Statistical Theory - Final Project

Ron Libman - 211540265
Omer Shadmi - 327722575

September 29, 2024

Abstract

This study explores the hypothesis that left footed footballers have distinctive advantages in specific aspects compared to their right footed counterparts. Utilizing a dataset from FIFA, which includes performance ratings across multiple features for ∼18,000 players, this research applies a variety of statistical tests and machine learning techniques to analyze and compare the abilities of left footed and right footed players. Detailed examinations using statistical tests reveal significant differences in key playing attributes such as dribbling, vision, and stability. The study also employs machine learning algorithms to evaluate the predictive importance of being left footed on players' attributes, supplemented by explainability analysis for feature impact transparency. The findings suggest that left footed players demonstrate enhanced capabilities in certain areas, contributing to the hypothesis of a unique left footed advantage in football. However, the results could be misleading and influenced by several confounding variables, highlighting the complexity of attributing performance to foot preference alone. The code for this study is available at `https://github.com/OmerShadmi/Statistical-Theory---Project`

1 Introduction

The debate on whether left footed footballers have distinctive advantages in professional football is not only a subject of fan speculation but also an area of growing academic interest. Historically, left footed players like Lionel Messi and Diego Maradona have captivated audiences with their exceptional skills, leading to a belief that foot preference in sports comes with certain advantages and disadvantages.

Recent studies have aimed to quantify observations by examining data across various leagues and competitions. One such study was conducted by Bozkurt et al. (2018) [1], and compared technical skills of youth soccer players according to foot preference. The study concluded that there were not significant differences between left and right footed players in the technical skills ($p>0.05$), and recommended further studies to clarify the comparing of technical skills of players according to preferred foot parameters.

Another study by Kocaoğlu et al. (2023) [2], investigated the effects of foot preference on postural control in soccer players. The findings revealed no significant differences in postural control between right footed and left footed players. However, they suggested that further research is needed.

For now, empirical research on this topic remains divided. It can be argued that the perceived advantage may be more reflective of cultural and strategic biases within sports coaching and talent development, rather than physiological or biomechanical benefits. Another possible reason for a left footed advantage could be the rarity of left footedness, which might disrupt typical defensive expectations.

Our study contributes to the topic by analyzing a dataset provided by FIFA, that includes a wide range of playing features across ∼18,000 football players. By employing statistical methods and advanced

machine learning techniques, this research seeks to find out how foot preference might correlate with specific football skills and overall player effectiveness, offering new insights into an old debate.

2 Methods

The key steps included data preprocessing, exploratory analysis, hypothesis testing, and machine learning model implementation. Below is a breakdown of the methods employed.

2.1 Data Preprocessing and Exploration

The initial dataset was loaded into Python using the **pandas** library, followed by basic exploration to understand its structure. Key features included age, overall rating, crossing, dribbling, ball control, and vision. We explored the distribution of each feature and checked for missing values (there were none in the features that were relevant to us). Since preferred foot is our target variable, we visualized the distribution of left footed vs. right footed players using bar plots created with the **seaborn** and **matplotlib** libraries. Additionally, we computed and visualized basic statistics (mean, standard deviation, etc.) of key features for both groups to get a clear comparison of how player attributes varied with foot preference.

2.2 Statistical Testing

To assess our hypothesis we used 3 different statistical tests:

- **Welch's t-test** - a parametric test used to compare the means of two independent groups. It is an adaptation of the Student's t-test and is used when the two groups have unequal variances and / or unequal sample sizes.
- **Mann-Whitney U-test** - a non-parametric test used to compare the distributions of two independent groups. It tests whether one group tends to have larger values than the other, rather than comparing means specifically. It is especially useful when the data do not meet the assumptions of parametric tests, such as normality.
- **Kolmogorov-Smirnov test**, which compares the distributions of two samples to determine if they come from the same underlying distribution.

Results from these tests were interpreted based on **p-values**, with a significance threshold of 0.05. These tests provided statistical evidence to our hypothesis. We have also drawn confidence intervals for the means of each preferred foot group.

In particular, we compared the differences in key performance metrics across various position and height groups. This was in order to try and somewhat neutralize the effect of other variables and make the tests as fair and unbiased as possible.

2.3 Machine Learning Analysis

To further investigate our hypothesis we employed **XGBoost**, a gradient boosting algorithm, using it to predict key features based on various player attributes. The features considered in the model included **foot preference** and other physical and technical attributes. The dataset was split into training and test sets, with 90% of the data used for training and 10% for testing. To understand the contribution of foot preference to the model's predictions, we evaluated the results in 2 ways:

- **Feature importance** of every attribute - a metric that quantifies how much each feature contributes to the model's accuracy.

- **SHAP values** - a method used to explain the contribution of each feature to a model's prediction, showing how much each feature influences the outcome.

These evaluation methods provided insights into the impact of each feature on the model's predictions and allowed us to understand whether a player's preferred foot was impactful on his skills and attributes.

3 Results

3.1 Comparing Key Attributes: Left-Footed vs Right-Footed Players

Our study aims to determine whether left footed players demonstrate superior abilities in key attributes compared to right footed players. We focused on four technical skills: **crossing**, **dribbling**, **curve**, and **free kick accuracy**, which are generally seen as requiring higher levels of precision and technique. We used a combination of statistical hypothesis testing and machine learning techniques to analyze the data. The **null hypothesis** in each test we conducted was that there is not difference between left and right footed players in each category.

3.1.1 Crossing

Crossing, an important skill for wingers and attacking players, showed one of the strongest differences between left footed and right footed players. The mean crossing score for left footed players was significantly higher, at 56.73, compared to 47.57 for right footed players. Welch's t-test, which accounts for unequal variances, yielded a p-value of less than 0.01, confirming the statistical significance of this difference.

The non-parametric Mann-Whitney U-test corroborated this result, with left footed players tending to occupy the higher percentiles of crossing ratings. The Kolmogorov-Smirnov test further revealed a difference in the distribution shape, with left-footed players showing a higher concentration of higher crossing scores.

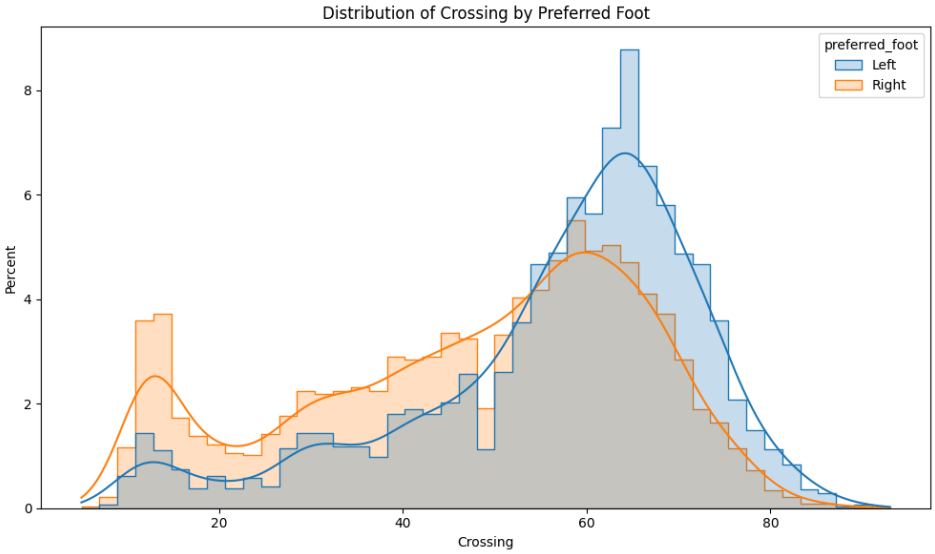

Figure 1: Distribution of crossing scores for left-footed and right-footed players. We can also notice a higher concentration of left-footed players in the higher values of crossing (above 60).

3.1.2 Dribbling

Dribbling, another skill vital to attacking play, also favored left footed players, with a mean score of 59.98 compared to 53.86 for right footed players. Welch's t-test produced a p-value of 2.8*E-94, indicating a statistically significant difference. The Kolmogorov-Smirnov test detected notable divergence in the distributions, with left footed players once again occupying higher percentiles in terms of performance. The Mann-Whitney U-test similarly supported these findings.

3.1.3 Curve

The curve, often essential for bending passes or shots around defenders, showed a clear advantage for left footed players. The mean curve rating for left footed players was 52.58, compared to 45.44 for right footed players. The Welch's t-test returned a p-value of less than 0.05, indicating that the difference was statistically significant. The non-parametric Mann-Whitney U-test yielded similar results, confirming that left-footed players tend to have higher curve ratings.

3.1.4 Free Kick Accuracy

Free kick accuracy, one of the most specialized skills in football, also showed a statistically significant advantage for left footed players, with a mean score of 47.57 compared to 41.21 for right footed players. Welch's t-test provided strong evidence of this difference (p less than 0.05). Additionally, the Kolmogorov-Smirnov test indicated a notable difference in the distribution of scores between left and right footed players, suggesting that left footed players generally perform better in free kick situations.

Attribute	**Welch T-Test**		**Mann-Whitney U Test**		**K-S Test**	
	t_statistic	**p_value**	**U_statistic**	**p_value**	**KS_statistic**	**p_value**
Crossing	30.91055	4.98921e-198	37385070.5	2.34651e-190	0.22981	3.42472e-149
Dribbling	20.85347	2.80666e-94	33642770.0	2.13786e-62	0.12764	5.95569e-46
Curve	22.78192	5.14085e-111	35060258.5	1.52533e-102	0.15962	9.28050e-72
Freekick	20.72628	1.28629e-92	34452639.5	4.32455e-84	0.14157	1.76791e-56

Table 1: Results of Welch T-Test, Mann-Whitney U Test, and Kolmogorov-Smirnov Test for key attributes.

Attribute	**Left footed Mean**	**Left footed 95% CI**	**Right footed Mean**	**Right footed 95% CI**
Crossing	56.72	(56.23, 57.21)	47.56	(47.25, 47.88)
Dribbling	59.97	(59.50, 60.44)	53.85	(53.52, 54.18)
Curve	52.57	(52.04, 53.10)	45.44	(45.13, 45.75)
Freekick	47.56	(47.04, 48.09)	41.21	(40.92, 41.49)

Table 2: Means and 95% Confidence Intervals for left footed and right footed players.

As we can see clearly, in each of the attirbutes we tested, we have rejected the null hypothesis, meaning that in each attribute the left footed players have an advantage.

3.2 Comparing Attributes by Different Subgroups

Another analysis we did aimed to explore the statistical differences between left footed and right footed players by categorizing players into positional groups (forwards, midfielders, defenders, and goalkeepers) and height groups. The study performed a Mann-Whitney U test to compare skill levels between left and right footed players within these subgroups (min. 30 players per subgroup). This

was in order to try and somewhat neutralize the effect of other variables and make the tests as fair and unbiased as possible.

Figure 2: P-values (U-test) of our null hypotheses, categorized by player subgroups.

The heatmap of p-values, shown in figure 2, highlights the results of these comparisons, with red areas indicating non-significant differences ($p>0.05$) between the two foot preferences, and green areas showing significant differences. This approach shows us that the significant advantage we have so far found may not be as strong - in dribbling for example, almost half of the subgroups show no significant difference between left and right footed players, even though we have initially found this attribute to be much higher among left footed players.

3.3 Machine Learning Analysis: XGBoost and SHAP Values

To complement the statistical tests, we employed an XGBoost model to predict various technical attributes based on physical characteristics and foot preference. Contrary to our expectations, the model's feature importance analysis identified foot preference as having a very small, almost insignificant impact on the attributes we examined.

We used this method on several features, in particular dribbling, as can be seen on figure 4. Using the model's feature importance and SHAP values to interpret the its predictions, we observed that while being left footed had a slight positive contribution to dribbling scores, the overall effect was minimal. This finding contrasts with the results from our statistical tests, which suggested a more substantial impact of foot preference on performance.

This could be the result of an unknown confouding variable - while on surface it seems like foot preference has a big role in affecting some of the player's abilities, in reality it might be caused by a completely different variable or multiple variables, that are by chance correlated to foot preference.

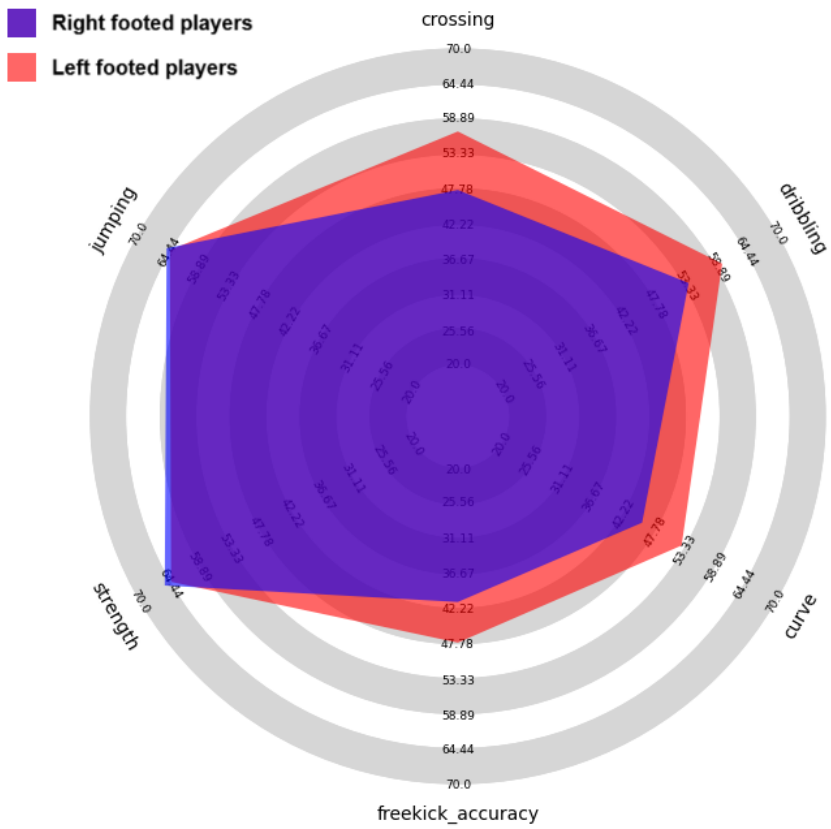

Figure 3: Key Attributes Radar scores for left-footed vs. right-footed players. In the plot you can see two of the only aspects in which right-footed are better on average: strength and jumping.

4 Discussion

The consistently higher ratings for left footed players in crossing and dribbling suggest that they may have an edge in one-on-one situations, particularly when playing on the wing. Crossing, in particular, is a skill that can be highly advantageous for left footed players in wide positions, as it allows them to deliver more accurate balls into the box from the left side of the pitch. This finding aligns with anecdotal observations from football coaches and fans.

Dribbling is another skill where left footed players seem to excel, possibly due to their ability to challenge defenders in ways that right footed players may not anticipate. The curve attribute also favors left footed players, which could give them an advantage in bending shots and passes around defenders or goalkeepers.

The findings of this study reveal some differences between left footed and right footed football players in key technical attributes such as crossing, dribbling, curve, and free kick accuracy. The initial results supported the hypothesis that left footed players have a unique advantage in certain aspects of the game, which may stem from their ability to surprise defenders due to the relative rarity of left-footedness, as well as natural physical or biomechanical factors. However, as we later saw, even though the initial results seemed clear cut, the clear advantage might not be the case at all.

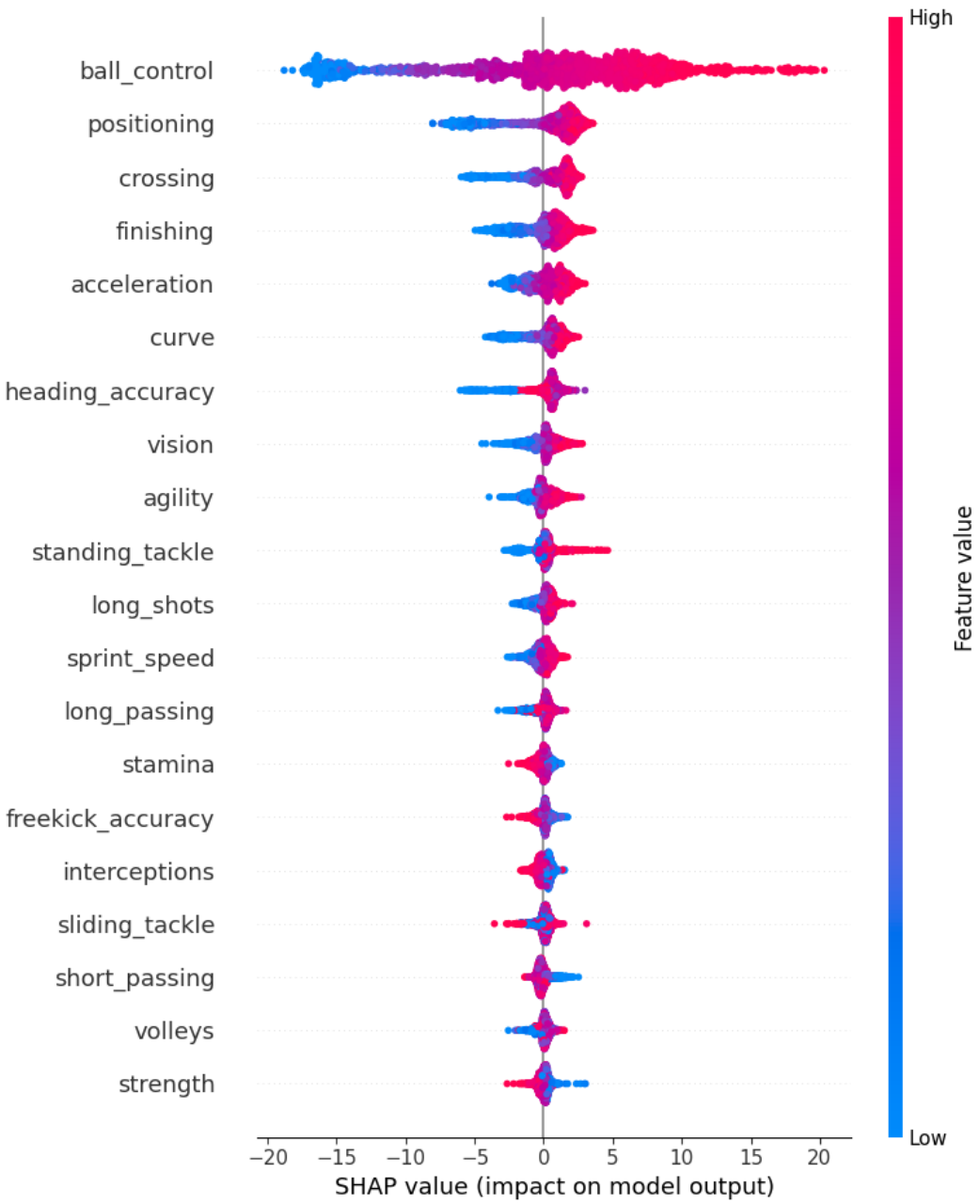

Figure 4: SHAP value analysis for dribbling. The preferred foot feature wasn't even in sight.

4.1 Limitations

While our results are compelling, there are limitations to our analysis. First, the dataset used is based on FIFA game ratings, which, although grounded in real-world player performance, may not fully capture all aspects of a player's abilities. These ratings are subjective to some extent, being determined by scouts and experts.

By using statistical tests on subgroups of players and using feature importance analysis on a machine learning model (parts 3.2 and 3.3) we tried to minimize bias and further explore the true effect of foot preference on different skills. We can almost certainly say that there remain other potential confounding variables that were not fully explored. Additionally, while the dataset provides comprehensive player ratings, it lacks more granular data such as playing time, tactical instructions, or match context, which could further clarify how foot preference impacts player effectiveness. Future studies could benefit from integrating these aspects to control for the nuances of football tactics and player utilization.

4.2 Future Research Directions

In our opinion, future research should focus on further isolating foot preference from the rest of a player's attributes, or find which other attributes it's highly correlated to, in order be clear cut about the foot preference's effect. Another possibility is to explore the impact of foot preference in other areas of football performance, such as defensive skills or goalkeeper performance. Studies could also expand to include data from other leagues or competitions, to determine whether the advantages observed for left-footed players in FIFA data hold true in broader contexts. Lastly, a long term study could examine how foot preference impacts player development over time.

4.3 Conclusion

In conclusion, while our study initially provides some evidence that left footed football players outperform their right footed counterparts in key technical attributes with high statistical significance, the advanced tools we used beg to differ. Separating players into more relevant groups and performing analysis using machine learning feature importance, showed less and even little evidence for a left footed advantage. Further research is needed to fully understand the underlying reasons for these differences and how they translate into real-world match performance.

References

[1] Sinan Bozkurt and Veysel Kucuk. Comparing of technical skills of young football players according to preferred foot. *International Journal of Human Movement and Sports Sciences*, 6(1):19–22, 2018.

[2] Yağmur Kocaoğlu and Yakup Girgin. Effects of foot preference on postural control in soccer players. *Turkish Journal of Kinesiology*, 9(1):67–76, 2023.