Project Part 3

Abhinay Dommalapati

```
chess = read.csv("games 2.csv")
```

Question

Is there evidence that white's elo rating is siginifcantly different from black's elo rating in any given online chess game?

Data Description

An elo rating can be understood as a performance metric that fluctuatues based on how a player does against lower and higher rated opponents. A player who loses to a lower rated opponent will likely lose a significant amount of elo and a player who beats a higher rated opponent will likely gain a signicant amount of elo. This "elo" tends to vary between 0 and 2500. Also, the terms "elo" and "rating" can be used interchangeably.

The chess dataset I used in Project Part 1 has 20,000 rows where each row represents an online chess game that holds many variables such as ID, Rated, Turns, Victory Status, Winner, etc. However, for this report, I have selected only the columns in the chess dataframe that I am concerned with. I have only selected white_rating and black_rating. These are the two vectors that I will use later in my paired t test. Moreover, my sample is a random 500 rows in this column-wise subsetted dataframe and my population is the entire 20,000 rows of white_rating and black_rating.

```
chess_rated = chess %>%
  dplyr::select(10, 12)
head(chess_rated)
```

```
##   white_rating black_rating
## 1         1500         1191
## 2         1322         1261
## 3         1496         1500
## 4         1439         1454
## 5         1523         1469
## 6         1250         1002
```

Data Relevance

In online chess games, players are matched based on their elo rating. Typically, players will be matched with other players of similar elo (Reference 4). This is to prevent an unfair advantage to either player and ensure an "even" game. For instance, a player with a 1400 elo will likely be matched with someone with an elo of 1300-1500. On the other hand, a player with a 1400 elo is highly unlikely to be matched with someone with an elo of 500 or 2300. Therefore, my question tries to find if there is a big difference amongst white and black's elo ratings within my sampled proportion for any given game. I can then apply my answer found from the sample to a generalized population.

Generalization

If the mean of the differences is low and the p-value is greater than 0.05, then my hypothesis that the system pairs players of similar elo holds true. This means I will accept the null hypothesis that the true difference in means is equal to or close to 0. Thus, I can conclude that the differences in ratings are statistically negligible. If I can make this conclusion for my sample, then I can make the following generalization for all online chess games: the match-making algorithm follows a seemingly fair system that pairs players based on their past performance. However, the system is not entirely cheater-proof. This system encourages "sandbaggers" - players who deceptively try to reduce their elo on purpose so that they can play in high-stakes online tournaments that involve low-elo players (Reference 3). In fact, many other e-sports such as Counter-Strike:Global Offensive and League of Legends face the same problem of high-skill players purposely reducing their elo so that they can participate in low-skill tournmanets and win money. I hope that my findings in this test will adequately highlight this problem.

Test Selection

I chose the paired-t test because I am testing the difference of two means by using the differences in pairs of data points that are dependent. White rating and black rating are dependent because as stated before it is highly likely for a player to be paired with someone of similar elo. I also chose the paired-t test because the two vectors are normally distributed. In relation to all online chess games on lichess.org, white rating and black rating are both normally distributed with an average rating of about 1500 (Reference 2). Lastly, I chose the paired-t test because my sample is big enough that the Central Limit Theorem holds. I have drawn a sample of 500 from my population of 20,000 games. CLT holds if $n >= 30$ (Reference 1) which is true here.

Test

```
sample_chess = chess_rated[sample(nrow(chess), 500),]
white_rating = sample_chess$white_rating
black_rating = sample_chess$black_rating
t.test(white_rating, black_rating, mu=0, alternative = "two.sided", paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  white_rating and black_rating
## t = 0.24159, df = 499, p-value = 0.8092
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.6304  23.8544
## sample estimates:
## mean of the differences
##                   2.612
```

Test Conclusions

Because the p-value is significantly greater than 0.05, I can accept the null hypothesis: the difference between the ratings of white and black in a given online chess game is negligible. Also note that the mean of the differences is very low. When these conclusions are generalized to either the population of 20,000 chess games or the entire current database of online chess games, one can make the conclusion that the online chess match-making system pairs players of similar elo. Although the chess elo system is a comprehensive way of evaluating a player's performance and has been adopted by many other video games to rank their players such as Counter-Strike:Global Offensive and League of Legends (Reference 5), the system has its flaws. This statistical report shows how someone can take advantage of the system. A sandbagger could deceptively lose games on purpose so that he/she can participate in high-stakes tournaments that involve cash-based prizes. Although his/her elo is low enough for him/her to be able to play in this tournament marketed for low-elo players, he/she will be able to easily beat everyone in the tournament because his/her real skill is much higher compared to the other competitive players. This has become a huge issue not only in the game of chess, but also in other competitive video games that have adopted the chess elo system. Players known as "smurfs" take advantage of the system to participate in high-stakes tournaments involving low-skill players to win cash or other prizes (Reference 5). This statistical report shows this problem in the competitive scene of e-sports. Either this current elo system needs to be changed to hinder sandbaggers and smurfs or an entirely new system of performance metrics needs to be adopted.

References

1. https://www.kaggle.com/datasnaek/chess
2. http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Probability/BS704_Probability12.html
3. https://gist.github.com/ornicar/2a402e46d05074f880d6
4. https://en.wikipedia.org/wiki/Cheating_in_chess#Sandbagging
5. https://www.reddit.com/r/chess/comments/4sva7d/lichess_matchmaking/
6. https://english.stackexchange.com/questions/17209/where-does-the-term-smurfing-come-from