Aaron Melton
Miguel Logarta

Investigation of Win Rate Difference Between Melee and Ranged Champions

Introduction

For our Hypothesis Testing Project, we decided to collect data from a popular esports video game, League of Legends. In this analysis, we aim to determine if there is a statistically significant difference in the win rate between melee and ranged characters within all characters/champions in the game. The data used for this investigation is sourced from blitz.gg, using the latest win rates of the champions in the Platinum+ rank division on May 24th, 2023.

Data Transparency

Our main data source:
https://blitz.gg/lol/champions/overview?queue=RANKED_SOLO_5X5&region=WORLD&tier=PLATINUM_PLUS

Complementary data source:
https://leagueoflegends.fandom.com/wiki/List_of_champions/Basic_attacks

The data used in this analysis was obtained from blitz.gg, a reputable source for League of Legends statistics. Blitz.gg uses an internal API provided by Riot Games, the creator of League of Legends, for use in data collection and statistics. Our dataset includes the latest win rates of champions in the Platinum+ rank division for Ranked Solo mode.

We used a python script to scrape data off the webpage for the top 200 entries. It should contain most champion data, including which lane they play in. To determine whether the champion was a ranged, or melee attack type champion, we had to use a complementary data source. This complementary data source contained character statistics that we scraped using python and mapped the data values to the champions names'.

Problem Statement

We want to investigate if there is a significant difference in win rate between melee and ranged champions within all champions in the game regardless of role. By examining this difference, we can gain insights into the performance and viability of different attack types in all roles of the game.

Population and Sample:

The population of interest for this analysis are all champions in the Plantium+ rank division. From this population, we have a sample consisting of the top 200 champions listed in the table above. Originally our population was the top 10 champions in the mid lane role, however that

sample was too small. Even if we used the whole table of mid lane champions, it only amounted to 52 entries. We felt that analyzing all 200 champions in all roles would lead to better results.

Populations:

- The population of win rates of all melee champions in all roles
- The population of win rates of all ranged champions in all roles

Samples:

- μ_1: A random sample of size $n_1 = 50$ for all melee champions
- μ_2: A random sample of size $n_2 = 50$ for all ranged champions

Limitations:

- Blitz.gg does not have all champion data for all ranks and all roles. There are 163 total characters in the game with 5 different roles (top, mid, adc, support, and jungle). That would make 815 rows of data. Since blitz.gg only provides a maximum of 200 rows of data, we will use hypothesis testing to measure our data.

Hypothesis and Setup

For our hypothesis, we will be using A/B hypothesis testing to determine if there is a significant difference in the average win rate between ranged and melee champions in all roles of the game.

Null hypothesis:

$H_0 = \mu_1 - \mu_2 = 0$

Our null hypothesis states that there is no difference in the true average win rates for ranged and melee champions

Alternative hypothesis:

$H_a = \mu_1 - \mu_2 > 0$

Our alternative hypothesis says that there is a significant difference in the true average win rates for ranged and melee champions. This means that your selection of a ranged or melee champion has a significant effect on your chances of winning a game.

Alpha Value / Skepticism:

We picked an alpha value / significance level of $\alpha = 0.05$ (corresponding to a 95% confidence level)

Assumptions:

- All of our sample data has n >= 40 (Satisfied)
- We do not know the true SD of either of our melee and ranged win rate populations (Satisfied)
- Samples are all independent (Satisfied). You cannot switch champions in the middle of a match.

Calculation

Initial Data:

After taking a random sample from our populations (melee and ranged attacks), we get:

- $n_1 = n_2 = 50$
- $Xbar_1$ (Sample mean of melee) = 0.50828
- $Xbar_2$ (Sample mean of ranged) = 0.50456
- SD_1 (Standard deviation of melee) = 0.01449566
- SD_2 (Standard deviation of ranged) = 0.01426879

Sample Data Set	
Melee Win Rates	0.504, 0.503, 0.519, 0.511, 0.527, 0.493, 0.500, 0.484, 0.496, 0.489, 0.507, 0.514, 0.523, 0.523, 0.519, 0.473, 0.496, 0.518, 0.510, 0.486, 0.515, 0.500, 0.506, 0.493, 0.535, 0.489, 0.519, 0.512, 0.495, 0.493, 0.474, 0.529, 0.511, 0.515, 0.520, 0.537, 0.503, 0.512, 0.508, 0.503, 0.530, 0.525, 0.506, 0.503, 0.512, 0.514, 0.522, 0.516, 0.515, 0.507
Ranged Win Rates	0.512, 0.497, 0.480, 0.508, 0.506, 0.510, 0.498, 0.511, 0.514, 0.528, 0.500, 0.510, 0.523, 0.509, 0.523, 0.479, 0.507, 0.510, 0.505, 0.498, 0.512, 0.508, 0.506, 0.489, 0.498, 0.491, 0.484, 0.515, 0.521, 0.517, 0.504, 0.494, 0.478, 0.498, 0.523, 0.471, 0.512, 0.533, 0.489, 0.490, 0.492, 0.508, 0.533, 0.503, 0.502, 0.489, 0.493, 0.509, 0.525, 0.513

Here is the R code:

Aaron Melton
Miguel Logarta

```r
setwd("C:/Users/potatochipse/Desktop/SFSU Classes/MATH 324/Final/miniProject2")

melee_statistics <- read.csv("meleeWinRates.csv")
ranged_statistics <- read.csv("rangedWinRates.csv")
sample_size <- 50

melee_sample <- sample(melee_statistics$winRate, sample_size, replace = FALSE)
ranged_sample <- sample(ranged_statistics$winRate, sample_size, replace = FALSE)

melee_sample
ranged_sample

```

Making samples

```r
# Samples I got
melee_sample <- c(0.504, 0.503, 0.519, 0.511, 0.527, 0.493, 0.500, 0.484, 0.496, 0.489, 0.507, 0.514, 0.523, 0.523, 0
                  0.486, 0.515, 0.500, 0.506, 0.493, 0.535, 0.489, 0.519, 0.512, 0.495, 0.493, 0.474, 0.529, 0.511, 0
                  0.508, 0.503, 0.530, 0.525, 0.506, 0.503, 0.512, 0.514, 0.522, 0.516, 0.515, 0.507)
ranged_sample <- c(0.512, 0.497, 0.480, 0.508, 0.506, 0.510, 0.498, 0.511, 0.514, 0.528, 0.500, 0.510, 0.523, 0.509,
                   0.498, 0.512, 0.508, 0.506, 0.489, 0.498, 0.491, 0.484, 0.515, 0.521, 0.517, 0.504, 0.494, 0.478,
                   0.489, 0.490, 0.492, 0.508, 0.533, 0.503, 0.502, 0.489, 0.493, 0.509, 0.525, 0.513)

xbar1 <- mean(melee_sample)
xbar2 <- mean(ranged_sample)
sd1 <- sd(melee_sample)
sd2 <- sd(ranged_sample)

```

Getting standard deviations

Standard Error

- $SE_1 = SD_1 / \text{sqrt}(n_1) = 0.002049996$
- $SE_2 = SD_2 / \text{sqrt}(n_2) = 0.002017912$

Calculating t-stat:

- t-stat = ((xbar1 - xbar2) - 0)/sqrt(se1^2+se2^2) = 1.293223

Calculating Degrees of Freedom:

- $df = ((SE_1 \text{ ^2} + SE_2 \text{ ^2})\text{^2})/((SE_1 \text{ ^4})/(n_1 - 1) + (SE_2 \text{ ^4})/(n_2 - 1)) = 97.97562$

Calculating the P-Value:

Looking back at our Hypothesis: $H_a = \mu_1 - \mu_2 > 0$, we will use the R function P-Value = 1 - pt(<t-stat>, df=<df>)

- P-Value = 0.09948745

Here is the R code:

```
32  tstat <- ((xbar1 - xbar2) - 0)/(sqrt(se1^2+se2^2))
33
34  tstat
35
36  df <- ((se1^2 + se2^2)^2)/((se1^4)/(sample_size - 1) + (se2^4)/(sample_size - 1))
37  df
38
39  pval <- 1 - pt(tstat, df=df)
40  pval
41
```

Calculating t-stat and P-Value

Conclusion

Since the P-Value = 0.09948745 > α = 0.05, we fail to reject the null hypothesis that there is no difference in the true average win rates for ranged and melee champions, and do not have enough evidence to conclude that the alternative hypothesis (that ranged and melee champions *do* have different win rates) is true.