Proposal: Analyzing Mens Tennis Grand Slam Performance

Kaitlyn Shevlin
Department of Statistics
University of Connecticut

October 9, 2022

Introduction There are four Grand Slam tournaments in tennis: the Australian Open, the French Open, Wimbledon, and the US Open, spanning from January to September **?**. These notourious tournaments feature the world's best tennis players and are each played over a course of two weeks. As with every sport, there are a variety of controllable, regulated factors as well as uncontrollable, extraneous factors that affect the event. For example, one of these is the type of court that the matches are played on, which vary at each tournament. This paper will evaluate Mens Grand Slam tennis championship matches. I have chosen this topic because tennis is an interest of mine and I am curious to further explore the different variables that impact player performance.

Specific Aims Using this data set, I will conduct exploratory analysis to determine which factors, if any, are useful in predicting the winner of a Grand Slam championship match. This includes examining if ATP ranking is indicative of who will win the match. In addition, I will conduct further research in attempt to determine if certain multi-title holders perform better at different tournaments due to the surface, as well as if being right handed or left handed aids performance at certain tournaments. There is the idea of home court advantage in all of sports, so it will be interesting to examine if that holds true in tennis Grand Slams, as well as further delving into the notion that Spanish players perform better at the French Open- which may be explained by Spanish players grow up practicing on clay courts. The objective of the analysis is to provide a better understanding of how certain factors affect players as well as potential predictive modeling.

Data There are 289 observations with 4 per year for each major Grand Slam with the exception of Wimbledon in 2020 as it was cancelled due to COVID-19. The dataset includes YEAR: year tournament started; TOURNAMENT: tournament type; WINNER: winner's name; RUNNER UP: runner-up's name; WINNER ATP RANKING: winner's ATP ranking, RUNNER UP ATP RANKING: runner-up's ATP Ranking; WINNER LEFT OR RIGHT HANDED: whether the winner is right or left handed; TOURNAMENT SURFACE: surface tournament is played on; and WINNER PRIZE: prize money won at each tournament in local

denominations. For the purpose of this research, winner prize money will not be considered. ATP stands for Association of Tennis Professionals and the ATP ranking is a player's ranking based on total points earned in 19 official ATP-certified men's events. It is capped at 19 events so if a player goes over, their best 19 results are counted **?**. In this study, the ATP ranking is the player's ranking going into the tournament.

Research Design and Methods I will look at a mixed-effect models which will allow for looking at a mass of information with different players to estimate the effects of common variables. Players appear in multiple Grand Slams throughout the years so they appear multiple times in the data. I will also hone in on individualized player models to see how specific factors affect them and look for trends among their matches.

Discussion I am expecting to find that factors such as ATP ranking and court type have some impact on who the winner is. I would expect that players are used to one court type over another and that has some impact on their performance. In addition, typically players of a higher rank have earned that title for a reason and are favored to win. Players' dominant hand also has an impact as it affects the spin and motion on the ball **?**

Conclusion This research project will provide insight into what elements affect players or a specific player in the four major tennis Grand Slams. Although analysis has been done on various aspects of this data, my research will look at data from a wider range of dates as well as a different combination of factors. The aim is to corroborate previous research done on this project and better understand how different factors affect different players and overall observations.