

# Project 1 158

Tesfa Asmara and Kevin Loun

2/20/2022

## 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.

## 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 @ref(fig:xyplot), we observe the linearity of the regression model, albeit with some curvature. In Figure @ref(fig:residyhatplot), we observe the nonconstancy of the error terms.

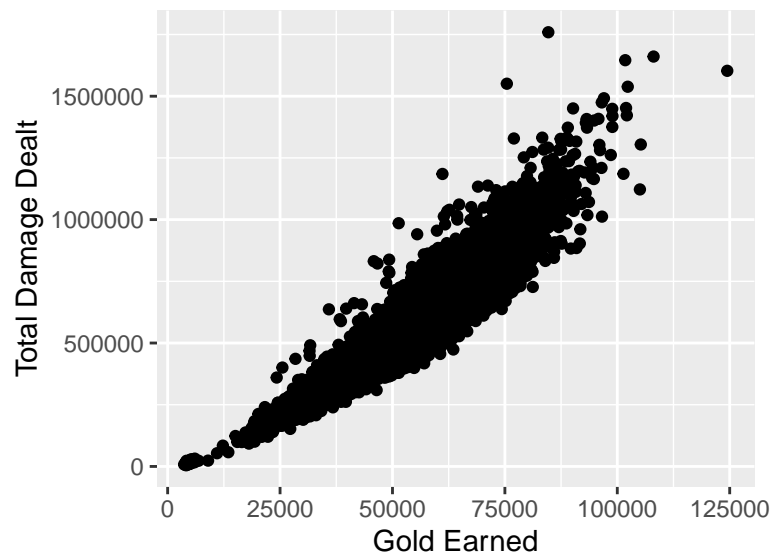


Figure 1: A scatterplot showing gold earned against total damage dealt on the blue team for 10,000 Gold-ranked matches

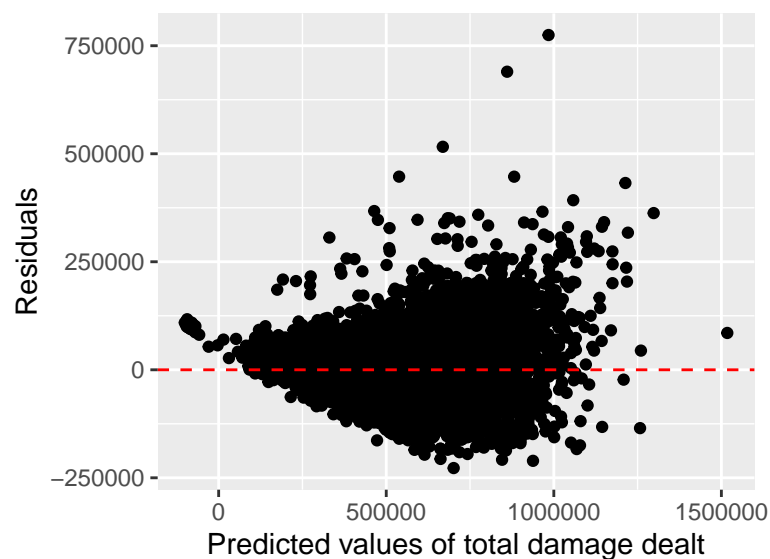


Figure 2: Residual plot for the model predicting total damage dealt from gold earned on the blue team for 10,000 Gold-ranked matches.

Looking at our plots we see that there is an  $R^2$  value of 0.8941 meaning that 89% of the variation in the total damage done can be explained by the amount of gold that a player earns in a game. Next, we consider  $H_0$ . In this instance, we have:  $H_0 : \beta_1 = 0$  and  $H_a : \beta_1 \neq 0$ . The 95% confidence interval for  $\beta_1$  is (0.066, 0.067); this means that we are 95% confident that  $\beta_1$  is in this range. Since the confidence interval for  $\beta_1$  does not contain 0, it can be concluded that there is evidence of a linear relationship between the gold earned and the total damage dealt for the blue team.

The residual plot has a semi-quadratic shape indicating that the model is not fit for regression as we would want a regression plot to be a random scatter rather than contain any trends. This along with the failure of

LINE conditions indicates that we should transform our data.

## Transformed Data

### Log Transformations

While we did observe some linearity in our initial plots, our plots still contained some curvature and we still observed nonconstancy of the error terms. We then sought to transform our data so that it would be more linear and have constant normality of errors. We approached this by taking the squareroot of both our response and explanatory variables.

```
## `geom_smooth()` using formula 'y ~ x'
```

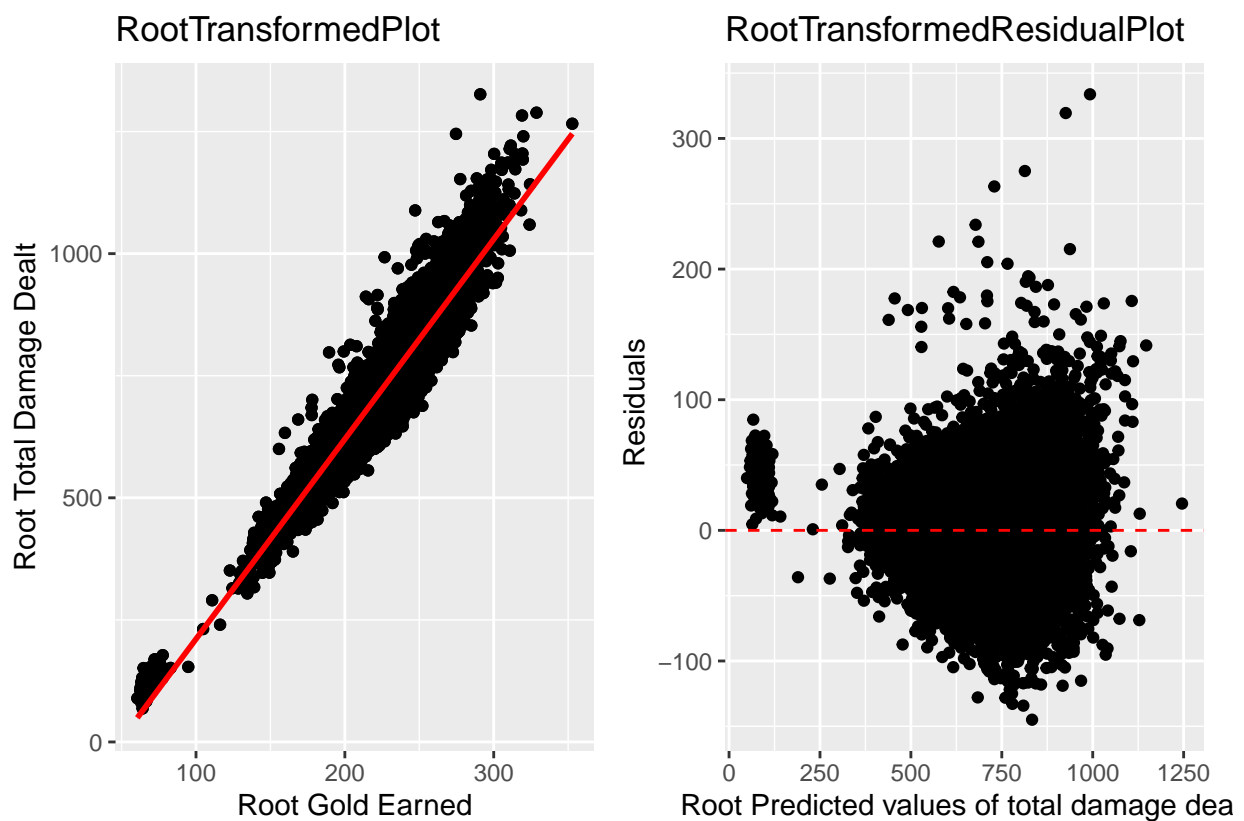


Figure 3: Transformed Residual and Scatter plots for the model predicting total damage dealt from gold earned on the blue team for 10,000 Gold-ranked matches.

From transforming and plotting our data we see that the data is much more linear and has less curvature and that the residual vs predicted plots remains largely the same in terms of shape. While our scatter plot now is linear and has constancy of the error terms, when we look at the residual plot of the transformed data, it maintains the same shape as before the transformation of the data. The residual plot maintains clustered shape which indicates that the model is not fit for regression as when we look at residual plots we look for random scatter of points.

When looking at the  $R^2$  value for the transformed model we find that is it 0.9239 indicating that 92% of the variability in  $\sqrt{\text{Total Damage Dealt}}$  can be explained by  $\sqrt{\text{Total Gold Earned}}$ . We can now choose to look at the confidence intervals for the mean and individual response at  $X = 200$  which is the center of our possible values for  $X$ . We found that the confidence interval for the mean response is  $(-145.5236, -135.665)$  and for the individual response at  $X=200$  is  $(-224.1521, -57.03654)$  meaning we are 95% confident that our true  $\sqrt{\text{Total Damage Dealt}}$  lies within this interval. While this does suggest a correlation between the variables, and the confidence interval for  $\beta_1$  suggest that there is a linear relationship between our variables, the residual plot suggests otherwise. The residual plot is clustered when we would expect it to be random and scattered if the model was a good fit for regression. This indicates that a linear model would not be appropriate for this data.

## Conclusion

Through looking at our residual plots for both transformed and untransformed data we have come to the conclusion that a linear model is not appropriate for this data. This was interesting as the initial plot of the data looked linear before and after transformation and because our interval for  $\beta_1$  suggested a linear relationship. Despite this our residual plots turned out to be clustered and the opposite of what a good linear model would look like. The data initially was what we would expect, the better a player did, indicated by the amount of gold they earned through out a game, the better they did in a match, indicated by the amount of damage they outputted. We are curious about the relationship that exist with the variables chosen in the data still, if a linear model isn't appropriate to measure this relationship, we wonder what kind of model would be best to showcase and explore this relationship.

## 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>.

## Descriptions of Variables

Table 1: Variables and their descriptions

Variable	Description
b_summoner1_role	The role of summoner 1 on the blue side of the map, taking on the values of TOP_LANE, MID_LANE, BOT_LANE, or JUNGLE.
b_summoner2_role	The role of summoner 2 on the blue side of the map, taking on the values of TOP_LANE, MID_LANE, BOT_LANE, or JUNGLE.
b_summoner3_role	The role of summoner 3 on the blue side of the map, taking on the values of TOP_LANE, MID_LANE, BOT_LANE, or JUNGLE.

Variable	Description
b_summoner4_role	The role of summoner 4 on the blue side of the map, taking on the values of TOP_LANE, MID_LANE, BOT_LANE, or JUNGLE.
b_summoner5_role	The role of summoner 5 on the blue side of the map, taking on the values of TOP_LANE, MID_LANE, BOT_LANE, or JUNGLE.
b_summoner1_championId	A number corresponding to the champion of summoner 1 on the blue side of the map.
b_summoner2_championId	A number corresponding to the champion of summoner 2 on the blue side of the map.
b_summoner3_championId	A number corresponding to the champion of summoner 3 on the blue side of the map.
b_summoner4_championId	A number corresponding to the champion of summoner 4 on the blue side of the map.
b_summoner5_championId	A number corresponding to the champion of summoner 5 on the blue side of the map.
b_summoner1_championId	A number corresponding to the first spell of summoner 1 on the blue side of the map.
b_summoner2_championId	A number corresponding to the first spell of summoner 2 on the blue side of the map.
b_summoner3_championId	A number corresponding to the first spell of summoner 3 on the blue side of the map.
b_summoner4_championId	A number corresponding to the first spell of summoner 4 on the blue side of the map.
b_summoner5_championId	A number corresponding to the first spell of summoner 5 on the blue side of the map.
b_summoner1_kills	The number of kills obtained by summoner 1 on the blue side of the map.
b_summoner2_kills	The number of kills obtained by summoner 2 on the blue side of the map.
b_summoner3_kills	The number of kills obtained by summoner 3 on the blue side of the map.
b_summoner4_kills	The number of kills obtained by summoner 4 on the blue side of the map.
b_summoner5_kills	The number of kills obtained by summoner 5 on the blue side of the map.
b_summoner1_deaths	The number of deaths obtained by summoner 1 on the blue side of the map.
b_summoner2_deaths	The number of deaths obtained by summoner 2 on the blue side of the map.
b_summoner3_deaths	The number of deaths obtained by summoner 3 on the blue side of the map.
b_summoner4_deaths	The number of deaths obtained by summoner 4 on the blue side of the map.
b_summoner5_deaths	The number of deaths obtained by summoner 5 on the blue side of the map.
b_summoner1_gold_earned	The gold obtained by summoner 1 on the blue side of the map.
b_summoner2_gold_earned	The gold obtained by summoner 2 on the blue side of the map.
b_summoner3_gold_earned	The gold obtained by summoner 3 on the blue side of the map.

Variable	Description
b_summoner4_gold_earned	The gold obtained by summoner 4 on the blue side of the map.
b_summoner5_gold_earned	The gold obtained by summoner 5 on the blue side of the map.
b_summoner1_total_damage_dealt	The total damage dealt by summoner 1 on the blue side of the map.
b_summoner2_total_damage_dealt	The total damage dealt by summoner 2 on the blue side of the map.
b_summoner3_total_damage_dealt	The total damage dealt by summoner 3 on the blue side of the map.
b_summoner4_total_damage_dealt	The total damage dealt by summoner 4 on the blue side of the map.
b_summoner5_total_damage_dealt	The total damage dealt by summoner 5 on the blue side of the map.
r_summoner1_role	The role of summoner 1 on the red side of the map, taking on the values of TOP_LANE, MID_LANE, BOT_LANE, or JUNGLE.
r_summoner2_role	The role of summoner 2 on the red side of the map, taking on the values of TOP_LANE, MID_LANE, BOT_LANE, or JUNGLE.
r_summoner3_role	The role of summoner 3 on the red side of the map, taking on the values of TOP_LANE, MID_LANE, BOT_LANE, or JUNGLE.
r_summoner4_role	The role of summoner 4 on the red side of the map, taking on the values of TOP_LANE, MID_LANE, BOT_LANE, or JUNGLE.
r_summoner5_role	The role of summoner 5 on the red side of the map, taking on the values of TOP_LANE, MID_LANE, BOT_LANE, or JUNGLE.
r_summoner1_championId	A number corresponding to the champion of summoner 1 on the red side of the map.
r_summoner2_championId	A number corresponding to the champion of summoner 2 on the red side of the map.
r_summoner3_championId	A number corresponding to the champion of summoner 3 on the red side of the map.
r_summoner4_championId	A number corresponding to the champion of summoner 4 on the red side of the map.
r_summoner5_championId	A number corresponding to the champion of summoner 5 on the red side of the map.
r_summoner1_championId	A number corresponding to the first spell of summoner 1 on the red side of the map.
r_summoner2_championId	A number corresponding to the first spell of summoner 2 on the red side of the map.
r_summoner3_championId	A number corresponding to the first spell of summoner 3 on the red side of the map.
r_summoner4_championId	A number corresponding to the first spell of summoner 4 on the red side of the map.
r_summoner5_championId	A number corresponding to the first spell of summoner 5 on the red side of the map.
r_summoner1_kills	The number of kills obtained by summoner 1 on the red side of the map.

Variable	Description
r_summoner2_kills	The number of kills obtained by summoner 2 on the red side of the map.
r_summoner3_kills	The number of kills obtained by summoner 3 on the red side of the map.
r_summoner4_kills	The number of kills obtained by summoner 4 on the red side of the map.
r_summoner5_kills	The number of kills obtained by summoner 5 on the red side of the map.
r_summoner1_deaths	The number of deaths obtained by summoner 1 on the red side of the map.
r_summoner2_deaths	The number of deaths obtained by summoner 2 on the red side of the map.
r_summoner3_deaths	The number of deaths obtained by summoner 3 on the red side of the map.
r_summoner4_deaths	The number of deaths obtained by summoner 4 on the red side of the map.
r_summoner5_deaths	The number of deaths obtained by summoner 5 on the red side of the map.
r_summoner1_gold_earned	The gold obtained by summoner 1 on the red side of the map.
r_summoner2_gold_earned	The gold obtained by summoner 2 on the red side of the map.
r_summoner3_gold_earned	The gold obtained by summoner 3 on the red side of the map.
r_summoner4_gold_earned	The gold obtained by summoner 4 on the red side of the map.
r_summoner5_gold_earned	The gold obtained by summoner 5 on the red side of the map.
r_summoner1_total_damage_dealt	The total damage dealt by summoner 1 on the red side of the map.
r_summoner2_total_damage_dealt	The total damage dealt by summoner 2 on the red side of the map.
r_summoner3_total_damage_dealt	The total damage dealt by summoner 3 on the red side of the map.
r_summoner4_total_damage_dealt	The total damage dealt by summoner 4 on the red side of the map.
r_summoner5_total_damage_dealt	The total damage dealt by summoner 5 on the red side of the map.