

# The Significance of the First Move Advantage in Chess

Labib Chowdhury

09/12/2020

Source Code: <https://github.com/labib-c/first-move-advantage-analysis>

## Abstract

## Keywords

## Introduction

In the world of Chess, one question which remains open is whether or not White has a significant advantage due to having the first move. The player who controls the White pieces begins the game on attack while the player who controls the Black pieces must defend, which is where this apparent advantage originates. If both players play identically and symmetrically, then White will win a significant majority of the time. The issue with this is that players do not play identically or symmetrically most of the time, in fact White wins only 37% of the time compared to Black's 28%, according to the chessgames.com database of chess games (Chess Statistics). This paper aims to analyze whether White has a significant advantage in Chess due to having the first move, and this will be done through considering multiple factors including rating level, game type, and first move.

Chess remains as one of the most popular games of all time, claiming 600 million fans worldwide (Cowen, 2018) hence the significance of Chess cannot be understated. Analyzing whether White has a definite advantage over Black will help novice and expert players understand the game better and modify the strategies utilized by every player.

To identify whether White has a significant advantage over Black in Chess, a logistic regression model will be applied over the LiChess Chess Game Dataset (Jolly, 2017). Logistic regression is best used to describe the relationships between one dependent binary outcome variable against one or more independent predictor variables. The logistic regression model will also provide estimates identifying how a certain predictor affects the outcome in an intuitive manner.

An outcome variable of the form `white_won` will be used in the logistic model, where 1 represents games where the White player won and 0 represents games where White lost. The Methodology section (Section 2) will dive deeper into which predictors were chosen and how the logistic regression model was developed. Results of the model can be seen in Section 3, Results, and further discussion on the outcome of the model will be highlighted in Section 4, Discussion.

## Methodology

Data: talk about how this is an experiment since Black and White are randomized

## **Results**

## **Discussion**

## **References**

Jolly, M. (2017). “Chess Game Dataset (Lichess)”. Retrieved December 7, 2020 from <https://www.kaggle.com/datasnaek/chess>.

Chess Statistics, Retrieved December 7, 2020 from [www.chessgames.com/chessstats.html](http://www.chessgames.com/chessstats.html).

Cowen, Tyler. (13 Nov. 2018). “Chess Is the Killer App; How and Why a 1,500-Year-Old Game Has Conquered the Internet.” Bloomberg, Retrieved December 7, 2020 from [www.bloomberg.com/opinion/articles/2018-11-13/world-chess-championship-2018-is-made-for-the-internet](http://www.bloomberg.com/opinion/articles/2018-11-13/world-chess-championship-2018-is-made-for-the-internet).