

Project 1 158

Tesfa Asmara and Kevin Loun

2/20/2022

Contents

1	Introduction	1
2	Hypothesis	1
3	Assessment	3
4	Conclusion	4
	Bibliography	4

1 Introduction

The dataset for this project contains 10,000 League of Legends ranked matches from the North American region with 775 variables offered through the Riot Games API, provided on Kaggle (Games 2021a)(James 2020). Each match is pulled from players who rank Gold in the League system, a ranking system that matches players of a similar skill level to play with and against each other. Amongst North American players, the Gold skill level was the second most common tier, achieved by 27.7 percent of players, or approximately 49.86 million players when considered against Riot Games' player base of 180 million ("Distribution of League of Legends (Lol) Summoners in North America as of October 2021, by Tier" 2021)(Games 2021b). This dataset will be referred to as `lol110`.

For this project, the following variables are of interest: lane, champion, map side, first summoner spell, kills, deaths, gold earned, and total damage dealt. A figure including all the relevant variables and their description is attached at the end.

2 Hypothesis

We consider the following research question: Does the amount of gold earned have an effect on the total damage dealt for the average Gold-ranked player on the blue team? We want to describe the relationship between the gold earned and the total damage dealt on the blue team in the `lol110` dataset using a line. We will use the gold earned across all summoners on the blue team as the predictor variable, x , to predict the total damage dealt across all summoners on the blue team, y . In Figure 1, we observe the nonlinearity of the regression model and the nonconstancy of the error terms.

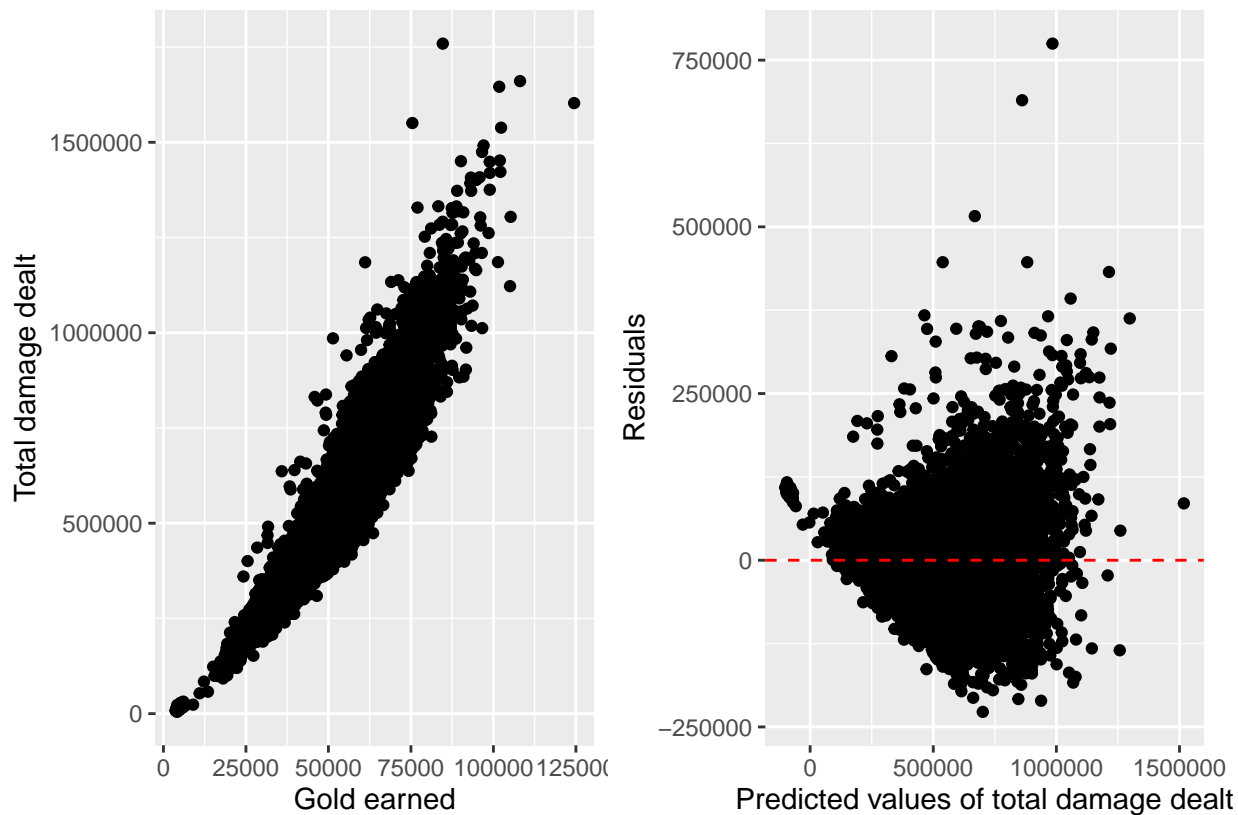


Figure 1: A scatterplot and residual plot showing gold earned against total damage dealt on the blue team for 10,000 Gold-ranked matches

If we apply a log transformation to both the predictor and outcome variable, a positive association between the variables is revealed.

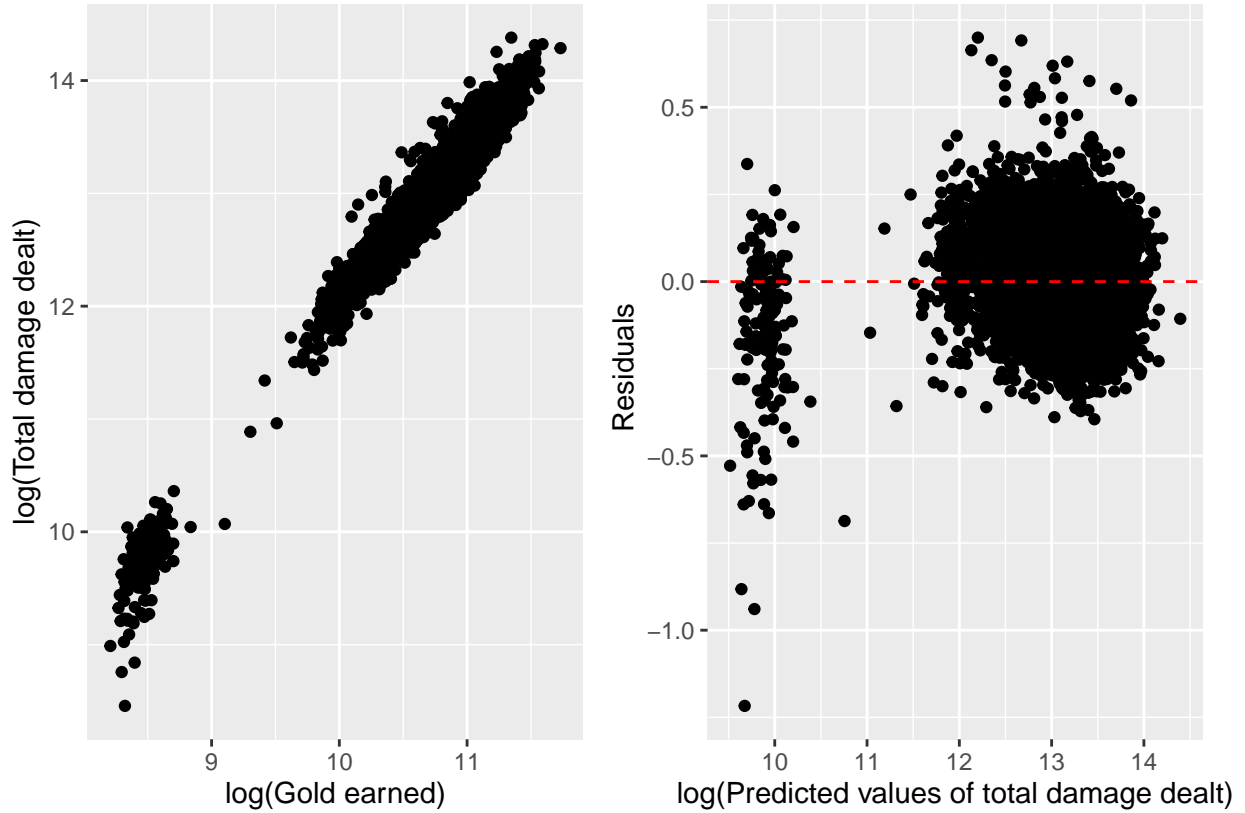


Figure 2: A scatterplot and residual plot for the log-log-transformed model predicting total damage dealt from gold earned on the blue team for 10,000 Gold-ranked matches.

Next, we consider H_0 . In this case, we have: $H_0 : \beta_1 = 0$ and $H_a : \beta_1 \neq 0$. The 95% confidence interval for β_1 is (1.378192, 1.3899095); this means that we are 95% confident that β_1 is in this range. Since the confidence interval for β_1 does not contain 0, it can be concluded that there is evidence of a linear relationship between the log(gold earned) and the log(total damage dealt) for the blue team.

3 Assessment

We can quantify the strength of this linear relationship by computing the coefficient of determination, $R^2 = 0.9553967$, to describe the amount of variation in the outcome variable that is explained by the least squares line; there was a reduction of 95.5396661% of the outcome variable's variation by using information about log(gold earned) for predicting the log(total damage dealt) for the blue team using a linear model.

We can now choose to look at the confidence intervals for the mean and individual response at gold earned = 10.8676537, which is the median of our possible values for log(gold earned). We found that the confidence interval for the mean response and individual response is (1.4093867, 1.5080972) and (1.2191978, 1.6982861), respectively. This means that we are 95% confident that our true log(total damage dealt) lies within this interval.

4 Conclusion

For the 1o110 data, we could write the equation of the least squares regression line as

$$\log(\widehat{\text{total damage dealt}}) = -1.8433137 + 1.3840508 \times \log(\text{gold earned})$$

This models our belief that multiplicative change in the predictor variable is associated with multiplicative change in the outcome variable. In this case, $\beta_1 = 1.3840508$. Hence, a one-percent increase in gold earned is associated with a 1.3840508 percent change in total damage dealt.

Bibliography

- “Distribution of League of Legends (Lol) Summoners in North America as of October 2021, by Tier.” 2021. *Statista*. <https://www.statista.com/statistics/807298/league-of-legends-player-tier/>.
- Games, Riot. 2021a. “Riot Games Api.” *Riot Developer Portal*. <https://developer.riotgames.com/apis>.
- . 2021b. *Twitter*. Twitter. <https://twitter.com/riotgames/status/1455172784938651649?s=20&t=AQmQGrTa1ijf6u3cEDPZcg>.
- James. 2020. “League of Legends Ranked Match Data from Na.” *Kaggle*. <https://www.kaggle.com/jamesbting/league-of-legends-ranked-match-data-from-na>.

Figure 3: Variables and their descriptions