers released on units sold for each game.

Table 2: ANOVA Table for Video Games Sales - Type II SSQs

Source	SS	df	MS	F	p-value	Partial Omega Sq.	Partial Eta Sq.	Partial Epsilon Sq.
Genre	93.2272	12	7.7689	24.0123	< 0.0001	0.0294	0.0309	0.0296
Platform.type	257.4958	4	64.3739	198.9678	< 0.0001	0.0800	0.0809	0.0805
Genre:Platform.type	60.5870	48	1.2622	3.9013	< 0.0001	0.0151	0.0203	0.0151
Residuals	2923.8260	9037	0.3235					

Results

After checking the data satisfy the assumptions for parametric ANOVA, our team has decided to use a two-way ANOVA model to investigate whether genre and platform of the game impact game sales, as well as set the Unusualness Threshold be to 0.05 before we do any tests. Also, since there are too many values, so it is not feasible to interpret all of them, as a team we decided to choose some extreme values to interpret.

We can see from the table that the genre of video games in sales accounts for 41.4865 (F-ratio = 41.4865) times as much variation as the residuals did when accounting for the impact of platform type and the interaction of the two. Under the null model of no platform type impact and the interaction impact, we would expect to see a result at least as extreme as <0.0001 of a percent of the time. Thus, we will reject the null hypothesis that genre doesn't impact game sales when our unusualness threshold is 0.05. We will decide to act as if the genre of game impact the sales.

Platform type accounts 198.9678 (F-ratio = 198.9678) times as much variation as the residuals did when accounting for the impact of genre of the game and the interaction. We would expect to see a result of at least as extreme as <0.0001 of percent of the time under the null model of no genre impact and the interaction impact. Therefore, we decide to reject the null hypothesis that platform has no impact on game sales when compared to UT=0.05 and act as if the platform of games impact the sales.

The interaction of genre and platform type accounts 3.9013 (F-ratio = 3.9013) times as much as variation as the residual when accounting the other two main effects in the model. Under the null model of no main effect and UT=0.05, we expect this result closely 0 percent of the time. Thus, we decide to reject the null that the interaction of genre and platform type doesn't impact game sales and act as if it does.

In particular, the genre of video games accounts for around 5% of the variation in the sales, whereas platform type accounts for 8% of the variation in sales and the interaction accounts for 1.51% of the variations in sales of video games.

Point Estimate

Table 3: Point Estimates from the Video Games Sales study

	Estimate
Grand Mean	5.08
Action	0.07
Action-Adventure	0.12
Adventure	-0.07
Fighting	-0.01
Misc	-0.01
Platform	0.00
Puzzle	-0.27
Racing	-0.05
Role-Playing	0.11
Shooter	0.16
Simulation	-0.01
Sports	0.12
Strategy	-0.16
Nintendo	0.10
Others	-0.19
PC	-0.41
Playstation	0.26
Xbox	0.24
Action x Nintendo	-0.01
Action-Adventure x Nintendo	-0.05
Adventure x Nintendo	0.09
Fighting x Nintendo	0.01
Misc x Nintendo	0.06
Platform x Nintendo	0.07
Puzzle x Nintendo	0.07
Racing x Nintendo	-0.06
Role-Playing x Nintendo	-0.01
Shooter x Nintendo	-0.25
Simulation x Nintendo	0.05
Sports x Nintendo	-0.05
Strategy x Nintendo	0.08
Action x Others	0.03
Action-Adventure x Others	-0.42
Adventure x Others	0.13
Fighting x Others	-0.16
Misc x Others	0.10
Platform x Others	0.10
Puzzle x Others	0.20
Racing x Others	0.00
Role-Playing x Others	0.01
Shooter x Others	-0.09
Simulation x Others	0.00

Ignoring all factors, our entire sample accumulated 5.08 times as many sales in millions per game as sampled video games this dataset. For the game genre of shooter, they accumulated an additional of 0.16 sales in millions per game whereas Sports games only accumulated an additional of 0.12 million sales per game and puzzle games accumulated -0.27 million sales per game. The platform of PlayStation accumulated 0.26 million sales per game whereas PC reduce the sale by 0.41 million per game. Action-Adventure game on platform of PC accumulated an additional of 0.27 million sales per game, while shooter game on Nintendo platform reduce video game sales by 0.25 million per game. This suggests that puzzle games and PC platform performing worse that baseline (GSAM), as well as shooter game on Nintendo platforms.

Pairwise comparison

Table 4: Marginal Means-Tukey 90% Adjustment

Genre	Platform type	Marginal Mean	SE	DF	Lower Bound	Upper Bound
Action	Nintendo	5.2323	0.0263	9037	5.1889	5.2756
Action-Adventure	Nintendo	5.2484	0.0722	9037	5.1296	5.3673
Adventure	Nintendo	5.2090	0.0320	9037	5.1563	5.2616
Fighting	Nintendo	5.1793	0.0648	9037	5.0726	5.2859
Misc	Nintendo	5.2345	0.0257	9037	5.1923	5.2768
Platform	Nintendo	5.2534	0.0410	9037	5.1859	5.3209
Puzzle	Nintendo	4.9725	0.0344	9037	4.9159	5.0292
Racing	Nintendo	5.0682	0.0420	9037	4.9990	5.1373
Role-Playing	Nintendo	5.2824	0.0428	9037	5.2121	5.3527
Shooter	Nintendo	5.0882	0.0499	9037	5.0062	5.1703
Simulation	Nintendo	5.2112	0.0303	9037	5.1613	5.2611
Sports	Nintendo	5.2498	0.0289	9037	5.2023	5.2973
Strategy	Nintendo	5.1022	0.0657	9037	4.9941	5.2102
Action	Others	4.9870	0.0624	9037	4.8843	5.0897
Action-Adventure	Others	4.5923	0.2011	9037	4.2614	4.9231
Adventure	Others	4.9526	0.1095	9037	4.7725	5.1326
Fighting	Others	4.7264	0.1642	9037	4.4562	4.9965
Misc	Others	4.9807	0.0813	9037	4.8471	5.1144
Platform	Others	4.9966	0.0717	9037	4.8787	5.1144
Puzzle	Others	4.8163	0.1305	9037	4.6016	5.0309
Racing	Others	4.8391	0.1056	9037	4.6654	5.0129
Role-Playing	Others	5.0112	0.0923	9037	4.8595	5.1630
Shooter	Others	4.9546	0.1642	9037	4.6845	5.2247
Simulation	Others	4.8766	0.1422	9037	4.6427	5.1106
Sports	Others	4.9718	0.1186	9037	4.7767	5.1669
Strategy	Others	4.8450	0.1715	9037	4.5628	5.1271
Action	PC	4.6894	0.0685	9037	4.5768	4.8021
Action-Adventure	PC	5.0539	0.1799	9037	4.7580	5.3498
Adventure	PC	4.4642	0.0821	9037	4.3292	4.5993
Fighting	PC	4.6034	0.2544	9037	4.1850	5.0219
Misc	PC	4.4259	0.1469	9037	4.1843	4.6675
Platform	PC	4.4602	0.1642	9037	4.1901	4.7303
Puzzle	PC	4.3213	0.1380	9037	4.0944	4.5482
Racing	PC	4.6007	0.0830	9037	4.4642	4.7372
Role-Playing	PC	4.8432	0.0690	9037	4.7297	4.9566
Shooter	PC	4.9262	0.0537	9037	4.8378	5.0146
Simulation	PC	4.8908	0.0628	9037	4.7875	4.9941
Sports	PC	4.7080	0.0867	9037	4.5653	4.8507
Strategy	PC	4.6484	0.0550	9037	4.5579	4.7388
Action	Playstation	5.4160	0.0227	9037	5.3786	5.4534
Action-Adventure	Playstation	5.5545	0.0621	9037	5.4525	5.6566
Adventure	Playstation	5.3073	0.0418	9037	5.2385	5.3761
Fighting	Playstation	5.4250	0.0419	9037	5.3560	5.4940

Table 5: Pairwise Comparisons of Genre-Tukey 90% Adjustment

Comparison	Platform type	Estimate	SE	DF	t Statistic	p-value
Action - (Action-Adventure)	Nintendo	-0.0161	0.0769	9037	-0.2098	1.0000
Action - Adventure	Nintendo	0.0233	0.0415	9037	0.5630	1.0000
Action - Fighting	Nintendo	0.0530	0.0700	9037	0.7578	0.9999
Action - Misc	Nintendo	-0.0023	0.0368	9037	-0.0613	1.0000
Action - Platform	Nintendo	-0.0211	0.0488	9037	-0.4329	1.0000
Action - Puzzle	Nintendo	0.2598	0.0434	9037	5.9918	0.0000
Action - Racing	Nintendo	0.1641	0.0496	9037	3.3077	0.0509
Action - (Role-Playing)	Nintendo	-0.0501	0.0502	9037	-0.9980	0.9988
Action - Shooter	Nintendo	0.1441	0.0564	9037	2.5534	0.3393
Action - Simulation	Nintendo	0.0211	0.0402	9037	0.5249	1.0000
Action - Sports	Nintendo	-0.0176	0.0391	9037	-0.4489	1.0000
Action - Strategy	Nintendo	0.1301	0.0708	9037	1.8386	0.8303
(Action-Adventure) - Adventure	Nintendo	0.0395	0.0790	9037	0.4996	1.0000
(Action-Adventure) - Fighting	Nintendo	0.0692	0.0971	9037	0.7126	1.0000
(Action-Adventure) - Misc	Nintendo	0.0139	0.0767	9037	0.1810	1.0000
(Action-Adventure) - Platform	Nintendo	-0.0050	0.0831	9037	-0.0600	1.0000
(Action-Adventure) - Puzzle	Nintendo	0.2759	0.0800	9037	3.4477	0.0325
(Action-Adventure) - Racing	Nintendo	0.1803	0.0836	9037	2.1567	0.6231
(Action-Adventure) - (Role-Playing)	Nintendo	-0.0340	0.0839	9037	-0.4048	1.0000
(Action-Adventure) - Shooter	Nintendo	0.1602	0.0878	9037	1.8247	0.8376
(Action-Adventure) - Simulation	Nintendo	0.0372	0.0783	9037	0.4751	1.0000
(Action-Adventure) - Sports	Nintendo	-0.0014	0.0778	9037	-0.0182	1.0000
(Action-Adventure) - Strategy	Nintendo	0.1462	0.0976	9037	1.4979	0.9569
Adventure - Fighting	Nintendo	0.0297	0.0723	9037	0.4107	1.0000
Adventure - Misc	Nintendo	-0.0256	0.0410	9037	-0.6239	1.0000
Adventure - Platform	Nintendo	-0.0445	0.0520	9037	-0.8541	0.9998
Adventure - Puzzle	Nintendo	0.2364	0.0470	9037	5.0303	0.0000
Adventure - Racing	Nintendo	0.1408	0.0528	9037	2.6647	0.2721
Adventure - (Role-Playing)	Nintendo	-0.0735	0.0534	9037	-1.3755	0.9779
Adventure - Shooter	Nintendo	0.1207	0.0593	9037	2.0369	0.7083
Adventure - Simulation	Nintendo	-0.0023	0.0441	9037	-0.0511	1.0000
Adventure - Sports	Nintendo	-0.0409	0.0431	9037	-0.9486	0.9993
Adventure - Strategy	Nintendo	0.1068	0.0731	9037	1.4615	0.9643
Fighting - Misc	Nintendo	-0.0553	0.0697	9037	-0.7929	0.9999
Fighting - Platform	Nintendo	-0.0741	0.0767	9037	-0.9663	0.9991
Fighting - Puzzle	Nintendo	0.2067	0.0734	9037	2.8167	0.1944
Fighting - Racing	Nintendo	0.1111	0.0773	9037	1.4380	0.9685
Fighting - (Role-Playing)	Nintendo	-0.1031	0.0777	9037	-1.3283	0.9834
Fighting - Shooter	Nintendo	0.0910	0.0818	9037	1.1129	0.9966
Fighting - Simulation	Nintendo	-0.0319	0.0716	9037	-0.4464	1.0000
Fighting - Sports	Nintendo	-0.0706	0.0710	9037	-0.9946	0.9988
Fighting - Strategy	Nintendo	0.0771	0.0923	9037	0.8353	0.9998
Misc - Platform	Nintendo	-0.0189	0.0484	9037	-0.3895	1.0000
Misc - Puzzle	Nintendo	0.2620	0.0429	9037	6.1015	0.0000
Misc - Racing	Nintendo	0.1664	0.0493	9037	3.3775	0.0409

Table 6: Pairwise Comparisons of Platform-Tukey 90% Adjustment

Comparison	Genre	Estimate	SE	DF	t Statistic	p-value
Action - (Action-Adventure)	Nintendo	-0.0161	0.0769	9037	-0.2098	1.0000
Action - Adventure	Nintendo	0.0233	0.0415	9037	0.5630	1.0000
Action - Fighting	Nintendo	0.0530	0.0700	9037	0.7578	0.9999
Action - Misc	Nintendo	-0.0023	0.0368	9037	-0.0613	1.0000
Action - Platform	Nintendo	-0.0211	0.0488	9037	-0.4329	1.0000
Action - Puzzle	Nintendo	0.2598	0.0434	9037	5.9918	0.0000
Action - Racing	Nintendo	0.1641	0.0496	9037	3.3077	0.0509
Action - (Role-Playing)	Nintendo	-0.0501	0.0502	9037	-0.9980	0.9988
Action - Shooter	Nintendo	0.1441	0.0564	9037	2.5534	0.3393
Action - Simulation	Nintendo	0.0211	0.0402	9037	0.5249	1.0000
Action - Sports	Nintendo	-0.0176	0.0391	9037	-0.4489	1.0000
Action - Strategy	Nintendo	0.1301	0.0708	9037	1.8386	0.8303
(Action-Adventure) - Adventure	Nintendo	0.0395	0.0790	9037	0.4996	1.0000
(Action-Adventure) - Fighting	Nintendo	0.0692	0.0971	9037	0.7126	1.0000
(Action-Adventure) - Misc	Nintendo	0.0139	0.0767	9037	0.1810	1.0000
(Action-Adventure) - Platform	Nintendo	-0.0050	0.0831	9037	-0.0600	1.0000
(Action-Adventure) - Puzzle	Nintendo	0.2759	0.0800	9037	3.4477	0.0325
(Action-Adventure) - Racing	Nintendo	0.1803	0.0836	9037	2.1567	0.6231
(Action-Adventure) - (Role-Playing)	Nintendo	-0.0340	0.0839	9037	-0.4048	1.0000
(Action-Adventure) - Shooter	Nintendo	0.1602	0.0878	9037	1.8247	0.8376
(Action-Adventure) - Simulation	Nintendo	0.0372	0.0783	9037	0.4751	1.0000
(Action-Adventure) - Sports	Nintendo	-0.0014	0.0778	9037	-0.0182	1.0000
(Action-Adventure) - Strategy	Nintendo	0.1462	0.0976	9037	1.4979	0.9569
Adventure - Fighting	Nintendo	0.0297	0.0723	9037	0.4107	1.0000
Adventure - Misc	Nintendo	-0.0256	0.0410	9037	-0.6239	1.0000
Adventure - Platform	Nintendo	-0.0445	0.0520	9037	-0.8541	0.9998
Adventure - Puzzle	Nintendo	0.2364	0.0470	9037	5.0303	0.0000
Adventure - Racing	Nintendo	0.1408	0.0528	9037	2.6647	0.2721
Adventure - (Role-Playing)	Nintendo	-0.0735	0.0534	9037	-1.3755	0.9779
Adventure - Shooter	Nintendo	0.1207	0.0593	9037	2.0369	0.7083
Adventure - Simulation	Nintendo	-0.0023	0.0441	9037	-0.0511	1.0000
Adventure - Sports	Nintendo	-0.0409	0.0431	9037	-0.9486	0.9993
Adventure - Strategy	Nintendo	0.1068	0.0731	9037	1.4615	0.9643
Fighting - Misc	Nintendo	-0.0553	0.0697	9037	-0.7929	0.9999
Fighting - Platform	Nintendo	-0.0741	0.0767	9037	-0.9663	0.9991
Fighting - Puzzle	Nintendo	0.2067	0.0734	9037	2.8167	0.1944
Fighting - Racing	Nintendo	0.1111	0.0773	9037	1.4380	0.9685
Fighting - (Role-Playing)	Nintendo	-0.1031	0.0777	9037	-1.3283	0.9834
Fighting - Shooter	Nintendo	0.0910	0.0818	9037	1.1129	0.9966
Fighting - Simulation	Nintendo	-0.0319	0.0716	9037	-0.4464	1.0000
Fighting - Sports	Nintendo	-0.0706	0.0710	9037	-0.9946	0.9988
Fighting - Strategy	Nintendo	0.0771	0.0923	9037	0.8353	0.9998
Misc - Platform	Nintendo	-0.0189	0.0484	9037	-0.3895	1.0000
Misc - Puzzle	Nintendo	0.2620	0.0429	9037	6.1015	0.0000
Misc - Racing	Nintendo	0.1664	0.0493	9037	3.3775	0.0409

From the Marginal Means Table, we can see that on the whole using Xbox as game platform tends to lead to more sales of video games and using PC as game platform tends to lead to lesser sales when compared to the other platforms in this sample.

As mentioned earlier that our overall Unusualness Threshold was 0.05. Since this is an enormous dataset, we decide to choose some comparisons from the Tukey HSD result to discuss. The pairwise comparison of Adventure-Puzzle is statistically significant with p-value closely 0 for all platform types except Others. No pairwise comparisons are statistically significant under the platform type of Others. Adventure – Shooter comparison under PC platform is statistically significant with p-value of 0.0002, as well as Puzzle-Shooter with p-value of 0.003. For PlayStation, Action-Puzzle, Puzzle-Shooter, and adventure-shooter with p-value closely 0 are statistically significant. Shooter-Simulation on Xbox platform is statistically significant with p-value closely 0.

Effect Sizes

Table 7: Effect Sizes for Genre

Comparison	Platform type	Cohen's d	Probability of Superiority
(Action - (Action-Adventure))	Nintendo	-0.028	0.492
(Action - Adventure)	Nintendo	0.041	0.512
(Action - Fighting)	Nintendo	0.093	0.526
(Action - Misc)	Nintendo	-0.004	0.499
(Action - Platform)	Nintendo	-0.037	0.490
(Action - Puzzle)	Nintendo	0.457	0.627
(Action - Racing)	Nintendo	0.289	0.581
(Action - (Role-Playing))	Nintendo	-0.088	0.475
(Action - Shooter)	Nintendo	0.253	0.571
(Action - Simulation)	Nintendo	0.037	0.510
(Action - Sports)	Nintendo	-0.031	0.491
(Action - Strategy)	Nintendo	0.229	0.564
((Action-Adventure) - Adventure)	Nintendo	0.069	0.520
((Action-Adventure) - Fighting)	Nintendo	0.122	0.534
((Action-Adventure) - Misc)	Nintendo	0.024	0.507
((Action-Adventure) - Platform)	Nintendo	-0.009	0.498
((Action-Adventure) - Puzzle)	Nintendo	0.485	0.634
((Action-Adventure) - Racing)	Nintendo	0.317	0.589
((Action-Adventure) - (Role-Playing))	Nintendo	-0.060	0.483
((Action-Adventure) - Shooter)	Nintendo	0.282	0.579
((Action-Adventure) - Simulation)	Nintendo	0.065	0.518
((Action-Adventure) - Sports)	Nintendo	-0.002	0.499
((Action-Adventure) - Strategy)	Nintendo	0.257	0.572
(Adventure - Fighting)	Nintendo	0.052	0.515
(Adventure - Misc)	Nintendo	-0.045	0.487
(Adventure - Platform)	Nintendo	-0.078	0.478
(Adventure - Puzzle)	Nintendo	0.416	0.616
(Adventure - Racing)	Nintendo	0.248	0.569
(Adventure - (Role-Playing))	Nintendo	-0.129	0.464
(Adventure - Shooter)	Nintendo	0.212	0.560
(Adventure - Simulation)	Nintendo	-0.004	0.499
(Adventure - Sports)	Nintendo	-0.072	0.480
(Adventure - Strategy)	Nintendo	0.188	0.553
(Fighting - Misc)	Nintendo	-0.097	0.473
(Fighting - Platform)	Nintendo	-0.130	0.463
(Fighting - Puzzle)	Nintendo	0.363	0.601
(Fighting - Racing)	Nintendo	0.195	0.555
(Fighting - (Role-Playing))	Nintendo	-0.181	0.449
(Fighting - Shooter)	Nintendo	0.160	0.545
(Fighting - Simulation)	Nintendo	-0.056	0.484
(Fighting - Sports)	Nintendo	-0.124	0.465
(Fighting - Strategy)	Nintendo	0.136	0.538
(Misc - Platform)	Nintendo	-0.033	0.491

Table 8: Effect Sizes for Platform type

Comparison	Genre	Cohen's d	Probability of Superiority
(Nintendo - Others)	Action	0.431	0.620
(Nintendo - PC)	Action	0.954	0.750
(Nintendo - Playstation)	Action	-0.323	0.410
(Nintendo - Xbox)	Action	-0.267	0.425
(Others - PC)	Action	0.523	0.644
(Others - Playstation)	Action	-0.754	0.297
(Others - Xbox)	Action	-0.699	0.311
(PC - Playstation)	Action	-1.277	0.183
(PC - Xbox)	Action	-1.222	0.194
(Playstation - Xbox)	Action	0.056	0.516
(Nintendo - Others)	Action-Adventure	1.154	0.793
(Nintendo - PC)	Action-Adventure	0.342	0.596
(Nintendo - Playstation)	Action-Adventure	-0.538	0.352
(Nintendo - Xbox)	Action-Adventure	-0.481	0.367
(Others - PC)	Action-Adventure	-0.812	0.283
(Others - Playstation)	Action-Adventure	-1.692	0.116
(Others - Xbox)	Action-Adventure	-1.635	0.124
(PC - Playstation)	Action-Adventure	-0.880	0.267
(PC - Xbox)	Action-Adventure	-0.823	0.280
(Playstation - Xbox)	Action-Adventure	0.057	0.516
(Nintendo - Others)	Adventure	0.451	0.625
(Nintendo - PC)	Adventure	1.309	0.823
(Nintendo - Playstation)	Adventure	-0.173	0.451
(Nintendo - Xbox)	Adventure	0.162	0.546
(Others - PC)	Adventure	0.859	0.728
(Others - Playstation)	Adventure	-0.624	0.330
(Others - Xbox)	Adventure	-0.288	0.419
(PC - Playstation)	Adventure	-1.482	0.147
(PC - Xbox)	Adventure	-1.147	0.209
(Playstation - Xbox)	Adventure	0.335	0.594
(Nintendo - Others)	Fighting	0.796	0.713
(Nintendo - PC)	Fighting	1.012	0.763
(Nintendo - Playstation)	Fighting	-0.432	0.380
(Nintendo - Xbox)	Fighting	-0.403	0.388
(Others - PC)	Fighting	0.216	0.561
(Others - Playstation)	Fighting	-1.228	0.193
(Others - Xbox)	Fighting	-1.199	0.198
(PC - Playstation)	Fighting	-1.444	0.154
(PC - Xbox)	Fighting	-1.416	0.158
(Playstation - Xbox)	Fighting	0.029	0.508
(Nintendo - Others)	Misc	0.446	0.624
(Nintendo - PC)	Misc	1.422	0.843
(Nintendo - Playstation)	Misc	-0.248	0.430
(Nintendo - Xbox)	Misc	-0.168	0.453
(Others - PC)	Misc	0.975	0.755

Suppose we focus on the following genre comparisons. For Adventure vs Puzzle, the estimated location parameter for Adventure Game is 0.416 pooled standard deviations above Puzzle game. For adventure vs shooter, the estimated location parameter for adventure game is 0.812 pooled standard deviation below Shooter game. Further, the probability of superiority for each pairing in the model is relatively close to 0.5, which we can say there is practical effect. As a result, I would say there is enough variation in those groups that the effect is statistically large enough to escape through the noise of the group. The same apply to the pairwise comparison of platform that the probability of superiority are quite close to 0.5 that apply practical effect.

Discussion and Limitations

We investigated our research question of how genre and platform affect video game sales. From our data, we found that genre, platform, and the interaction between genre and platform affect the sales of all video games (specifically the log of the sales in terms of unit sold).

Our study has several limitations. Firstly, we could not take into account the impact of price on sales. We should set a price range, and then compare sales within the same price range by genre and platform. This will make our results more accurate with less bias. If the price of a video game is much higher than the market average price, then regardless of the platform and type, the sales of the game will be relatively low.

Secondly, we need to consider the quality of the game by adding the rating variable. Selecting video games in the rating range and build a model can reduce the impact of high-quality or low-quality games on sales. If a video game's rating is lower than one out of five points, this video game must have some major flaws, such as bugs that cannot be cleared, etc. Remove extreme ratings will reduce the impact of changes in sales due to game quality.

Furthermore, since this was not an experiment, our conclusions cannot be extrapolated to the population. Therefore, there is potential for more studies on this subject that could investigate even more potential factors as stated earlier. This is analysis was in determining potentially significant factors in the sales of video games.

Author Contribution

The authors of this report would like to acknowledge their individual contributions to the report.

- Zining Yin contributed to writing the executive summary and wrote the introduction.
- Ben contributed writing the code for all the models/visualizations. Additionally he explored and cleaned the data and wrote down the findings.
- L contributed to the Literature Review, references, and assessing the assumptions of ANOVA.
- Yurui Niu contributed to the description of the study and wrote down the final discussion and limitations.
- Luciana Lai contributed assessing the results of the ANOVA omnibus, post hoc, point estimates, and effect sizes.

References

Davidovici-Nora, M., & Bourreau, M. (2012). Two-sided markets in the video game industry. Réseaux, 173-174(3), 97–135.

Bellovich, Matthew, "The Effects of Cross-Platform Gaming and Multihoming in a Two-Sided Market: A

Focus on the Sixth, Seventh and Eighth Generation of the Video Game Industry" (2020). Senior Independent Study Theses. Paper 9001.

Landsman, V., & Stremersch, S. (2011). Multihoming in two-sided markets: An empirical inquiry in the video game console industry. Journal of Marketing, 75(6), 39–54. https://doi.org/10.1509/jm.09.0199

J. Strube, S. Schade, P. Schmidt and P. Buxmann, "Simulating Indirect Network Effects in the Video Game Market," 2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07), 2007, pp. 160b-160b, doi: 10.1109/HICSS.2007.483

Code Appendix

```
# Setting Document Options
knitr::opts_chunk$set(
  echo = FALSE,
  warning = FALSE,
  message = FALSE,
  fig.align = "center"
)

# Add additional packages by name to the following list
packages <- c(
  "tidyverse", "knitr", "kableExtra",
  "psych", "car", "parameters", "scales",
  "hasseDiagram", "DescTools", "emmeans"
  )
lapply(X = packages, FUN = library, character.only = TRUE)

# Loading Helper Files and Setting Global Options
options(knitr.kable.NA = "")
options("contrasts" = c("contr.sum", "contr.poly"))
source("https://raw.github.com/neilhatfield/STAT461/master/rScripts/ANOVATools.R")

source("https://raw.github.com/neilhatfield/STAT461/master/rScripts/shadowgram.R")

library(tinytex)

# Hasse Diagram

modelLabels <- c("1 Sell Video Game 1", "13 Genre 12", "5 Video Game Platform Type 4", "65 Genre × Vide
modelMatrix <- matrix(
  data = c(FALSE, FALSE, FALSE, FALSE, FALSE, TRUE, FALSE, FALSE, FALSE, FALSE, TRUE, FALSE, FALSE, FALS
  nrow = 5,
  ncol = 5,
  byrow = FALSE
)
hasseDiagram::hasse(
 data = modelMatrix,
 labels = modelLabels
)

# FROM FACTORIAL DESIGNS HANDOUT *
# 1. Load and clean data

vgsales.raw <- read.csv("vgsales2019update.csv")

# remove rows with incomplete information
vgsales <- vgsales.raw[which(vgsales.raw$Platform > " " & vgsales.raw$Global_Sales > 0 & vgsales.raw$ESI

# Check variable to study: Global_Sales by Year
Year <- vgsales$Year
counts <- data.frame(table(Year))
# create a temp data frame including only Global_Sales summarized by Year
```

```r
p <- vgsales %>%
  dplyr::select(Year, Global_Sales) %>%
  group_by(Year) %>%
  summarise(Sales = sum(Global_Sales))
q <- cbind.data.frame(p, counts[2]) # Add counts to data frame
names(q)[3] <- "count"
q$count <- as.numeric(q$count)
# create histogram
ggplot(q, aes(x = Year, y = Sales, label = count)) +
  geom_col(fill = "steelblue1") +
  theme(axis.text.x = element_text(angle = 90),
  panel.grid.major = element_blank(),
  panel.grid.minor = element_blank())
# The histogram shows an outlier: an observation from year 1970. Probably a mistake, will remove.

# Remove years with far away observations
vgsales <- vgsales[which(vgsales$Year > 1970),]


# CREATING PLATFORM TYPE FACTOR
# create platform type to group families of platforms
pc <- c("PC")
xbox <- c("X360", "XB", "XOne")
nintendo <- c("Wii", "WiiU", "N64", "GC", "NES", "3DS", "DS")
playstation <- c("PS", "PS2", "PS3", "PS4", "PSP", "PSV")
vgsales <- vgsales %>%
  mutate(Platform.type =  ifelse(Platform %in% pc, "PC",
                          ifelse(Platform %in% xbox, "Xbox",
                          ifelse(Platform %in% nintendo, "Nintendo",
                          ifelse(Platform %in% playstation, "Playstation", "Others")))))

# create plot to see how yearly sales are distributed among platform types
p <- vgsales %>%
  group_by(Platform.type, Year) %>%
    summarise(total = sum(Global_Sales))
p$YearR <- as.numeric(as.character(p$Year))
ggplot2::ggplot(p, ggplot2::aes(x = YearR, fill = Platform.type)) +
  ggplot2::geom_density(position = "fill") +
  labs(x = "Year", y = "% Global Sales") +
  scale_y_continuous(labels = percent_format())
# we observe great instability in the market share among platforms in the early years. This could be pl

# we will consider data from 2004-2019 when the market share pattern seem stable.
vgsales <- vgsales[which(vgsales$Year > 2004 & vgsales$Year < 2019),]
p <- vgsales %>%
  group_by(Platform.type, Year) %>%
  summarise(total = sum(Global_Sales))
p$YearR <- as.numeric(as.character(p$Year))
ggplot2::ggplot(p, ggplot2::aes(x = YearR, fill = Platform.type)) +
  ggplot2::geom_density(position = "fill") +
  labs(x = "Year", y = "% Global Sales") +
  scale_y_continuous(labels = percent_format())
```

```r
# now we observe the distribution of the Global Sales in the filtered data frame
# histogram of Global Sales

ggplot(vgsales, aes(x=Global_Sales)) +
  geom_histogram(binwidth = 0.25, color="darkblue", fill="lightblue")




# Global sales seem to follow an exponential distribution. We apply a log transformation to global sales
vgsales$global.log <- log10((vgsales$Global_Sales*1000000))

# Now we observe the distribution of the transformed Global sales considering 2 factors of interest (bo
ggplot2::ggplot(vgsales, ggplot2::aes(x = Genre, y = global.log, fill=Genre)) +
  ggplot2::geom_boxplot() + theme(axis.text.x = element_text(angle = 90)) +
  guides(fill="none") + labs(y = "Log Global Sales")
ggplot2::ggplot(vgsales, ggplot2::aes(x = Platform.type, y = global.log, fill=Platform.type)) +
  ggplot2::geom_boxplot() + guides(fill="none") + labs(y = "Log Global Sales")

# First we need to remove rare genre types (very few observations on them) WE CREATE A NEW CLEAN DATA F
vgsales.clean <- vgsales[which(vgsales$Genre != "Visual Novel" & vgsales$Genre != "Music" ),]

#factors
vgsales.clean$Platform.type <- factor (x=vgsales.clean$Platform.type)
vgsales.clean$Genre <- factor( x=vgsales.clean$Genre)

# BOX PLOTS revisited

# professional looking boxplot depicting combined factors
ggplot(
  data = vgsales.clean,
  mapping = aes(
    x = Genre,
    y = global.log,
    fill = Platform.type
  )
) +
  geom_boxplot() +
  theme_bw() +
  xlab("Genre") +
  ylab("Log Global Sales (units)") +
  labs(
    fill = "Platform Type"
  ) +
  theme(
    legend.position = "bottom",
    text = element_text(size =12),
    axis.text.x = element_text(angle = 90)
  )

# DESCRIPTIVE STATISTICS

vgsalesStats <- psych::describeBy(
```

```
  x = vgsales.clean$global.log,
  group = paste(vgsales.clean$Genre, vgsales.clean$Platform.type, sep = " x "),
  na.rm = TRUE,
  skew = TRUE,
  ranges = TRUE,
  quant = c(0.25, 0.75),
  IQR = TRUE,
  mat = TRUE,
  digits = 4
)

vgsalesStats %>%
  tibble::remove_rownames() %>%
  tibble::column_to_rownames(
    var = "group1"
  ) %>%
  dplyr::select(
    n, min, Q0.25, median, Q0.75, max, mad, mean, sd, skew, kurtosis
  ) %>%
  knitr::kable(
    caption = "Summary Statistics for Video Games Log(Global Sales)",
    digits = 3,
    format.args = list(big.mark = ","),
    align = rep('c', 11),
    col.names = c("n", "Min", "Q1", "Median", "Q3", "Max", "MAD", "SAM", "SASD", "Sample Skew", "Sample
    booktabs = TRUE
  ) %>%
  kableExtra::kable_styling(
    font_size = 12,
    latex_options = c("Hold_position", "scale_down")
  )

#check interactions

interaction.plot(
  x.factor = vgsales.clean$Genre,
  trace.factor = vgsales.clean$Platform.type,
  response = vgsales.clean$global.log,
  fun = mean,
  type = "b",
  col = c("black","red","blue","#009999","maroon"),
  pch = c(19,18,17,16,15),
  fixed = TRUE,
  legend = TRUE,
  xlab = "Genre",
  ylab = "log Global Sales (units)",
  las = 2,
  trace.label = "Platform Type"
)

vgsalesModel <- aov(
  formula = global.log ~ Genre*Platform.type,
  data = vgsales.clean,
```

```
  na.action = "na.omit"
)

#assessing the Gaussian Assumption
#assumptionPlots, fig.cap="Assessing Gaussian Assumptions for Video Game Sales Study", fig.subcap=c("QQ

#qqplot for residuals
car::qqPlot(
  x = vgsalesModel$residuals,
  distribution = "norm",
  envelope = 0.90,
  id = FALSE,
  pch = 20,
  ylab = "Residuals"
)

## Homoscedasticity
ggplot(
  data = data.frame(
    residuals = residuals(vgsalesModel),
    fitted = fitted(vgsalesModel)
  ),
  mapping = aes(x = fitted, y = residuals) )+
  geom_point(size = 2) +
  geom_hline(
    yintercept = 0,
    linetype = "dashed",
    color = "grey50"
  ) +
  geom_smooth(
    formula = y ~ x,
    method = loess,
    method.args = list(degree = 1),
    se = FALSE,
    size = 0.5
  ) +
  theme_bw() +
  xlab("Fitted values (units)") +
  ylab("Residuals (units)")



parameters::model_parameters(
  model = vgsalesModel,
  omega_squared = "partial",
  eta_squared = "partial",
  epsilon_squared = "partial",
  type = 2,       # Type II SSQs
  drop = "(Intercept)",
  verbose = FALSE
) %>%
  dplyr::mutate(
    p = ifelse(
```

```
      test = is.na(p),
      yes = NA,
      no = pvalRound(p)
    )
  ) %>%
  knitr::kable(
  digits = 4,
  col.names = c(
    "Source", "SS", "df", "MS", "F", "p-value",
    "Partial Omega Sq.", "Partial Eta Sq.", "Partial Epsilon Sq."),
  caption = "ANOVA Table for Video Games Sales - Type II SSQs",
  booktabs = TRUE,
  align = c("l", rep("c", 8))
  ) %>%
  kableExtra::kable_styling(
    bootstrap_options = c("stripped", "condensed"),
    font_size = 12,
    latex_options = c("scale_down")
  )


pointEst <- dummy.coef(vgsalesModel)
pointEst <- unlist(pointEst)
names(pointEst) <- c(
  "Grand Mean",
  levels(vgsales.clean$Genre),
  levels(vgsales.clean$Platform.type),
  outer(
    levels(vgsales.clean$Genre),
    levels(vgsales.clean$Platform.type),
    FUN = paste,
    sep = " x "
  )
)
data.frame("Estimate" = pointEst) %>%
  knitr::kable(
    digits = 2,
    caption = "Point Estimates from the Video Games Sales study",
    booktabs = TRUE,
    align = "c"
  ) %>%
  kableExtra::kable_styling(
    bootstrap_options = c("stripped", "condensed"),
    font_size = 12,
    latex_options = c("HOLD_position")
  )

vgsalesMeans <- emmeans::emmeans(
  object = vgsalesModel,
  specs = pairwise ~ Genre | Platform.type,
  adjust = "tukey", # chosen method
  level = 0.9  # type I risk
)
```

```r
as.data.frame(vgsalesMeans$emmeans) %>%
  knitr::kable(
    digits = 4,
    col.names = c("Genre", "Platform type", "Marginal Mean", "SE", "DF", "Lower Bound", "Upper Bound"),
    caption = "Marginal Means-Tukey 90\\% Adjustment",
    align = rep("c",7),
    booktabs = TRUE
  ) %>%
  kableExtra::kable_styling(
    bootstrap_options = c("stripped", "condensed"),
    font_size = 12,
    latex_options = c("HOLD_position")
  )

vgsalesGenre <- emmeans::emmeans(
  object = vgsalesModel,
  specs = pairwise ~ Genre | Platform.type,
  adjust = "tukey", # chosen method
  level = 0.9  # type I risk
)

as.data.frame(vgsalesMeans$contrasts ) %>%
  knitr::kable(
    digits = 4,
    col.names = c("Comparison", "Platform type", "Estimate", "SE", "DF", "t Statistic", "p-value"),
    caption = "Pairwise Comparisons of Genre-Tukey 90\\% Adjustment",
    align = rep("c",7),
    booktabs = TRUE
  ) %>%
  kableExtra::kable_styling(
    bootstrap_options = c("stripped", "condensed"),
    font_size = 12,
    latex_options = c("HOLD_position")
  )

vgsalesPlatform <- emmeans::emmeans(
  object = vgsalesModel,
  specs = pairwise ~ Platform.type | Genre,
  adjust = "tukey", # chosen method
  level = 0.9  # type I risk
)

as.data.frame(vgsalesMeans$contrasts ) %>%
  knitr::kable(
    digits = 4,
    col.names = c("Comparison", "Genre", "Estimate", "SE", "DF", "t Statistic", "p-value"),
    caption = "Pairwise Comparisons of Platform-Tukey 90\\% Adjustment",
    align = rep("c",7),
    booktabs = TRUE
  ) %>%
  kableExtra::kable_styling(
    bootstrap_options = c("stripped", "condensed"),
    font_size = 12,
```

```
    latex_options = c("HOLD_position")
  )


as.data.frame(
  eff_size(
    object = vgsalesGenre,
    sigma = sigma(vgsalesModel),
    edf = df.residual(vgsalesModel)
  )
) %>%
  dplyr::mutate(
    ps = probSup(effect.size),
    .after = effect.size
  ) %>%
  dplyr::select(contrast, Platform.type, effect.size, ps) %>%
  knitr::kable(
    digits = 3,
    col.names = c("Comparison", "Platform type", "Cohen's d", "Probability of Superiority"),
    caption = "Effect Sizes for Genre",
    align = "lccc",
    booktabs = TRUE
  ) %>%
  kableExtra::kable_styling(
    bootstrap_options = c("stripped", "condensed"),
    font_size = 12,
    latex_options = c("HOLD_position")
  )

as.data.frame(
  eff_size(
    object = vgsalesPlatform,
    sigma = sigma(vgsalesModel),
    edf = df.residual(vgsalesModel)
  )
) %>%
  dplyr::mutate(
    ps = probSup(effect.size),
    .after = effect.size
  ) %>%
  dplyr::select(contrast, Genre, effect.size, ps) %>%
  knitr::kable(
    digits = 3,
    col.names = c("Comparison", "Genre", "Cohen's d", "Probability of Superiority"),
    caption = "Effect Sizes for Platform type",
    align = "lccc",
    booktabs = TRUE
  ) %>%
  kableExtra::kable_styling(
    bootstrap_options = c("stripped", "condensed"),
    font_size = 12,
    latex_options = c("HOLD_position")
  )
```