SDS 320E – Preliminary Analysis Report

Name: Jason Kim
EID: *jk46965*

Introduction

The goal of this project is to determine whether securing the first baron kill and total gold earned has an impact on the total damage dealt to the enemy team's champions in a world renown video game called League of Legends.

The data collection comes from Kaggle, specifically a dataset that details challenger ranked matches in 2020. The individual sampling units for this study will be League of Legends matches, otherwise referred to as "games."

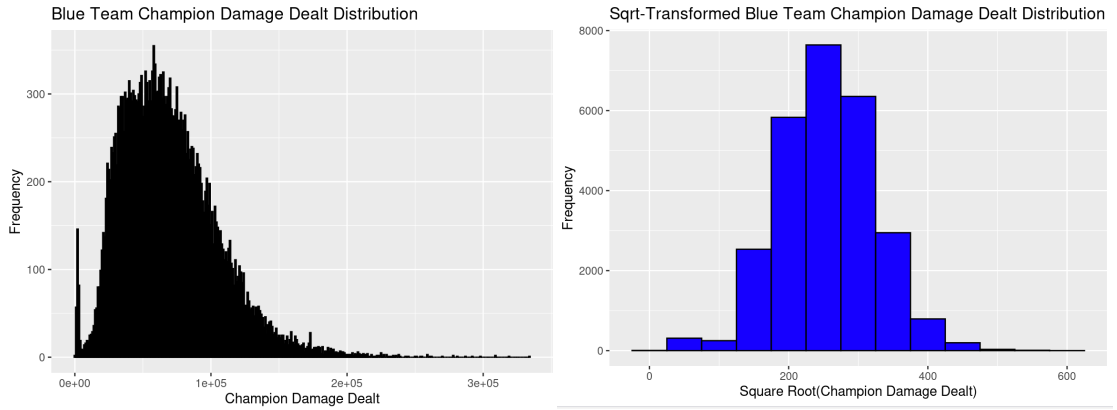
The final sample size is n=26904 matches.

Research Questions

RQ1 – In League of Legends, does securing the first baron kill (binary explanatory) increase the likelihood of dealing more damage to the other team's champions (response)?

RQ2 – In League of Legends, how does the total amount of gold earned (numerical explanatory) by a team relate to the total damage dealt to the opposing team's champions (response)?

Descriptive Analysis of Response Variable



The left graph of the response's marginal distribution shows a right-skewed distribution.

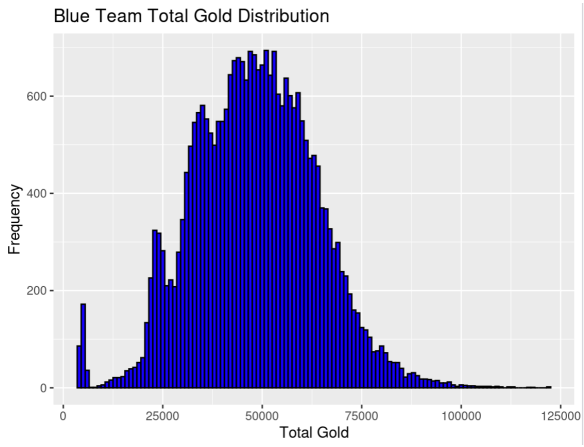
Looking at the univariate statistics obtained from R studio, we can further describe the shape:

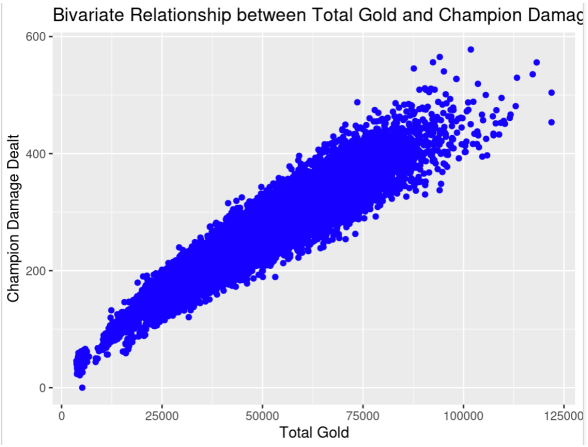
- Minimum: 0
- 1st Quartile: 43205
- Median: 65013
- Mean: 69746
- 3rd quartile: 90160
- Maximum: 333957

As a result of the distribution not being normal, I performed a square root transformation. As shown on the 2nd histogram above, indeed shows normality improvement in the distribution.

SDS 320E – Preliminary Analysis Report

Investigation of Numeric Explanatory Variable





The left histogram displays an overall normal distribution with a slight right skewness, hence no required transformation.

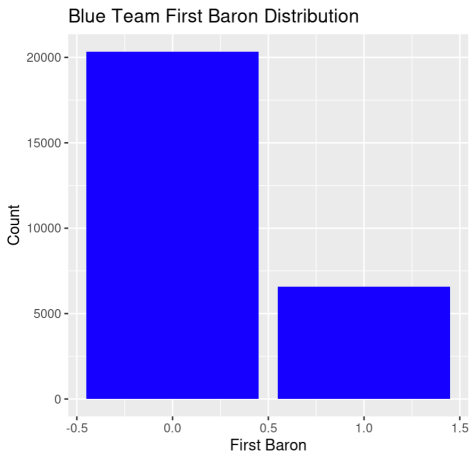
Looking at the univariate statistics obtained from R studio, we can further describe the shape:

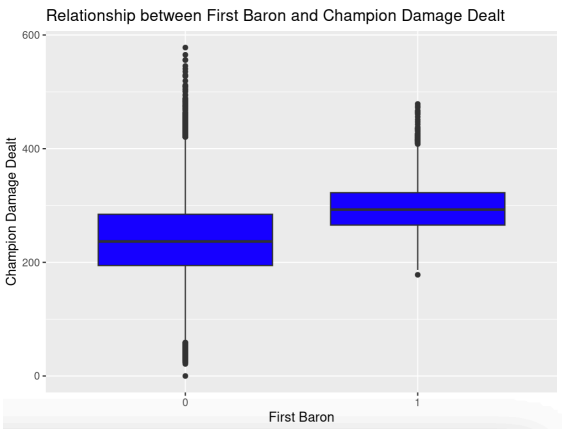
- Minimum: 3704
- 1st Quartile: 37191
- Median: 48104
- Mean: 48169
- 3rd Quartile: 58656
- Maximum: 121920

From the above statistics, we can infer that the distribution is centered around the median and mean (around 48000) given that the range is 0 – 125000, which indicates a broad span of values.

According to the right graph there seems to be a positive correlation between total gold and the champion damage dealt. Notably, the spread of data increases as total gold increase, suggesting potential heteroscedasticity in the relationship.

Investigation of Categorical Explanatory Variable





Based on the left plot, displaying the counts of no first baron (0) and yes first baron (1), we observe frequency of each category.

From our statistical calculations in R studio and looking at the plot, we note that for:

- Category "0" (no first baron): count = 20337
- Category "1" (yes first baron): count = 6567.

From the second plot (bivariate graph), we can see that teams that secure the first baron kill during a match tend to have a more concentrated and narrower distribution of 'total damage dealt' around a higher central tendency than matches where a team does not.

R Code

```
league <- read.csv('Challenger_Ranked_Games.csv')

league_subset = league[, c('gameId', 'blueFirstBaron', 'blueTotalGold',
'blueChampionDamageDealt')]

### RESPONSE VARIABLE 'blueChampionDamageDealt':

# Visualize the marginal distribution of response variable
ggplot(league_subset, aes(x = blueChampionDamageDealt)) + geom_histogram(binwidth = 1000, fill =
"blue", color = "black") + labs(title = "Blue Team Champion Damage Dealt Distribution",x =
"Champion Damage Dealt",y = "Frequency")

# Calculate univariate statistics for response variable
summary(league_subset$blueChampionDamageDealt)

# Apply a square root transformation to response variable
league_subset$sqrt_blueChampionDamageDealt <-sqrt(league_subset$blueChampionDamageDealt)

# Create a histogram of the transformed data of response variable
ggplot(league_subset, aes(x = sqrt_blueChampionDamageDealt)) + geom_histogram(binwidth = 50, fill
= "blue", color = "black") + labs(title = "Sqrt-Transformed Blue Team Champion Damage Dealt
Distribution",x = "Square Root(Champion Damage Dealt)",y = "Frequency")

### NUMERIC EXPLANATORY VARIABLE 'blueTotalGold':

# Create a histogram to visualize the marginal distribution of 'blueTotalGold'
ggplot(league_subset, aes(x = blueTotalGold)) + geom_histogram(binwidth = 1000, fill = "blue",
color = "black") + labs(title = "Blue Team Total Gold Distribution",x = "Total Gold",y =
"Frequency")

# Calculate univariate statistics for 'blueTotalGold'
summary(league_subset$blueTotalGold)

# Visualize the relationship between 'blueTotalGold' and 'sqrt_blueChampionDamageDealt'
ggplot(league_subset, aes(x = blueTotalGold, y = sqrt_blueChampionDamageDealt)) +
geom_point(color = "blue") + labs(title = "Bivariate Relationship between Total Gold and Champion
Damage Dealt",x = "Total Gold",y = "Champion Damage Dealt")

### CATEGORICAL EXPLANATORY VARIABLE 'blueFirstBaron':

# Get counts for each category
category_counts <- table(league_subset$blueFirstBaron)
print(category_counts)

# Create a bar plot to visualize the marginal distribution of 'blueFirstBaron'
ggplot(league_subset, aes(x = blueFirstBaron)) + geom_bar(fill = "blue") + labs(title = "Blue
Team First Baron Distribution", x = "First Baron", y = "Count")

# Visualize the relationship between 'blueFirstBaron' and 'sqrt_blueChampionDamageDealt'
ggplot(league_subset, aes(x = factor(blueFirstBaron), y = sqrt_blueChampionDamageDealt)) +
geom_boxplot(fill = "blue") + labs(title = "Relationship between First Baron and Champion Damage
Dealt", x = "First Baron", y = "Champion Damage Dealt")
```