Video Game Capstone Inferential Statistical Analysis

For the inferential statistics in this project, I decided to break down essentially all of the important relationships in the data. To start, I ran t-tests for both the Genre and Platform of the video games against the Global as well as regional sales. Then, for the categorical data like Genre, Platform, Rating and Developer, I used a chi-squared test to determine their relationships with each other and also dug deeper in certain situations. Lastly, Pearson correlation tests were used to see the relationship between the different sales of games as well as their relationships with Critic and User scores. For all tests, the alpha was set to 0.05 but the null hypothesis was somewhat different for each. In t-tests, it was that the factor had no impact on sales. The null hypothesis for the chi-squared tests was that there was no association between the factors, which is similar to the correlation tests that stated there was no relationship between each factor. All together, there were many p-values that showed statistical significance and explained a lot about the sale of video games.

When it came to the t-tests for Genre vs. sales, most of them showed statistical significance for at least one type of sale. Racing and Platform games were the only ones that did not, which means they do not sell better or worse in any of the measured areas. One thing that stood out though was that while most of the significant areas had a positive relationship, Japan showed a negative relationship for most types of games. For example, Sports games sold better in North America but sold worse in Japan. Also, Shooters sold better in Global, North American and European sales but sold worse in Japanese Sales. Even Action games only showed significance in Japanese Sales, where they sold worse. The lone exception was Role-Playing games, which sold better in Japan while selling worse in North America and Europe.

For the other genres, Miscellaneous games sold better globally and in North America while Fighting games struggled in Global and European sales. Puzzle and Simulation games only showed statistical significance in North America, where they sold worse. Finally, Adventure and Strategy games had negative t-statistics globally and regionally, which explains why they are the weakest selling genres in the Data Story plot. These tests add credibility to the plots in explaining which genres do better in each area and how important they are to overall sales. Sports games do the best because they sell better in North America while Shooter games are stuck in the middle since they struggle in Japan. The bottom three (Role-Playing, Adventure and Strategy games) do worse in most areas so it makes sense why they are at the end of the spectrum.

The t-tests for Platform vs. sales had many results that showed statistical significance in all types of sales. Games for consoles like Xbox, PC and GameCube sold worse in all areas while the Wii and original PlayStation games sold better in each. The Xbox 360 games sold better in all areas except Japan, where it sold worse, while the Dreamcast games only did better in Japan and struggled in the other areas. Also, Japan's passion for Nintendo was exemplified as the DS and 3DS games only showed statistical significance there and each sold better. Otherwise, PlayStation 3 games sold better globally as well as in North America and Europe while PSP and PlayStation Vita games sold worse; PlayStation 4 games did well in Global and European sales but struggled in Japanese Sales. Lastly, Xbox One games only showed

statistical significance in Japan, where they sold worse, and Game Boy Advance games struggled in both areas they showed significance, Global and European sales.

Altogether these t-tests further illustrate what the Data Story plots displayed earlier. Nintendo does very well in Japan but can be hit-or-miss in other areas. Microsoft, on the other hand, consistently struggled in Japan with a lot of success elsewhere when it came to Xbox 360 games. Games for Sony's home consoles did well overall especially in Global and European sales. One thing that stood out to me, however, was how both the PlayStation 2 and Wii U games did not show statistical significance in any area. For the PS2, that makes some sense, as it was so successful overall that the games did not sell better or worse in any area. However, the Wii U is considered a disappointment for Nintendo, implying it sold worse but none of the areas showed any significance when it came to games sold. While this does not necessarily make the console a success, it does go to show that there is more to sales than just the overall numbers.

Moving to the chi-squared tests, all showed statistical significance. The associations between Genre, Platform, Rating and Developer were not random. This means that there is some statistical relationship between any two. To delve further, I compared some of the top genres, platforms and ratings with each other to see just how significant the relationships are. For Genre, I looked at Action, Shooter and Sports games while PlayStation 2, Nintendo Wii and Xbox 360 games were the focuses in Platform. Lastly, the key ratings I delved into were E (Everyone), T (Teen) and M (Mature), which are the standard types of ratings games receive. When these chi-squared tests were run again and only focused on the main three factors, they all still showed statistical significance. This means that even the top genres, platforms and ratings all have non-random associations with each other. Although sales may not have been a focus in these tests, they still explain the importance of other factors in the field.

Similar to the chi-squared tests, all of the Pearson correlation tests showed statistical significance with some showing stronger positive relationships than others. The relationships between Global Sales as well as North American and European sales were each over 0.9, which is very high. Its relationship with Japanese Sales was a little weaker, around 0.61. Between North America and Europe, the relationship was also very strong, close to 0.84 while their relationships with Japan were also weaker, around 0.47 and 0.52 respectively. These tests help support the earlier plots that showed the same strong relationships when it came to fraction of sales. North America and Europe have a strong connection with each other as well as with Global Sales.

For Critic and User scores, all were statistically significant but the positive relationships were much weaker. For Global Sales, the relationship with Critic Score was around 0.24 and User Score around 0.09. In North American Sales, they were around 0.23 and 0.09 while European Sales had relationships close to 0.21 and 0.06 respectively. Japanese Sales had similarly low relationships but had a lower one with Critic Score (around 0.15) and a higher one with User Score (around 0.13). While these scores do not have much of a predictive relationship between the factors, it does show that User scores are more important in Japan while Critic scores are less critical. Although the positive relationships are not strong, there is still some significance for Critic and User scores when it comes to sales.

All of these statistical tests continued to stress the important trends from the Data Story. The driving force for sales in Genre tended to be North America while Japan saw many sell worse including Shooters. When it came to consoles, it became more clear how much Microsoft's systems struggled in Japanese game sales unlike Nintendo while Sony probably did the best overall. All of the categorical data showed statistical association between each other including when focused on the biggest types of genres, platforms and ratings. Lastly, the relationships between the different types of sales further exemplified what the plots from the Data Story illustrated with North American Sales having a strong relationship with Global and European sales while Japanese Sales had the weakest relationships. And while Critic and User scores did have statistically significant relationships with sales, they were not strongly correlated. When you put all of these tests together, they show just how deep the statistical relationships are when it comes to video game sales.