Math 3030 Module 1 Final Report

Noah Dollimount

January 2026

1 Introduction

Marvel Rivals is a highly popular multiplayer video game featuring 6 versus 6 matches. At the end of each match it is possible to view the performance of each player, measured by various stats such as kills, deaths, etc. Depending on a win or loss, players have a tendency to attribute this outcome to the value of some particular stat. My intention is to perform a logistic regression, evaluating the correlation between these various statistics and a player winning or losing a match. In essence, the goal is to determine the degree to which an individual players metrics statistically contributes towards the success or failure of their team. This is an area of mathematics that sits firmly in the realm of statistics and data science, utilizing data and computer software to make predictions about a players performance based on their end game stats.

E-sports are a rapidly growing interest among people who enjoy and play video games. People who are able to perform at the highest levels of play are often compared to celebrity athletes for their skills, talent, and ability to entertain large crowds. Like athletes, people often want to measure and track the performance of these players in order to make predictions and even bets on the outcome of competitive matches. Being able to make more accurate predictions gives prospective betters an advantage over others.

2 Methods

The main work being performed here was done using a software package that can be downloaded and used with Google sheets, called XLMiner Analysis ToolPak. This software enables users to perform a multitude of statistical tests on recorded data, in particular, the logistic regression tool was used as we are chiefly interested in the correlations of various stats to a binary value representing a win or loss.

Although the initial plan was to record the necessary information using web scraping tools, the website that was ultimately selected for data gathering, Tracker.gg, forbids the use of such tools. This means that the data had to be recorded manually in a spreadsheet, which was assisted by the use of AI tools to transcribe snips of pictures into usable data.

The data gathered pertain to the following statistics

1. KDA (Kill-Death-Assist Ratio): The ratio obtained by adding kills and assists and dividing by deaths

2. Kills: Enemy player deaths as a result of damage the player personally contributed

3. Deaths: The number of times that the players health is reduced to 0
4. Assists: Kills that were acquired by players the player was directly aiding in some manner, typically through healing or buffs
5. Final Hits: Kills in which the player directly reduced an enemy's health to 0, meaning that you were the last person to cause them damage
6. Damage: The total number of hit points dealt by a player character during a match
7. Damage Blocked: Damage that was absorbed through methods that ultimately prevented teammates or the player from receiving said damage, typically obtained through shields or walls utilized by the player
8. Healing: The total number of health points granted to teammates by the player
9. Accuracy: The percentage of attacks or abilities which successfully landed on a player or players

3 Results

Table 1: Regression Results (100 Games)

Variable	**Coefficient**	**Standard Error**	**P-value**
Intercept	0.1844384524	1.746393028	-
KDA	0.9167161854	0.5068814792	0.0705226555
Kills	-0.0318750374	0.134542985	0.8127239663
Deaths	-0.1877311738	0.1878119424	0.3175186717
Assists	0.2282007551	0.1440593947	0.1131768589
Final Hits	0.4682867582	0.2354443076	0.04670681465
Damage	-0.00263653037	0.001475225897	0.07390444124
Blocked	0.0001084086170	0.000378567813	0.7745984873
Healing	-0.00338539688	0.001665284296	0.04206017361
Accuracy	-1.613471347	2.547238406	0.5264595113

The most important value that should be considered first is the P-values, a low P-value means the coefficient obtained for a given variable is less likely to have occurred due to luck. In the first 100 games, the variables that appeared to show an immediate predictive effect, filtering by a P-value of less than 0.1, were KDA, Final Hits, Damage, and Healing. While the coefficients for KDA and Final Hits are positive, indicating that these values do indeed make a positive impact on the odds of winning a game, the coefficients for Damage and Healing are not only miniscule, but actually negative.

The second table should be given more heedance, as information on more games is acquired, we can reasonably expect to get a more accurate picture of the relevance of these various stats. Here, the only values that are still statistically significant, referring to the previous P-value threshold, are KDA and Deaths. It is interesting to note that the coefficient for KDA has halved in significance, as

Table 2: Regression Results (205 Games)

Variable	Coefficient	Standard Error	P-value
Intercept	0.9520489363	0.9826923872	-
KDA	0.4405337199	0.2268237135	0.05211465183
Kills	0.05783906867	0.07082845753	0.414152614
Deaths	-0.1903163241	0.09087293153	0.03623214403
Assists	0.05811998859	0.06384175008	0.3626243203
Final Hits	0.1496702625	0.1158446787	0.1963603203
Damage	-0.00130590078	0.000814581962	0.108900535
Blocked	-0.00003714291	0.000244546510	0.8792777357
Healing	-0.00086799003	0.000569189248	0.1272694274
Accuracy	-1.847733807	1.195844618	0.1223151428

it indicates that KDA may have been a significant predictor for the first 100 matches tested but in the long term is not as relevant as initial testing may originally indicate. Interpreting these results as is, it implies that the optimal strategy is one which focuses on self preservation over aggression.

4 Conclusion

As it stands, this project can currently be considered incomplete or inconclusive. Although the P-values tend to decrease as the number of games tested increases, the pattern of decrease is currently erratic and hard to predict with just two tests run. The obvious solution is to test with more games in the hope that over the long run clear patterns should form, but until a better method is found for gathering data, the number of games to be tested is currently limited by manual effort, which is slow at best. Another point which could be expanded upon is filtering games by the role played, Marvel Rivals currently has 41 different characters that are categorized into 3 unique roles. This further exacerbates the difficulty of assigning predictive weights to the data collected, as each character having unique strengths and weaknesses means that the optimal or key stats to focus on will likely differ from one character to another. This research could be expanded on by narrowing the focus to a particular role or character.

5 Appendices

5.1 AI Usage

Artificial Intelligence was primarily used to assist in gathering and recording information in google sheets. It was also used to generate the Latex code that was used to create the two tables in the results section. Finally, it was also used for general assistance in formatting of the report, such as in ensuring the references were in proper MLA format.

5.2 References

References

[1] Bailey, Kevin. "Statistical Learning for Esports Match Prediction." 2024.

[2] Casella, George, and Roger L. Berger. *Statistical Inference*. 2nd ed., Duxbury, 2002.

[3] "Logistic Regression in Machine Learning." *GeeksforGeeks*, 21 June 2023, www.geeksforgeeks.org/logistic-regression-in-machine-learning/.

[4] "Goodhart's Law." *Wikipedia*, Wikimedia Foundation, 15 Oct. 2024, en.wikipedia.org/wiki/Goodhart%27s_law.

[5] Jim. "Logistic Regression Overview with Example." *Statistics by Jim*, 2023, statisticsbyjim.com/logistic-regression/logistic-regression-overview/.

[6] Lee, SangHyeok, et al. "Machine Learning Methods for Predicting League of Legends Game Outcome." *Proceedings of the 2020 ACM Southeast Conference*, ACM, 2020, pp. 229–232.

[7] "Marvel Rivals Game Stats." *Tracker.gg*, 2024, tracker.gg/marvel-rivals.