

Performance of Video Games in Different Markets

Introduction

Video games are electronic games that involve interaction between a user interface to generate visual feedback on some sort of display device. Video games started reaching popularity in the 1970's and 1980's. Since then, they have become one of the most popular forms of entertainment around the globe with sales of games reaching millions worldwide. This report uses data that was scrapped from VGChartz.

VGChartz is a video game sales tracking website that provides weekly sales figures of video games based on region. The sales data collected by VGCharts is based on platform of the games release (i.e. PC, PS4, etc.), year the game was released, genre of the game, and publisher of the game. Based on these variables VGChartz provides sales data in the millions for regions in which the games are sold. Included in the data set we use are sales in North America (NA_Sales), sales in Japan (JP_Sales), sales in Europe (EU_Sales), sales in the rest of the world (Other_Sales), and total worldwide sales (Global_Sales).

Our study looks at 2 specific aspects of sales. Firstly, we consider how well different publishers perform in each market. Second, we examine to see how well different genres perform in each market.

Data

Sources

The Data was derived by a scrape of VGChartz. Data was collected on video games with sales greater than 100, 000 copies. Information about all the video games and their resulting sales data are available on the VGChartz website. For all the video games data posted on the VGChartz website users at Kaggle (An online community of data scientists) extracted information about the video games and recorded the data from the VGChartz website and compiled for each video game the publisher, the year, the genre, the platform, and sales data.

My analyses are based on a single dataset created by the users at Kaggle. My analyses only included video games with sales greater than 100, 000 copies.

Variables

The data included information about each video game, including the video games name, the platform on which the game was released, the year the game was released, the genre of the game, the publisher of the game, the rank of the game based on overall sales, and the sales data in millions for North America, Japan, Europe, and all other counties, it also includes the global sales in millions for each video game.

Brief Summary

There were 16,660 video games in our data between the years of 1980 and 2020, of whom 9% were released in 2009 and 9% were released in 2008. We only look at the top 2000 video games with the highest sales. There are 12 genres that a video game can fall under. Table one shows

that data for these 12 genres. The data shows the total sales in North America, Europe, Japan, other regions, and globally based on the genre the video game fall under. Table two shows the sales data for the publishers.

	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
	Sum	Sum	Sum	Sum	Sum
Genre					
Action	514.70	338.85	75.80	122.57	1051.93
Adventure	45.27	30.45	10.18	7.09	92.98
Fighting	141.59	65.73	38.23	23.67	269.21
Misc	233.95	140.48	52.02	45.47	471.85
Platform	342.94	155.16	111.72	37.71	647.58
Puzzle	78.32	32.38	35.93	7.40	154.09
Racing	220.81	157.16	44.30	53.18	475.31
Role-Playing	230.59	140.74	221.09	41.47	633.89
Shooter	436.21	228.65	21.61	76.71	763.18
Simulation	93.10	79.24	33.38	17.87	223.62
Sports	375.20	250.36	63.29	90.77	779.64
Strategy	33.71	19.57	8.98	3.66	65.91

Table 1: Summary of sales data based on genre of video games

	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
	Sum	Sum	Sum	Sum	Sum
Publisher					
3DO	1.10	0.14	0.00	0.04	1.28
505 Games	13.04	8.78	0.31	2.43	24.57
989 Studios	7.29	1.99	0.10	0.43	9.82
ASC Games	0.73	0.50	0.00	0.09	1.31
ASCII Entertainment	0.00	0.00	5.21	0.22	5.42
Acclaim Entertainment	17.87	6.01	0.10	1.17	25.11
Accolade	1.05	0.20	0.00	0.05	1.30
Activision	267.48	158.69	5.91	52.80	484.82
Activision Value	1.58	0.00	0.00	0.12	1.70
Agetec	0.53	0.41	0.00	0.14	1.08
Alchemist	0.59	0.46	0.01	0.15	1.22
Arena Entertainment	3.70	0.88	0.00	0.14	4.72
Atari	57.14	11.61	6.12	3.68	78.59
Atlus	0.46	0.34	0.36	0.21	1.38
Banpresto	0.00	0.00	1.69	0.01	1.70
Bethesda Softworks	31.09	23.74	1.20	8.22	64.25
Black Label Games	0.59	0.46	0.00	0.15	1.20
Capcom	47.55	23.55	42.78	9.10	123.01
City Interactive	0.54	0.51	0.00	0.12	1.16
Codemasters	1.47	11.75	0.47	2.32	16.02
Coleco	1.36	0.08	0.00	0.02	1.46
Compile	0.00	0.00	1.07	0.07	1.15
Crave Entertainment	0.66	0.52	0.15	0.17	1.50
DSPublisher	1.24	1.38	0.99	0.39	3.99
Deep Silver	5.09	2.63	0.25	0.97	8.94
Disney Interactive Stu	31.44	15.45	0.21	5.94	53.08
Eidos Interactive	28.20	22.40	5.15	4.43	60.14
Electronic Arts	361.59	259.98	9.74	92.37	723.77
Empire Interactive	0.56	0.44	0.00	0.15	1.15
Enix Corporation	0.53	0.14	26.20	0.05	26.93
Focus Home Interactive	0.08	1.02	0.00	0.13	1.23
Fox Interactive	2.42	2.37	0.00	0.30	5.10
GSP	0.23	0.74	0.00	0.10	1.06
GT Interactive	6.65	4.49	0.02	0.60	11.76
Global Star	1.26	0.12	0.00	0.21	1.59

Gotham Games	0.98	0.53	0.00	0.06	1.57
GungHo	0.00	0.00	1.51	0.00	1.51
Hasbro Interactive	8.25	2.11	0.00	0.39	10.75
Hello Games	0.58	0.74	0.02	0.26	1.60
Hudson Soft	0.00	0.00	4.80	0.00	4.80
Imagic	4.16	0.25	0.00	0.04	4.45
Infogrames	2.99	0.60	0.05	0.11	3.74
JVC	2.10	0.24	0.00	0.07	2.40
Konami Digital Enterta	40.08	38.07	28.38	19.92	127.27
Level 5	0.00	0.00	6.23	0.00	6.23
LucasArts	38.09	21.22	0.20	8.63	68.14
MTV Games	7.46	2.46	0.01	1.04	10.98
Majesco Entertainment	1.54	1.07	0.00	0.28	2.88
Mastertronic	0.95	0.15	0.00	0.08	1.18
Maxis	1.88	1.30	0.00	0.18	3.36
Microprose	0.18	0.81	0.00	0.08	1.06
Microsoft Game Studios	132.56	56.93	2.24	15.82	207.54
Midway Games	16.76	7.17	0.00	2.53	26.47
Mindscape	3.12	0.41	0.00	0.27	3.80
NIA	3.25	1.89	0.00	0.53	5.67
NCSoft	0.95	1.06	0.00	0.29	2.30
Namco Bandai Games	33.23	21.74	27.38	6.99	89.34
Natsume	0.99	0.13	0.10	0.04	1.25
Nintendo	763.84	397.48	395.86	89.87	1647.14
Ocean	0.74	0.38	0.02	0.03	1.17
Oxygen Interactive	0.82	0.64	0.00	0.22	1.68
Palcom	3.38	0.44	0.31	0.04	4.17
Parker Bros.	2.06	0.12	0.00	0.02	2.20
Play It	0.52	0.40	0.00	0.14	1.06
Pygnosis	0.85	0.44	0.00	0.08	1.17
RTL	0.45	0.68	0.00	0.12	1.26
Red Orb	1.54	3.61	0.00	0.10	5.24
Red Storm Entertainmen	0.93	0.63	0.00	0.08	1.64
RedOctane	6.40	1.28	0.03	0.90	8.61
Rising Star Games	0.79	0.53	0.00	0.13	1.44
Russel	0.00	1.12	0.00	0.00	1.12
Sega	64.58	49.28	10.83	13.37	138.12
Sony Computer Entertai	197.18	142.91	52.16	58.63	450.88
Sony Online Entertainm	0.63	0.49	0.00	0.16	1.28
SouthPeak Games	1.00	0.04	0.00	0.16	1.20
Square	0.24	0.09	1.33	0.12	1.77
Square Enix	28.76	19.62	25.99	9.15	83.73
Squaresoft	9.47	4.07	28.30	1.14	42.96
Success	0.64	0.43	0.00	0.07	1.15
TDK Mediative	1.58	0.85	0.00	0.19	2.62
THQ	83.38	47.74	3.52	15.05	149.66
Taito	0.00	0.00	1.02	0.07	1.09
Take-Two Interactive	161.47	94.06	5.43	45.95	306.85
Tecmo Koel	0.87	0.39	2.04	0.19	3.50
UEP Systems	1.52	0.46	0.20	0.08	2.25
Ubisoft	131.36	96.18	3.71	29.57	260.72
Universal Interactive	4.87	4.92	0.24	1.39	11.41
Unknown	0.60	1.56	0.42	0.49	3.06
Valve	0.63	0.63	0.02	0.25	1.74
Valve Software	1.41	0.51	0.01	0.17	2.10
Video System	0.46	0.96	0.09	0.08	1.60
Virgin Interactive	13.79	9.12	5.30	2.31	30.49
Vivendi Games	15.12	6.76	0.17	1.64	23.68
Warner Bros. Interacti	45.96	27.80	0.55	10.27	84.60
Westwood Studios	1.55	0.00	0.00	0.00	1.55
Zoo Digital Publishing	0.57	0.45	0.00	0.15	1.17
Zoo Games	1.22	0.00	0.00	0.09	1.31

Table 2: Summary of sales data based on publisher

Data

Specific Question 1:

i. **Methods:**

The first question I chose to dig into with our data set is to see how well developers perform in different international markets. I wanted to investigate to see if some of the biggest publishers outperform one another in certain countries and see if there are any indicators as to whether a publisher would perform better in a country more so than another.

The relative parameter in our data is to observe the sales across four different market variables: NA_Sales, representing sales in North America, EU_Sales, representing sales in Europe, JP_Sales, representing sales in Japan, and Other_Sales, representing sales in all other countries and territories. Of course, the Publisher variable will be an important parameter. For the sake of shrinking down our data size from the hundreds of publishers our data has to offer, I decided to pick some of the highest performing publishers to observe across the four different markets. Those publishers are Activision, Atari, Capcom, Nintendo, and Sega.

The analysis I used to conduct this analysis was through the visualization of the data in bar graphs. Here are the procedures I used in SAS and their function in the code:

- PROC SORT
 - Used to sort our data by publisher
- PROC TRANSPOSE
 - Used to manipulate the data in such a way that makes the “region” variable into a categorical variable so it is compatible with PROC SGPLOT to make our bar graph that will be used for analysis
- PROC DELETE
 - Used for data deletion
- PROC SGPLOT
 - Used to create the bar plot used in our analysis of the data

ii. Results:

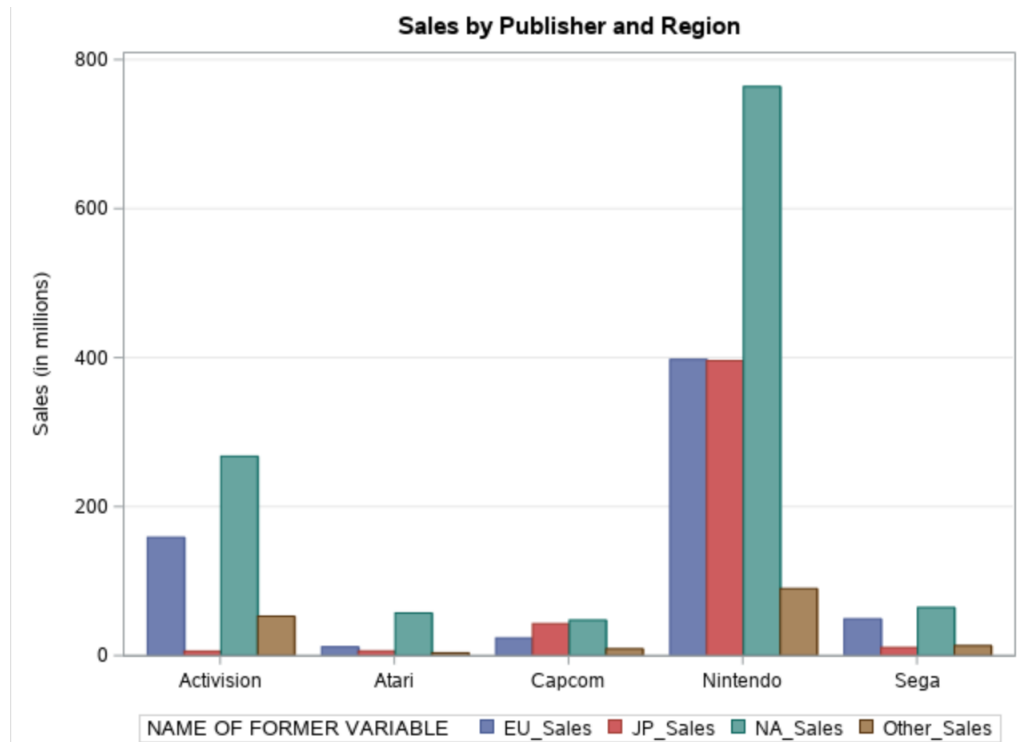


Figure 1: Bar Graph of sales data based for each publisher based on region.

Above is the graphical analysis that we will use to answer the question to my specific question. As stated before, we have the five publishers that we decided to analyze and their respective sales data in the four different markets we have data for. The legend shows correspondence between bars and the market they represent.

One initial conclusion we can draw across all the sales data we have is that North America is the highest selling market across all five of our publishers. North America yields the most sales for all the publishers and by a significant margin for most.

The general trend that seems to follow the largest market (NA) is that Europe is the next highest, then Japan, and finally Other. While this is the most common trend, we notice that This doesn't seem to be the case with Capcom. Similarly, Europe is beating Japan for Nintendo's sales by a very small margin. These are immediate outliers that needed investigating.

One possible explanation for these anomalies is the country of origin of the publisher. Let's observe the country of origin of each of the publishers we chose to observe:

Nintendo	Japan
Capcom	Japan

Atari	North America
Activision	North America
Sega	Japan

We see that the Publishers whose Japan market sales come close to or exceeding Europe both have their origin in Japan, which could explain the higher Japanese market sales relative to those Publishers who do not have their origin in Japan. However, it is important to note that Sega also has its origin in Japan but does not follow a similar pattern of high Japanese sales like that of Nintendo or Capcom.

Some possible problems with our conclusion are that we are using graphical analyses to judge the general patterns across different markets. We would need to understand the general trends across many other publishers to make any assumptions about the general trends of publishers across different markets. For example, we noted that 2 of 3 Japanese publishers have relatively high market sales in Japan. Could the same be said about European developers? Other market developers? Since our analysis has none, it is hard to say.

Specific Question 2:

The second question I chose to look at based on the data was how well the different genres perform in each market. I wanted to see if the action genre (this is considered the most popular genre of video games in the United States) outperformed other genres in different regions of the world. In this section we answer the question: Do certain genres perform better in different markets?

i. Methods:

The relative parameters in our data include the four sales variables: NA_Sales, representing sales in North America, EU_Sales, representing sales in Europe, JP_Sales, representing sales in Japan, and Other_Sales, representing sales in all other countries and territories. The Genre variable is also an important parameter. Since there were only 12 genres there was no need to shrink the data down and look at a few. The 12 different genres are Action, Adventure, Fighting, Platform, Puzzle, Racing, Role Playing, Shooter, Simulation, Sports, Strategy, and Miscellaneous.

The analysis I used to was through visualization of data and ANOVA to look at the means of the sales data of each genre based on region. 4 different ANOVA tables were created to see how sales in genres varied based on region.

The procedure I used for analysis of data:

- PROC SORT

- Used to sort our data by publisher
- PROC TRANSPOSE
 - Used to manipulate the data in such a way that makes the “region” variable into a categorical variable so it is compatible with PROC SGPLOT to make our bar graph that will be used for analysis
- PROC DELETE
 - Used for data deletion
- PROC SGPLOT
 - Used to create the bar plot used in our analysis of the data
- PROC ANOVA
 - Analysis of variance for the balanced data (balanced data is data that has the same number of observations)

ii. Results:

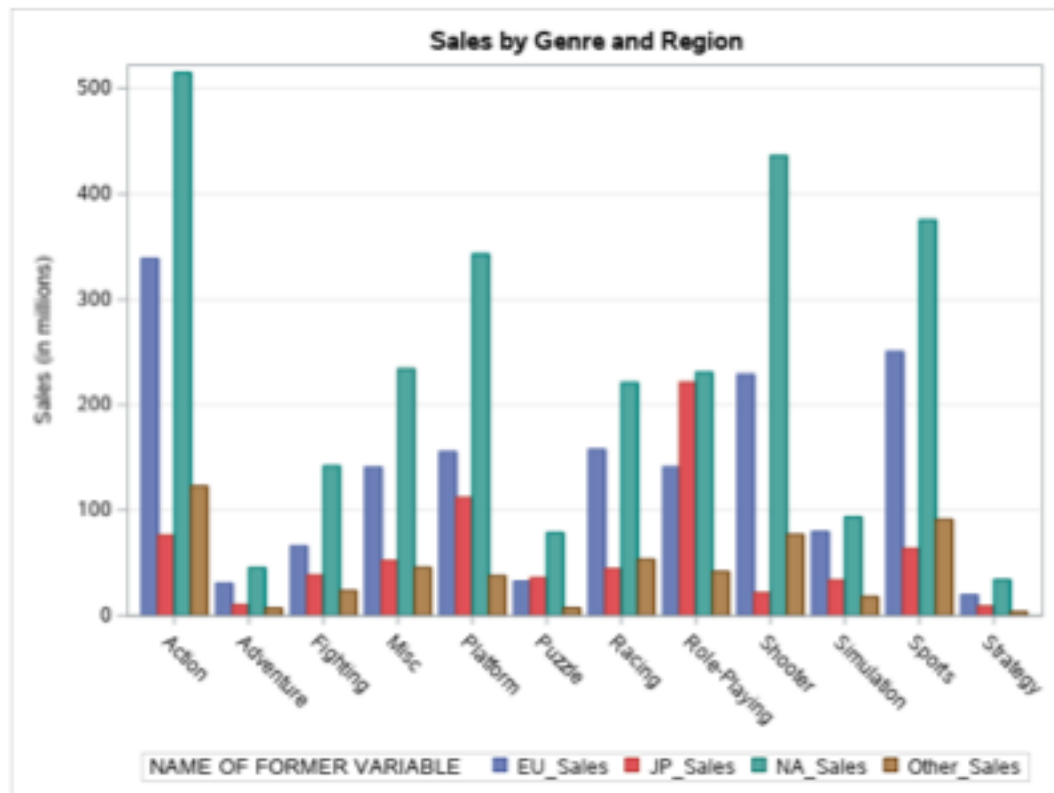


Figure 2: Bar graph of sales data for each region based on genre

My question examines the association between 2 categorical variables (Genre and Region) to a nominal variable (Sales). The Figure 2 conveys the relationships between the variables in form of a bar graph. As previously stated, there are 12 genres we are analyzing.

An initial conclusion we can draw from figure 2 is that North America is the highest selling market across all 12 genres. North America yields the most sales for all the genres by a significant margin. We also see that out of all the genres Action has the highest sales in each region except Japan where we see that role playing games have the highest sales.

Looking more in depth at our data we do an analysis of our variance for each region.

Let us look at the ANOVA table for North America:

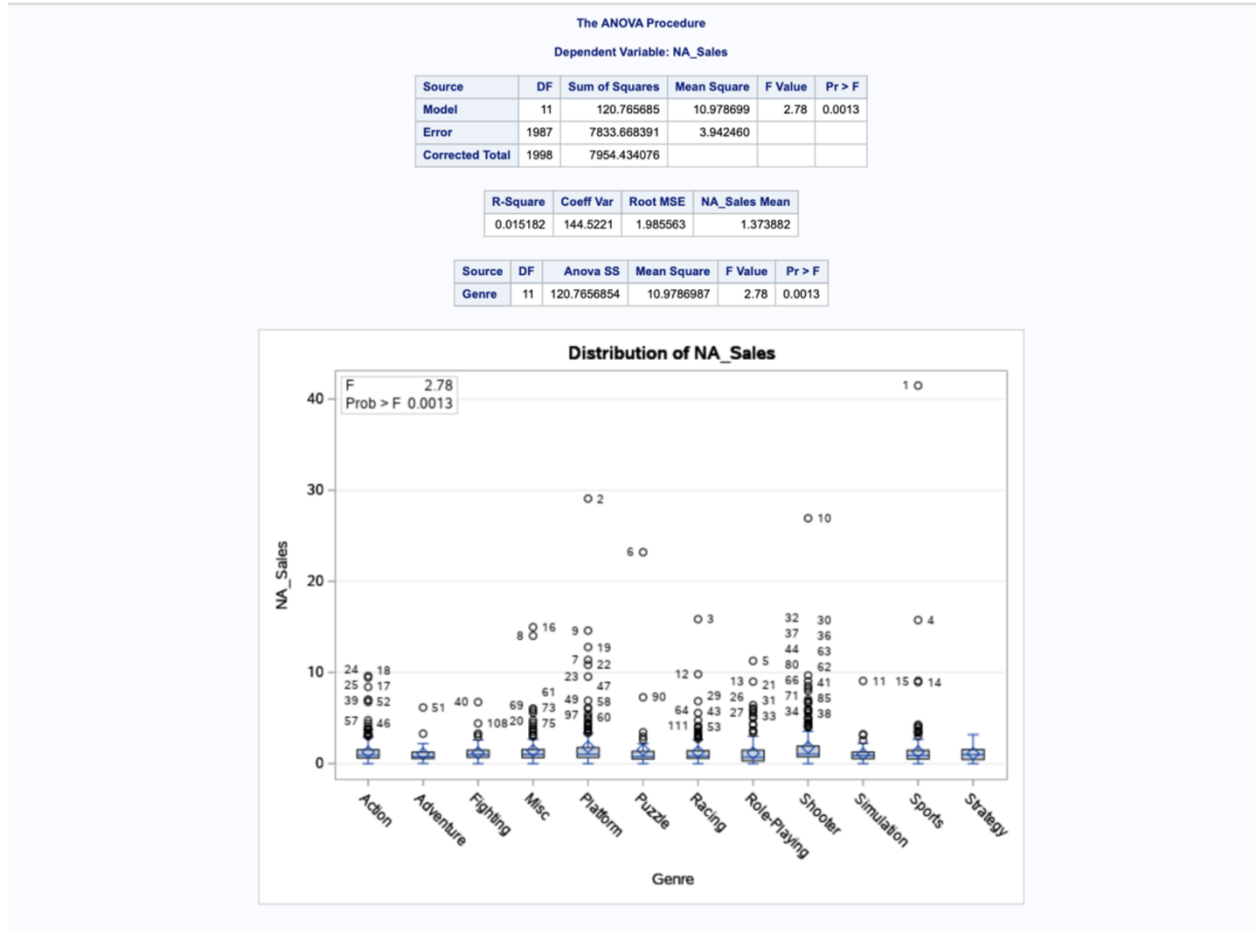


Figure 3: ANOVA table of North America sales data based on genre

State the hypothesis being tested by this model:

H_0 : All North American Sales are same for each genre

H_1 : There is some difference between mean sales in north america between the genre

I decided to do an ANOVA table to compare the mean of the sales data of each genre in North America. We can see with this that the null hypothesis is rejected. Because we reject the null hypothesis (all the North American Sales is the same for each genre), we can say assume that the alternate hypothesis that states there is some difference between mean sales. After doing the pairwise test I can see that the mean difference of 0

is in none of the genres in North America (The result of the pairwise test is in the results file). Hence this tells us that the sales of each genre in north America is different. Also, we can see that North American sales are highest in the Action Genre.

ANOVA table for Japan:

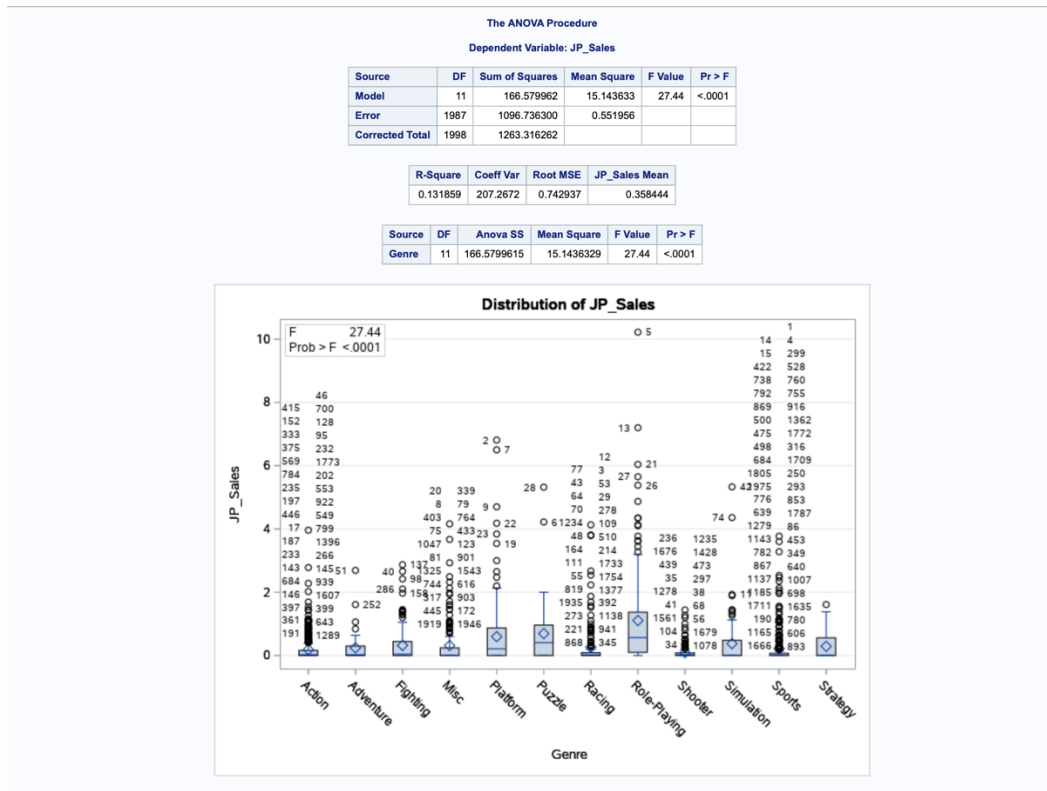


Figure 4: ANOVA table of Japan Sales data for each genre

State

the hypothesis being tested by this model:

H_0 : All Japanese Sales are same for each genre

H_1 : There is some difference between mean sales in Japan between the genres

Another ANOVA table to compare the mean of the sales data of each genre in Japan. We can see with this that the null hypothesis is rejected. Because we reject the null hypothesis (all the Japanese Sales is the same for each genre), we can say assume that the alternate hypothesis that states there is some difference between mean sales has occurred. After doing the pairwise test I can see that the mean difference of 0 does occur between some of the genres (The result of the pairwise test is in the results file). Hence this tells us that the sales between some genres are different. Also, we can see that Japanese sales are highest in the Role – Playing Genre.

Let us look at the ANOVA table for Europe:

State the hypothesis being tested by this model:

H_0 : All European Sales are same for each genre

H_1 : There is some difference between mean sales in Europe between the genres

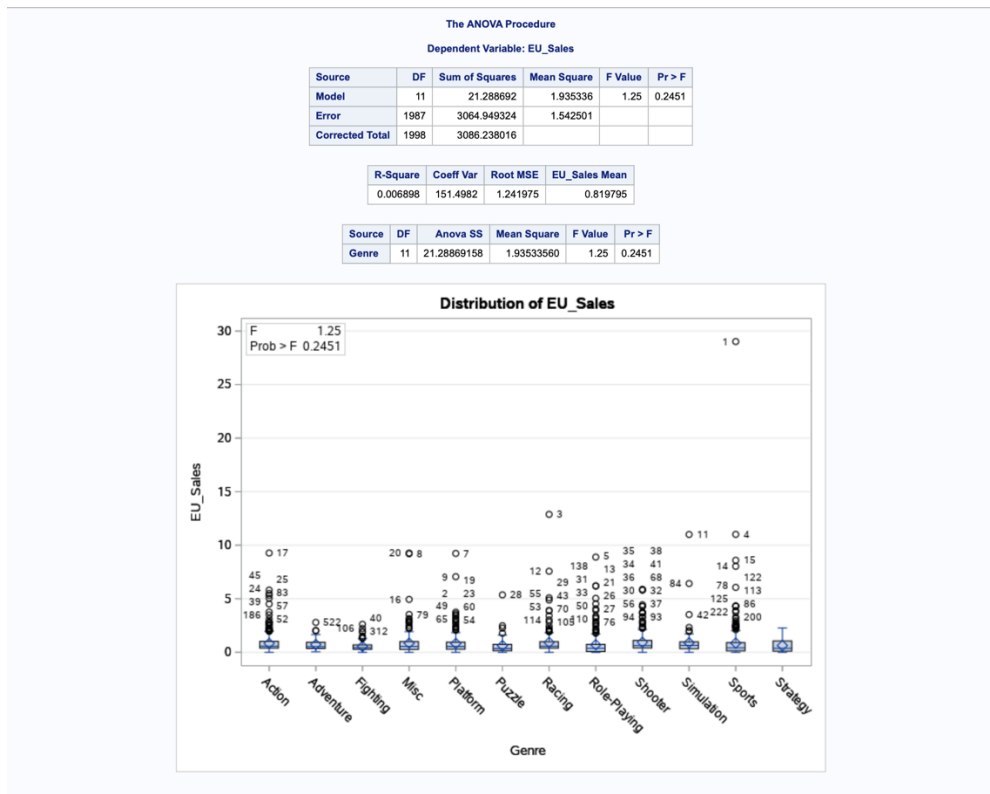
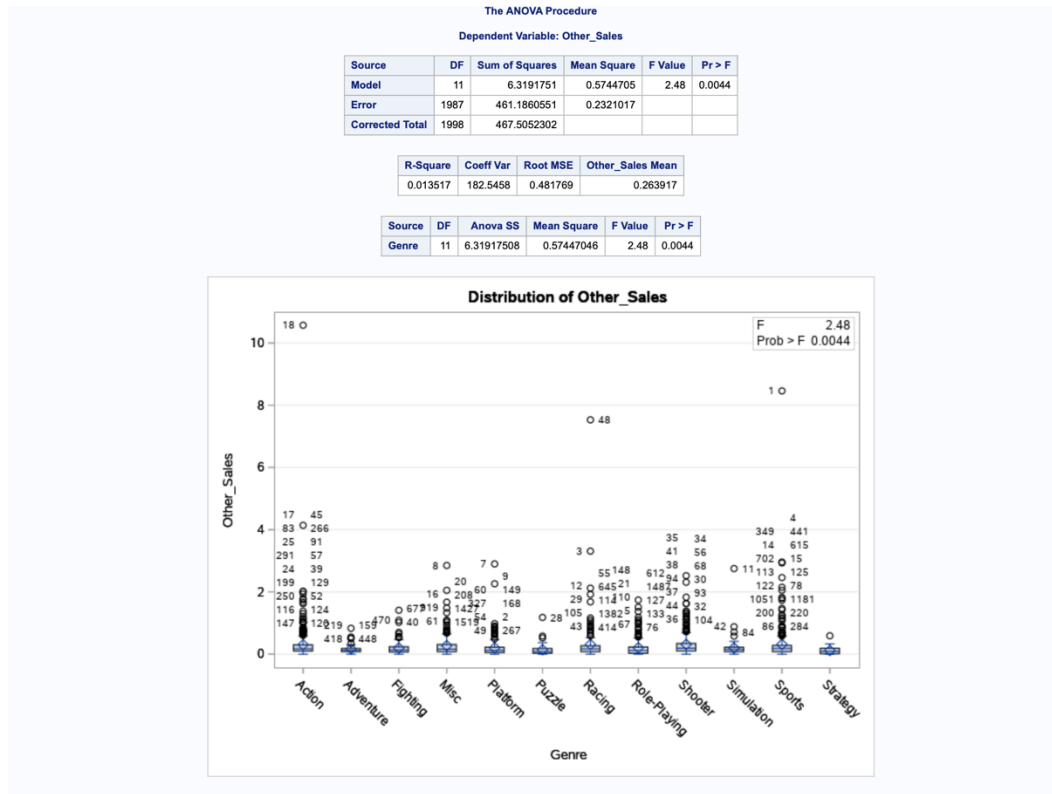


Figure 5: ANOVA table for Europe sales data in each genre

Another ANOVA table to compare the mean of the sales data of each genre in Europe. We can see with this that the null hypothesis is not rejected in this case. Because we don't reject the null hypothesis, we can't assume that the alternate hypothesis that states there is some difference between mean sales has occurred.

Let us look at the ANOVA table for Other Sales:



State the hypothesis being tested by this model:

H_0 : All other Sales are same for each genre

H_1 : There is some difference between mean sales in other countries between the genres

Another ANOVA table to compare the mean of the sales data of each genre in other countries. We can see with this that the null hypothesis is rejected. Because we reject the null hypothesis (all the other Sales is the same for each genre), we can say assume that the alternate hypothesis that states there is some difference between mean sales has occurred. After doing the pairwise test I can see that the mean difference of 0 is in none of the genres in Other Countries (The result of the pairwise test is in the results file). Hence this tells us that the sales of each genre in Other Countries is different. Also, we can see that other countries sales are highest in the Action Genre.

We see that the Role-Playing games are the most popular genre of games in Japan. The reason behind this could be that Japan is a highly successful producer of Role-Playing games. Also, this genre of video games has a long-standing history in Japan. All in all, we see that Action games are the most popular genre video game in all the other regions.

Discussion

Our analysis showed that the Publishers whose Japan market sales come close to or exceeding Europe both have their origin in Japan, which could explain the higher Japanese market sales relative to those Publishers who do not have their origin in Japan. However, it is important to note that Sega also has its origin in Japan but does not follow a similar pattern of high Japanese sales like that of Nintendo or Capcom. For the other Publishers North America is highest in Sales.

Our analysis also shows that the action genres is most popular among all the regions besides Japan. Japan has its highest sales in the role- playing genre. The reason behind this could be that Japan publisher are some of the highest producers of role-playing games.

Some possible problems with our conclusion are that we are using graphical analyses to judge the general patterns across different markets. We would have to take a closer look at some of the trends among the different variable to gauge a better understanding.