Statistical Theory
Chess Dataset Analysis

Dor Boker, Itamar Nakar

I.D: , 325829000

Email: , itamar.nakar@gmail.com

I. INTRODUCTION

Chess is one of humanity's oldest board games, played by two players on an 8×8 grid. Each player, controlling 16 pieces of their color, aims to checkmate the opponent's King, making it impossible for the King to escape capture. Unlike many games, chess does not involve luck or hidden information; the outcome is determined solely by the players' knowledge, strategy, and analytical skills.

The length of a chess match is influenced by several factors, including the players' skill levels, strategic choices, and in-game dynamics. In this project, we investigate the relationship between player ratings and the duration of chess matches, aiming to determine whether higher-rated players tend to play shorter or longer games. In addition to player ratings, we examine other variables that may impact match length, such as opening strategies, game outcomes (win, loss, draw), and time controls.

Our dataset comprises 19,113 games from Lichess, an online chess platform, each described by 16 features. The most relevant features to our analysis are:

- **Turns**: A positive integer representing the length of the game, where one turn includes a move by both White and Black.
- **ELO Rating**: A system that quantifies player skill, widely used by FIDE and online chess platforms, which we use to assess its correlation with match length.
- **Time Controls**: Competitive games are categorized into four main time control classes: Bullet, Blitz, Rapid, and Classical, based on initial clock times and increments.
- **Winner and Victory Status**: These columns capture whether White or Black won and the method of victory (checkmate, resignation, draw).

Through statistical analysis, we aim to understand how these factors influence the duration of chess matches and to provide insights into the relationship between skill level and game dynamics.