

Predictions of the Outcomes for League of Legends and Diamond Price Using Regression Method

Qizhi Fang^{1, *, †} and Minghao Zhang^{2, †}

¹Canada British Columbia Hefei No.1 Secondary School, Hefei, Anhui, China

²Secondary High School Attached Beijing Normal University, Qinghe campus, Beijing, China

*15020140310@xs.hnit.edu.cn

Abstract. Intensified worldwide inflation has added panic to the public about confidence of monetary currency value. People devote themselves to searching for relative constant measurement on country or world economic level evaluation frequently. Under economic uncertainty and on account of precious metal being a great hedge against inflation, gold or diamond are usually applied into economic analysis or prediction. The Diamond, as an equivalent to currency, has extensive acceptance worldwide. Hence, predicting diamond price has the same function compared to economic future forecasting through financial fluctuation. Advances in information technology have facilitated the development of video games and provided opportunities for video game competitions. More and more international tournaments are recognized by sports and sports organizations, and some competitive sports have become official events of the Asian Games. As a sport, fairness is essential, and League of Legends as a 5V5 combat game, fairness on both sides is the basis to maintain the game environment, which means that the winning rate of both sides will be close to 50%. This paper intends to predict diamond price by using simple linear regression. Besides, the forecasting of outcome for League of Legends will be also displayed in this paper, based on the platform R as analysis tool and standard indication. These results shed light on two practical applications of regression models in statistics and the corresponding explanations of their causes will be discussed.

Keywords: Diamond Price; League of Legends; Regression Method; Prediction.

1. Introduction

Statistics play an essential role with extensive use in different areas, and the underlying causes of many things can be explained by statistics. Furthermore, as one of basic statistical research methods, regression model always apply to more realistic problem, especially are used to explain the correlation of multiple dependent variables in different situations and to calculate the relevant expressions. The progress of information technology has promoted the development of video games and provided opportunities for video game tournaments. More and more international tournaments are recognized by sports and athletic organizations, and some game sports have become official projects of the Asian Games. As a sport, fairness is essential, which is often affected by more unstable situations (e.g., the internet), thus causing unfairness in the game and violating the spirit of Olympic sports. League of Legends (LoL) is a multiplayer online battle arena video game developed and published by Riot Games [1]. As a 5V5 battle game, the fairness of both sides is the basis for maintaining the game environment, which means that the winning rate of both sides will approach 50 percent.

Diamond, as a precious metal, has the function of measuring changes in the economy. As the public needed standards to assess the level of the economy due to the constant interference of inflation or deflation in the value of money, they eventually introduced precious metals as a measure. Since most precious metals represent a store of value, they are often used to examine economic trends and levels of the economy. One of the most representative examples is diamond. In addition, due to its own properties such as durability or extreme fitness, diamond has also received a great deal of attention and popularity [2]. Therefore, diamond prices can be applied to economic or commercial domains as a forecasting method for those domains to predict possible tendencies and avoid economic depression or commercial damage. Diamond price forecasts can also be a guide to future economic trends to a certain extent. To analyze the relationship between price and specific parameters of

diamonds, regression models are used. For linear regression, it can determine the correlation between two variables. Besides, it can reflect the fitting degree for independent and dependent variables as much.

This study will focus on using the data in League of Legends and using regression method to calculate the winning rate of the game. As for the diamond price prediction, this paper will use simple regression to interpret interior relation for price and three parameters of diamond, respectively. Model and plot building will rely on Microsoft Excel, involving linear regression models and fitting degree plots for each parameter. Based on the dataset from Kaggle, the diamond analysis will discuss three factors, i.e., X, Y, and Z. Each represent the length, width, and depth of a diamond to enhance reliability. The Microsoft Excel cannot deal with exceedingly data, so each parameter is limited into a specific range elevating the model's confidence level.

2. Prediction of the Outcomes for League of Legends

Logistic regression, also known as logistic regression analysis, is a generalized linear regression analysis model. To explore the factors that influence stock price the probability of the price change based on the risk factors [3]. Formally, the two dependent variable values are marked as "0" and "1" in the logistic regression. Logistic regression will provide the coefficients that best predict the probability of any team winning a past game based on these past data. We expect this relationship to remain stable in future games.

Table. 1 Player Game Datas.

gameId	blueWins	blueTotalGold	blueAvgLevel	blueGoldDiff	redTotalGold	redAvgLevel	redGoldDiff
4519157822	0	17210	6.6	643	16567	6.8	-643
4523371949	0	14712	6.6	-2908	17620	6.8	2908
4521474530	0	16113	6.4	-1172	17285	6.8	1172
4524384067	0	15157	7	-1321	16478	7	1321
4436033771	0	16400	7	-1004	17404	7	1004
4475365709	1	15899	7	698	15201	7	-698
4493010632	1	16874	6.8	2411	14463	6.4	-2411
4496759358	0	15305	6.4	-2615	17920	6.6	2615
4443048030	0	16401	7.2	-1979	18380	7.2	1979
4509433346	1	15057	6.8	-1548	16605	6.8	1548
4452162573	0	15474	6.6	331	15143	7.2	-331
4453038156	0	16695	7	-1505	18200	7	1505
4515594785	1	17865	7.4	3274	14591	6.8	-3274
4524924257	0	14979	6.6	-3414	18393	7.2	3414
4516505202	1	15722	6.8	-470	16192	7	470
4482120064	0	15015	6.8	-1996	17011	7.2	1996
4523758462	1	19733	7.6	5228	14505	6.8	-5228
4503636905	0	14852	6.8	-1975	16827	6.8	1975

R provides a number of statistical programs that allow users to perform statistical analysis by simply specifying a database and a number of parameters. It provides not only several integrated statistical tools, but also a variety of basic functions for mathematical and statistical calculations, thus giving users the flexibility to analyze data and even create new statistical calculations that meet their needs. Users can use R to construct linear or non-linear models, perform some common statistical tests, analyze time series, or perform classification and clustering analysis of data. In this study, R will be the main research method in which a regression value will be brought into the calculation, facilitating the derivation of a strong relationship between multiple factors and the thesis.

The dataset is collected from Kaggle, and the selected part shows the winning ratio of different players. 0 represents the defeat of the blue side, 1 represents the victory of the blue side, and the other data are directly related to the game results, such as the total amount of gold coins, the economic difference of the red and blue sides, etc. [4]. The table 1 counts 20 of players at Diamond and above

in the League of Legends game (higher levels mean that there will be less uncertainty in the outcome of the game). Values such as total red side economy, total blue side economy, and experience gap are represented in the table. However, not all data will be included in the statistical cohort, and the data and reasons that will most important. Two independent variables that can most affect the final outcome of the game to start the study. Through the test, we can combine two types of data to predict future game outcomes: the percentage of wins in past games and the difference in the percentage of coins spent in past games. Logistic regression will provide the coefficients that best predict the probability of any team winning a past game based on these past data. We expect this relationship to remain stable in future games. We define the model as:

$$\begin{aligned} \text{Probability Blue Side Team Wins the Game} \sim & \text{Intercept} + \\ & \text{Coefficient 1} * (\text{Difference in Blue} - \text{Red Win Percentage}) + \\ & \text{Coefficient 2} * (\text{Difference in Blue} - \text{Red Gold Spent Percentage Difference}) \end{aligned} \quad (1)$$

Table 2. Results from R

	Estimate	z value
Intercept	0.2384	7.808
Win Percentage Difference	2.38	6.241
Gold Spent Percentage Difference	5.43	4.504

Plugging it in to R, the results are summarized in Table 2. To estimate the probability of the blue team winning the game, all else being equal, keeping the other two numbers at 0 and inserting an intercept of 0.2384 in the chart, we get the odds of the blue team winning the game as $1 / (1 + \exp(-0.2384)) = 0.559$. Here, 0.559 means that in an otherwise evenly matched game, the blue team should be predicted to win 55.9% of the time. This is the reason why the blue side is always preferred in many League of Legends international tournaments [5]. The blue and red sides of League of Legends are not the same, they show the mapping as mirror images. This means that the map has no axis of symmetry at any point. There is also an argument that certain champions will perform better on the blue line because of the difference in terrain [6]. In addition, the terrain aspect will be more advantageous. The game interface is more favorable to the blue side, so it will be easier psychologically. Red side is looking from top to bottom and blue side is looking from bottom to top, which is more comfortable visually. It is perfectly possible to win the game by selecting and banning stages. The red team will have an advantage in the pick and ban phase. The blue side is more dominant in terms of BP, the impact on BP is obvious, the two sides of the ban, can ban 10 heroes, each ban a person; blue side 1 floor first pick, after the red side 1, 2 floor selection, and then turn to the blue side 2, 3 floor selection, according to such z-word selection [7]. After both players have finished picking, they enter the game. A champion is more likely to be popular if he has a lower ban rate, lower damage tolerance, and lower damage dealing ability, or if a champion has a higher ban rate, a higher win rate, and is able to kill enough of the opponent's followers. Based on the advantage of being on the blue side will increase the possibility of winning to the maximum [8].

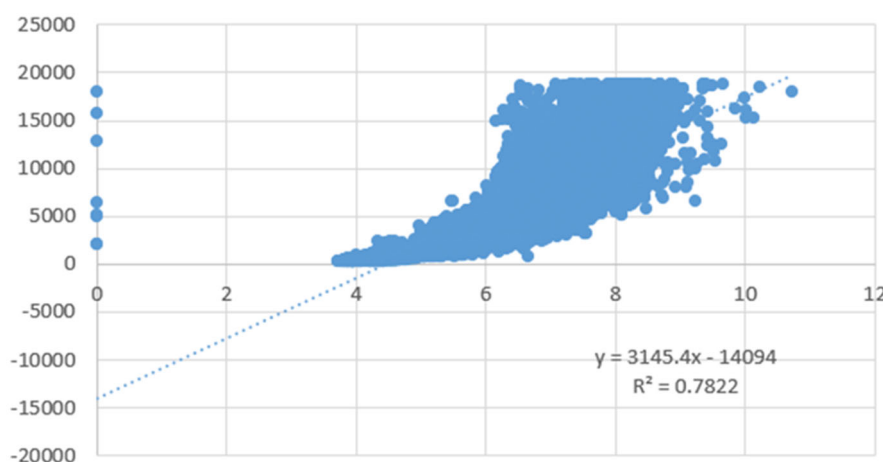
3. Mythology & Analysis for Diamond Price Prediction

As one of propaedeutic machine learning method, simple linear regression can apply in various fields. According to Zou, simple linear regression inly focuses on one specific predictor by analyzing specific effect for particular result. Furthermore, simple linear regression could reflect the relation between one dependent variable and various independent variables under certain datasets [9, 10]. Linear regression model especially marks with its linear in the parameters, and it contains some crucial parameters. Seen from Table 3, these parameters include R square, adjusted R square, P value, Significance F, and Intercept Coefficients for both X and Y. They can indicate the fitting degree and confidence level of the established model of variables. Simple linear regression model only includes a single independent variable, so it has primary difference with multivariate linear regression.

Table 3. Microsoft Excel Linear Regression Parameters

	X	Y	Z
Multiple R	0.884432769	0.865418674	0.86124881
R Square	0.782221324	0.748949482	0.74179513
Adjusted R Square	0.782221729	0.748944898	0.741744725
Significance F	0	0	0
Intercept Coefficients	-14094.11883	-13402.09076	-13296.63622
X Variable Coefficients	3145.411956	3022.886933	4868.801106
P Value	0	0	0

Machine learning is necessary when dealing with abundant data, especially for supervised machine learning models. As one of the spreadsheet platforms, Microsoft Excel can directly manifest independent and dependent variables. Additionally, the outcome of Microsoft Excel is more decipherable, and standardizing data can be easily chosen for demonstration. Data source comes from Kaggle, and the dataset of diamond analysis is available for regression models. The total number of the dataset is 53940. The dataset includes 9 parameters totally. Each of these last parameter quantities is emblematic of diamond values such as carat, color, and clarity. Except previous measurement, the natural dimension of diamond was also involved in the evaluation of values. Consistent with the dataset, research only select three aspects of diamond dimension to accomplish the forecasting. And each parameter has its corresponding range (X: 0-10.74, Y: 0-58.9, Z:0-31.8, unit: mm) to reduce model error and operate greatly [11]. The analysis process contains three steps: data collection, model building, and result evaluation. Data preprocessing and clearing have fulfilled before data gathering. The analysis has ensured all the prerequisites like other 6 parameters are identical. Primarily, raw data are cited from dataset based on Kaggle. After data preprocessing and cleaning, conditional data will be used as basis for subsequent analysis. Since the total amount of data is 54,000, the conditional data is 53,940 after data selection. All conditional data then is applied into linear regression calculation on Microsoft Excel. Each parameter has a separate model, and finally receive a different result. Any valuable value soon afterwards gathers altogether into one collection. The points of the plot shown in Fig. 1, Fig. 2 and Fig. 3 are to denote the difference of each parameter in relation to the price under the same circumstances. Moreover, each model can thus be studied to obtain the scatter plot. Each scatter plot represents the degree of fit and the rough trend under the range limit. The manifestation of each parameter is revealed in Table. R square, adjusted R square, and multivariate R square represent the goodness of fit between x and y. Moreover, P value and significance F reflect the confidence level of the linear regression model for the dependent and independent variables [12].


Fig. 1 Fitting Plot of X (length, unit: mm) as a function of Price.

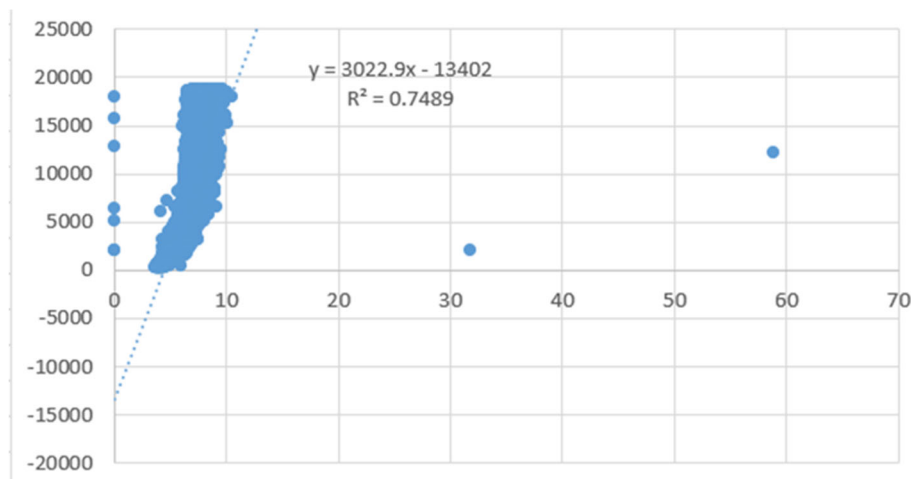


Fig. 2 Fitting Plot of Y (width, unit: mm) as a function of Price

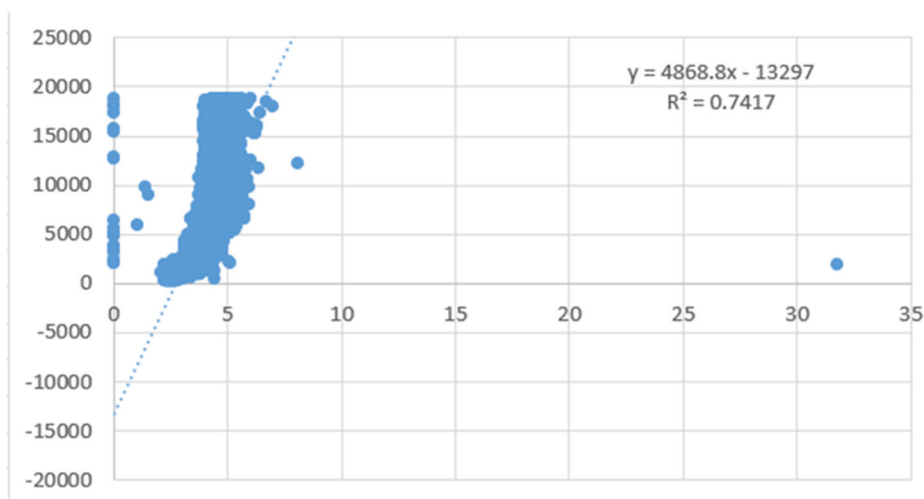


Fig. 3 Fitting Plot of Z (depth, unit: mm) as a function of Price.

Besides, each parameter has its scatter plot versus price. The univariate plot reveals the distinction between two variables and assimilates with its correlation, especially determine the probability distribution of a continuous value [13]. Referring to the results of data collection and plots, the conclusions can reflect the effect of highlighted length on diamond prices. First, the collection of data demonstrates the outstanding numerical value of length compared to width and depth. As X (length) has greater R square, adjusted R square, and multivariate R square, these parameters manifest the length of diamond fitting better with its relation with price [14]. Thus, in terms of the relationship with price, X has a higher positive correlation. Furthermore, based on fit diagram, diamond length appears extensive outcome by distributing relative vastly o on the range of limitation. Comparing to width and depth, the scatter diagram reflection figures out more equally-distributed X instead of restrict in a certain area [15]. Since the variation trend of length is more obvious than other two parameters, it proves the leverage of length have on price. Taking account of each parameter range into consideration, only length has relative perfect normal distribution [16]. Because the distributions of width and length are restricted into a more confined area. To conclude, length produces a better overall outcome on the effect of diamond price due its preferable fitting degree and equal distribution through its range of study. Hence, a diamond length has relative decisive impact on its price.

4. Conclusion

In summary, this paper investigates predictions of the outcomes for League of Legends and diamond price using regression methods. Specifically, diamond price prediction uses the Microsoft

Excel to discuss one of the most influential factors among X(length), Y(width), and Z(depth), whereas predictions of outcomes rely on platform R to analyze the ideal solution for gaming. As for the results of League of Legends, Riot's only solution was to find a way to mirror the camera to both sides. Until Riot takes this into account, there will always be an advantage on the blue side. Riot has paid attention and tried to solve this problem in several version updates, such as increasing the combat power of red minions in the early game, fine-tuning the terrain map of the red side, and so on. Major League of Legends matches are played in a five-game, three-win format, with the losing side choosing the side they want (most losers choose the blue side in the second game). However, according to the latest version of the experiment, the blue side still has a 5.59% higher win rate than the red side, a value that can be decisive in winning sports matches. Using the advantage of being on the blue side will increase the possibility of winning to the maximum.

Regarding to diamond price prediction, it seems complicated since this project originally included many parameters in evaluating the price of a single diamond. Nonetheless, the linear regression model and Microsoft Excel platform facilitate computation and analysis. This paper discusses the influence of the most preferred factor on the price of diamonds. In the future, more algorithms and machine learning methods should be involved in the study. More standardized parameters need to be added into price forecasting. For instance, regression models such as multivariate linear regression, Random Forest regression, decision tree, Lasso regression plus machine learning platforms (e.g., python, R studio) should be employed in the task to enhance the prediction accuracy and minimize the error of possible underfitting. Besides, basic componential parameter of diamond price as this paper previously mentioned should also consider of future price change of diamond. Overall, these results offer a guideline for applications of mathematical statistics to life and recreation, and specific examples of using regression models to solve relationships between different independent and dependent variables. In particular, it provides ideas and solutions for safeguarding the fair game competition of LOL, and offers some valuable means for the diamond price forecasting based on its three main characters (i.e., length, width, and depth).

References

- [1] Blue Rosario. League of Legends: a beginner's guide. TechRadar. Archived from the original on September 4, 2020.
- [2] Pandey A C, Misra S, Saxena M. Gold and diamond price prediction using enhanced ensemble learning. In 2019 Twelfth International Conference on Contemporary Computing (IC3) 2019, August, pp. 1-4.
- [3] Tolles J, Meurer W J. Logistic Regression Relating Patient Characteristics to Outcomes. JAMA, 2016, 316 (5): 533-4.
- [4] Ma Yilan. League of Legends Diamond Ranked Games (10 Min). Kaggle, 13 Apr. 2020, Retrieved from: <https://www.kaggle.com/datasets/bobbyscience/league-of-legends-diamond-ranked-games-10-min>.
- [5] Carroll Timothy. Predicting the Winner of League of Legends Games. Medium, Analytics Vidhya, 14 Sept. 2020, Retrieved from: <https://medium.com/analytics-vidhya/predicting-the-winner-of-league-of-legends-games-c6fb3513b3d4>.
- [6] Investigating the Impact of Game Features on Champion Usage in League ...<https://myslu.stlawu.edu/~clee/docs/lol.champion.usage.fdg.2015.pdf>.
- [7] Zhang X, Yue Y, Gu X, et al. Investigating the impact of champion features and player information on champion usage in League of Legends. Proceedings of the 2017 International Conference on Information Technology. 2017: 91-95.
- [8] UnrankedSmurfs.com. Blue Side Advantage: A League of Legends Study. Unranked Smurfs, Retrieved from: <https://www.unrankedsmurfs.com/blog/lol-blue-side-advantage>.
- [9] Zou K H, Tuncali K, Silverman S G. Correlation and simple linear regression. Radiology, 2003, 227(3), 617-628.
- [10] Maulud D, Abdulazeez A M. A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 2020, 1(4), 140-147.

- [11] Sharma G, Tripathi V, Mahajan M, Srivastava A K. Comparative analysis of supervised models for diamond price prediction. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering, 2021, January pp. 1019-1022.
- [12] Orlov M L. Multiple linear regression analysis using Microsoft Excel. Chemistry Department, 1996.
- [13] Kavitha S, Varuna S, Ramya R. A comparative analysis on linear regression and support vector regression. In 2016 online international conference on green engineering and technologies (IC-GET) 2016, November, pp. 1-5.
- [14] Acharya M S, Armaan A, Antony A S. A comparison of regression models for prediction of graduate admissions. In 2019 international conference on computational intelligence in data science (ICCIDS) 2019, February, pp. 1-5.
- [15] Shao L, Schleicher T, Behrisch M, Schreck T, Sipiran I, Keim D A. Guiding the exploration of scatter plot data using motif-based interest measures. Journal of Visual Languages & Computing, 2016, 36, 1-12.
- [16] Marill K A. Advanced statistics: linear regression, part I: simple linear regression. Academic emergency medicine, 2004, 11(1), 87-93.