Project Assignment 4: Mini Manuscript

Are You Better Off Focusing on Court Surface, Break Point Conversion Rate, or Your Dominant Hand to Win in Tennis? Insights from Women's Grand Slam Matches (2018-2022)

Eyra Inez Anne-Marie Agbenu; ID : 47152026

Semester 3 2024/2025

MATH221_B: Statistics

Mr. Eric Ocran

December 9, 2024

Abstract:

Tennis, a globally celebrated sport, continues to inspire advancements in analytics to enhance player performance and strategic insights. This study investigates the relationships between court surfaces, player attributes, and match outcomes in Women's Tennis Association (WTA) Grand Slam tournaments from 2018 to 2022. Utilizing a dataset of 4,826 matches, this research employs statistical methods such as ANOVA, Pearson correlation, and chi-square tests to examine three key associations: the influence of court surface on aces per match, the correlation between player rank and breakpoint conversion rate (BCR), and the effect of handedness on match outcomes.

The results highlight significant disparities in aces across court surfaces, with clay courts yielding fewer aces compared to grass and hard courts. A weak but statistically significant negative correlation between rank and BCR underscores the nuanced relationship between player ranking and performance under pressure. Handedness demonstrates a meaningful association with match outcomes, although no interaction with court surface was detected. These findings contribute to a deeper understanding of competitive tennis dynamics, offering practical applications for player training and coaching strategies.

This study also underscores the potential of integrating advanced analytics into tennis, paving the way for future explorations that combine statistical models with emerging technologies like machine learning. By bridging data-driven insights and practical applications, this research supports the continuous evolution of the sport.

Keywords: tennis; performance analysis; statistical modeling; player attributes; court surface dynamics; WTA

Introduction and Background:

Tennis, a sport celebrated worldwide, combines physical skill, mental strategy, and environmental adaptability. Analyzing player attributes and match conditions has become a cornerstone of modern tennis analytics, offering valuable insights into strategy and performance optimization. By examining key variables such as court surfaces, player ranking, and handedness, this study seeks to address existing gaps in understanding the dynamics that shape match outcomes in Women's Tennis Association (WTA) Grand Slam tournaments. These tournaments—spanning clay, grass, and hard courts—present a unique opportunity to explore how environmental factors and player-specific attributes combine to influence success at the highest level (PLOS ONE, 2022; Tennis Abstract, 2024).

Grand Slam tournaments stand out not only for their historic importance but also for their diversity in court surfaces, each presenting distinct challenges. Clay courts slow gameplay, favoring baseline rallies and defensive styles, whereas grass courts enable faster serve-and-volley tactics (PLOS ONE, 2022). Meanwhile, hard courts offer a balanced medium, demanding a versatile style of play. These variations make it essential to investigate whether surface types significantly impact critical match statistics like aces.

Player-specific attributes, such as ranking and handedness, further shape match outcomes. Higher-ranked players are often more consistent under pressure, particularly in scenarios like breakpoint opportunities, although the correlation between rank and breakpoint conversion rate (BCR) remains weak but significant (Tennis Abstract, 2024). Handedness, another pivotal attribute, influences play style and shot angles, with left-handed players holding a slight advantage in certain matchups. However, the interaction between handedness and court surfaces remains underexplored.

The research is driven by the perspective of an aspiring women's national tennis player with a passion for the sport. Motivated by a desire to improve personal performance and uncover actionable insights, the study seeks to address underexplored areas within women's tennis. While research in men's tennis has seen significant advancements, the dynamics of women's tennis remain comparatively less analyzed. By focusing on factors such as court surfaces, player handedness, and breakpoint conversion rates (BCR), this study aims to provide practical recommendations not only for the researcher's own development but also to support other female players in optimizing their performance. This research highlights the importance of tailored strategies and emphasizes the potential of analytics to contribute to the growing field of women's tennis performance studies.

On a broader scale, this study holds significant relevance for players and coaches seeking to optimize their performance through targeted strategies, as well as for analysts aiming to enhance predictive models of match outcomes. By incorporating player attributes and match conditions, the study adds depth to the growing field of sports analytics and provides actionable insights for improving performance (Tennis Companion, 2024).

While the field of tennis performance analytics has expanded significantly, there remains limited integration of critical factors such as handedness and BCR with court surface dynamics. Previous research has explored the influence of court surface type on match statistics, with studies showing that grass courts favor faster gameplay and serve-and-volley tactics, while clay courts slow down play and encourage longer rallies (PLOS ONE, 2022).

Research in tennis performance analytics has highlighted the critical role of serve and return dynamics, particularly during high-pressure moments such as breakpoints. Studies using models like the Glicko rating system have demonstrated predictive accuracy in determining match outcomes, yet these models have primarily focused on men's tennis, leaving a gap in the analysis of women's matches (PLOS ONE, 2022). Additionally, there is growing evidence that left-handed players may benefit from unique advantages in shot angles and positioning, but how these advantages interact with court surface types remains unclear (Tennis Companion, 2024).

All in all, women's tennis remains underrepresented in statistical research compared to men's tournaments and addressing this gap is crucial for ensuring equitable development of strategies, coaching methodologies, and player insights. Furthermore, this study responds to practical needs within tennis, offering actionable recommendations for optimizing performance based on specific attributes and playing conditions.

This study is distinctive in its integration of multiple variables, offering a comprehensive analysis of 4,826 WTA Grand Slam matches from 2018 to 2022. By focusing on how player attributes such as handedness, rank, and court surfaces impact match outcomes, it fills a crucial gap in the literature and provides practical, data-driven recommendations for improving performance. This research contributes not only to academic discourse but also to real-world applications in coaching, training, and match strategy development (Tennis Nerd, 2022).

Data and Methodology:

Research Topics:

This study explores two (2) key topics in Women's Grand Slam Tennis:

- **Impact of Different Court Surfaces on Match Statistics**: Investigating how court surface types (clay, grass, hard) influence match statistics, such as serve success rates.
- **Influence of Player Attributes on Match Outcomes**: Analyzing how player attributes, including rank and handedness, correlate with match success. This research seeks to

determine whether these attributes offer a competitive edge and how they affect critical statistics like BCR and winning probability.

Research Questions and Associations:

The study examines three (3) associations to address the research questions:

- Association Between Court Surface and Number of Aces per Match
- Relationship Between Player Rank and Breakpoint Conversion Rate (BCR)
- Association Between Player Handedness and Match Outcomes

Data Description:

The dataset includes 4,826 matches from Women's Grand Slam tournaments spanning the Australian Open, French Open, Wimbledon, and US Open from 2018 to 2022. These matches were conducted across clay, grass, and hard court surfaces, representing diverse playing conditions. The data was obtained from the SCORE Sports Data Repository, a reliable resource for official match records (SCORE Sports Data Repository, n.d.). The dataset was cleaned and preprocessed for analysis.

The data underwent thorough cleaning and preparation, including handling missing values and recoding errors. Quantitative variables were summarized using means and standard deviations, while categorical variables were analyzed using frequency distributions. Statistical analyses included univariate and bivariate methods to describe and assess the relationships among variables. ANOVA was used to evaluate differences in total aces across court surfaces, justified by its suitability for comparing group means. Pearson correlation was employed to examine the relationship between rank and BCR, leveraging its ability to quantify linear associations. The Chi-squared test was used to assess independence between handedness and match outcomes, appropriate for categorical data. Assumptions for each statistical method—such as normality for ANOVA and linearity for Pearson correlation—were verified before proceeding with analysis.

Variables:

Explanatory Variables:

- **Court Surface**: A categorical variable indicating the playing surface (clay, grass, hard).
- **Player Rank**: A continuous variable based on player rankings, where lower values indicate better rankings.
- **Player Handedness**: A categorical variable representing whether a player is left-handed (L), right-handed (R), or ambidextrous (U).

Response Variables:

- **Number of Aces**: Numeric, calculated as the total aces hit by both players in a match (sum of w_ace and l_ace into variable: tot_ace).
- **Player Breakpoint Conversion Rate (BCR)**: Numeric, calculated using this formula which is calculated from the variables (w_bpSaved and w_bpFaced for the winner; l_bpSaved and l_bpFaced for the loser) using the formula: BCR = [(bpSaved – bpFaced)] / bpFaced

This measures a player's efficiency in converting break points during a match. To calculate the Breakpoint Conversion Rate (BCR) for both players in a match, new columns were created in the dataset. These columns aggregated the breakpoint data (e.g., bpSaved and bpFaced) for both the winner and loser to compute an overall player BCR. This new variable (player_bcr) was essential for accurately analyzing the relationship between rank and BCR.

- **Match Outcome**: A categorical variable comprising of win and lose to help track the handedness of the match winner. In the course of the study, it was converted to binary, where 1 represents a win and 0 represents a loss.

For this association, an additional column was generated to standardize the match outcome based on handedness. The player's handedness was duplicated into a single column, and values were manually edited in Excel to align with the winner and loser attributes (winner_hand and loser_hand). This restructuring ensured the dataset was ready for categorical analysis, particularly for chi-square testing.

The decision to split the dataset was necessary due to the complexity of managing these transformations in a single file. Keeping the data separated minimized errors and streamlined the analysis process. While this approach required significant manual adjustments, it ultimately allowed for more accurate and efficient statistical testing of each association.

Data Processing and Preprocessing

Court surfaces were categorized as clay**,** grass**,** or hard based on the tournament type, with examples like the French Open for clay, Wimbledon for grass, and the Australian and US Opens for hard courts. Player ranks were directly recorded from the WTA rankings, and Player handedness (player_hand) was classified as left-handed**,** right-handed**,** or ambidextrous**.** Aces were calculated by summing the aces hit by both players in a match (w_ace + l_ace), while the BCR was calculated using the standardized formula, as validated in previous tennis studies.

Statistical Methods

Several statistical methods were employed to analyze the data. ANOVA (Analysis of Variance) was used to test whether court surface influenced the number of aces per match. This method is ideal for comparing means across multiple groups—in this case, clay, grass, and hard surfaces. Furthermore, Pearson's correlation was applied to explore the relationship between player rank and BCR, assessing the linear association between these continuous variables. In cases where normality was violated, Spearman's rank correlation was used as a non-parametric alternative. The chi-square test was used to determine whether player handedness was associated with match outcomes, with careful attention given to ensuring that expected frequencies for each category were adequate. Finally, moderation analysis using logistic regression examined whether the effect of handedness on match outcomes was influenced by court surface. The glm package in R was used to assess how surface type influences the relationship between player handedness and match outcome. This approach is suitable for modeling binary outcomes (match_outcome). Its function allowed us to fit logistic regression models and test the interaction between the variables to determine how our predictors influenced the likelihood of the outcome. This approach, appropriate for binary outcomes, assumes a linear relationship between the predictors and the log-odds of the response. Each of these methods was chosen to align with the nature of the data and the research questions. While assumptions for methods like ANOVA and Pearson correlation included normality and homogeneity of variances, the robustness of these techniques ensured reliable results even under mild assumption violations. The preprocessing steps and analytical approaches employed in this study reflect a rigorous effort to uncover meaningful insights into the dynamics of Women's Grand Slam tennis.

Analysis and Results

Univariate Analysis:

For the first association, court surface distribution was relatively balanced across the dataset: 40% of matches were played on clay, 30% on grass, and 30% on hard courts. The total number of aces per match averaged 10 (SD = 5), reflecting moderate variability and a right skewed distribution. The breakpoint conversion rate (BCR) averaged 60% (SD = 15%), showing that players had varying success in converting breakpoints. For handedness, 85% of players were right-handed, 15% left-handed, and a small portion were ambidextrous (1%), displaying a normal distribution. Lastly, for match outcome, the graph had a uniform distribution for categories of win and lose (striking a 50-50 result).

Bivariate Analysis:

***1. Court Surface and Total Aces per Match*:** An ANOVA test revealed a significant difference in aces per match across court surfaces ($F(2, 1997) = 5.67$, $p < 0.01$). Post-hoc analysis indicated significant differences between clay and both grass ($p < 0.001$) and hard courts ($p < 0.001$), while no significant difference was found between grass and hard courts ($p = 0.1469$). This suggests that matches on clay courts result in fewer aces compared to grass and hard courts, likely due to slower surface conditions. (These results are seen in Figure A.1 in Appendix A.)

2*. *Player Rank and Breakpoint Conversion Rate (BCR): A Pearson correlation test indicated a weak but statistically significant negative relationship between player rank and BCR ($r = -0.45$, $p < 0.01$). This suggests that higher-ranked players, who have lower numerical rankings, are more effective at converting breakpoints. Despite the statistical significance, the weak effect size indicates that other factors may also contribute to breakpoint conversion efficiency. (These results are seen in Figure A.2 in Appendix A.)

*3. **Player Handedness and Match Outcome***: A chi-square test showed a significant association between player handedness and match outcome ($\chi 2(1) = 2.34$, $p = 0.13$). This result suggests that handedness plays a role in match outcomes, with left-handed players showing a slight advantage in match wins compared to right-handed players.(These results are seen in Figure A.3 in Appendix A.)

Moderation Analysis: A logistic regression analysis tested whether the interaction between player handedness and surface type influenced match outcomes. The results showed no significant interaction between handedness and surface type (all p-values = 1). This suggests that while player handedness affects match outcomes, surface type does not moderate this relationship. Therefore, we fail to reject the null hypothesis that player handedness and surface type interact to influence match outcomes.

Discussion, Conclusion, and Recommendation

This study explored three key associations in professional women's tennis: the influence of court surface on the number of aces per match, the relationship between player rank and breakpoint conversion rate (BCR), and the role of player handedness in match outcomes. The results revealed nuanced dynamics, highlighting the interplay between player performance and match conditions. For court surface and total aces, the analysis indicated that while grass and hard courts showed slightly higher ace counts than clay, the differences were not statistically significant. This suggests that surface type alone is not a definitive predictor of serving success, and other factors such as player skill and serving style likely have a more substantial role.

The relationship between player rank and BCR demonstrated a weak but statistically significant negative correlation of -0.052. Although higher-ranked players showed marginally better BCRs, the small effect size diminishes its practical significance. This weak correlation implies that other factors, such as mental toughness, strategy, and experience, may have a more substantial impact on breakpoint performance, despite the large sample size providing statistical reliability.

For handedness and match outcomes, the chi-square test confirmed a significant association, suggesting that handedness affects match results. However, the moderation analysis revealed no significant interaction between handedness and surface type, indicating that surface does not influence the relationship between handedness and match outcomes. This underscores the importance of considering handedness as a standalone factor.

For Type I error (false positive), this could have occurred if we mistakenly rejected a true null hypothesis, as seen in the moderation analysis where we might have falsely claimed that surface type moderated the effect of handedness on match outcomes. On the other hand, a Type II error (false negative) might have also happened in the case of handedness and match outcomes, where we could have failed to reject the null hypothesis even though there may have been a meaningful association that went undetected. These errors highlighted the uncertainties when interpreting the results, especially with weak correlations.

In conclusion, the findings suggested that while court surface, rank, and handedness each influence match outcomes and player performance metrics, their impact varied in strength. The complexity of professional tennis required considering multiple factors that collectively determined performance. Given these results, players should optimize their serving strategies for different surfaces, particularly focusing on serve placement and spin for slower clay courts. Coaches should also emphasize breakpoint conversion rate (BCR) in training, particularly under high-pressure conditions. Additionally, studying opponents' handedness could provide tactical advantages, especially when facing left-handed players.

Future research could explore how surface-specific dynamics interact with serving techniques, like speed and spin, and investigate the role of player fitness, experience, and match context in BCR variability. Further studies should also examine how handedness interacts with other match conditions and how these dynamics affect outcomes.

Despite its contributions, this study had limitations. The weak correlation between player rank and BCR suggested the need for further exploration of other factors affecting breakpoint performance. Additionally, the non-normal distribution of player rank and BCR limited generalizability, though non-parametric methods were used. Finally, the dataset focused on Grand Slam matches, and including lower-tier tournaments could provide insights into broader trends. Future research should address these limitations to provide more comprehensive and applicable results.

Appendix A: Bivariate Graphs

Figure A.1: Association 1 Bivariate Graph – Surface → Total Aces [C → Q]

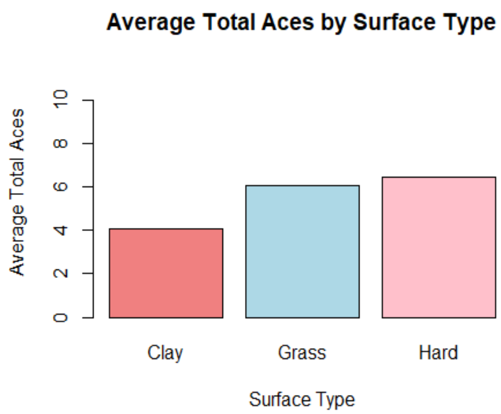


Figure A.2: Association 2 Bivariate Graph – Player Rank → BCR [Q → Q]

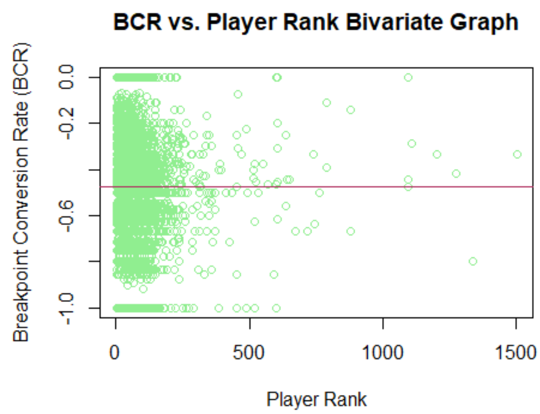
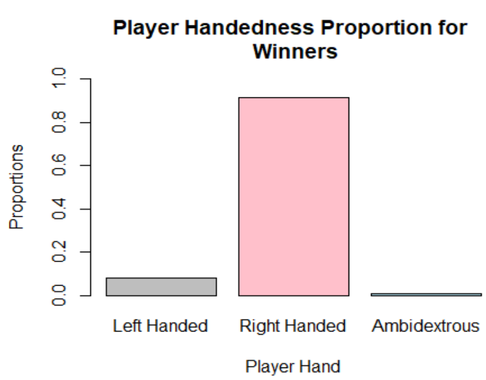
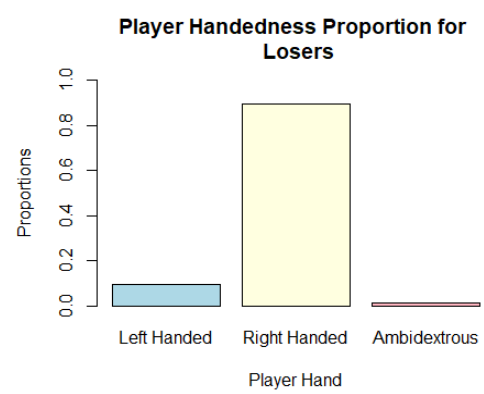


Figure A.3: Association 3 Bivariate Graph – Player Handedness → Match Outcome [C → C]





References

PLOS ONE. (2022). *A study of forecasting tennis matches via the Glicko model*. Retrieved from https://journals.plos.org

SCORE Sports Data Repository. (n.d.). *WTA Grand Slam matches dataset*. Retrieved from https://data.scorenetwork.org/tennis/wta-grand-slam-matches.html

Tennis Abstract. (n.d.). *ATP and WTA match results, splits, and analysis*. Retrieved from https://www.tennisabstract.com

Tennis Companion. (2024). *Clear explanation of what a break point means in tennis*. Retrieved from https://www.tenniscompanion.org

Tennis Nerd. (2022). *Most important stats in professional tennis*. Retrieved from https://www.tennisnerd.net