

Predicting Football Match Outcomes*

Sinan Ma

March 20, 2024

Introduction

In this essay, we explore the predictive power of shots on target and corner kicks on football match outcomes, focusing on the likelihood of a home team win and the number of goals scored by the home team. We employ logistic regression for the binary outcome of home team wins and compare Poisson and negative binomial regressions for count data of home team goals.

The dataset for the 2023/2024 Bundesliga 1 season, encompassing full-time and half-time results, match statistics, as well as match, total goals, and Asian Handicap odds, was sourced from Football-Data (Football-Data.co.uk 2024). Code is based on book *Telling Stories with Data* (Alexander 2023).

Data

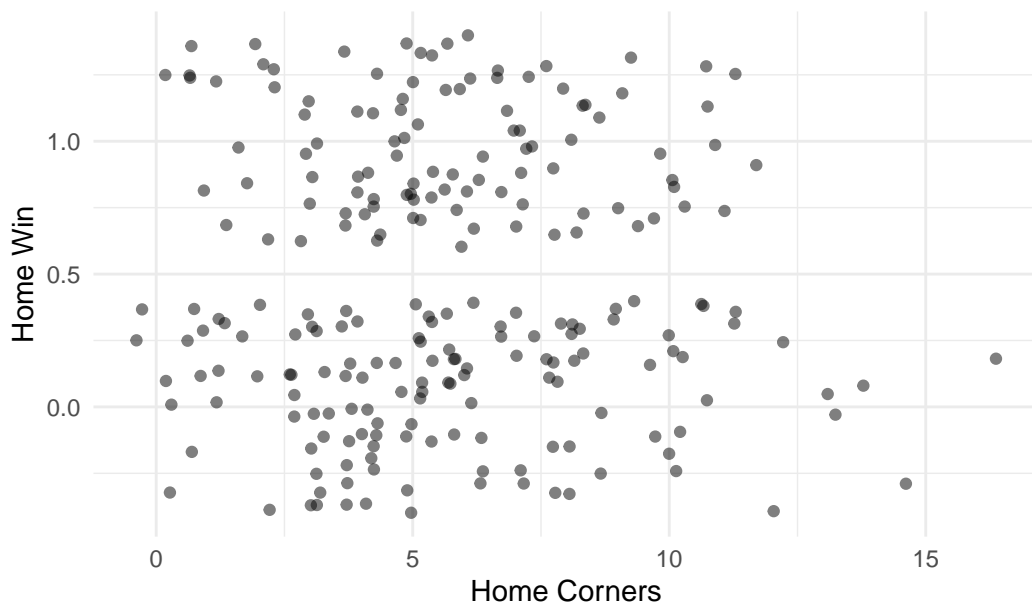
Our dataset comprises detailed match statistics from a football league, including shots on target (SOT) and corners (C) for each game. Initial exploratory data analysis revealed a positive correlation between the number of shots on target and home team victories, suggesting that teams with higher SOT are more likely to win. Similarly, corners showed a moderate positive correlation with home team wins, indicating their potential influence on the game's outcome.

*Code and data are available at: <https://github.com/Sinanma/Mini-essay-10.git>

Relationship between Home Shots on Target and Home Wins



Relationship between Home Corners and Home Wins



	Logistic	Poisson
(Intercept)	-1.492	-0.081
HST	0.327	0.161
HC	-0.101	-0.056
Num.Obs.	234	234
R2	0.141	
Log.Lik.	-143.817	-344.142
ELPD	-146.7	-346.5
ELPD s.e.	5.8	7.7
LOOIC	293.4	693.0
LOOIC s.e.	11.5	15.3
WAIC	293.4	693.0
RMSE	0.46	1.08

Logistic Regression for Home Team Wins

In examining the impact of match activities on home team victories, our logistic regression analysis revealed some insightful trends. Notably, each additional home shot on target (HST) is associated with a notable increase in the probability of a home team win, as indicated by a coefficient of 0.327. This positive relationship suggests that more aggressive offensive play significantly enhances the chances of victory. Contrary to expectations, corners (HC) are inversely related to the probability of winning, with a coefficient of -0.101. This could imply that while corners represent offensive opportunities, they do not necessarily translate to wins and may require more information to understand this dynamic.

Poisson Regression for Home Team Goals

The Poisson model we used to predict the count of home team goals, echoes the importance of shots on target. The coefficient of 0.161 for HST confirms a positive correlation with the number of goals, implying that direct shots on goal are an effective metric for success in scoring. The corner kicks' coefficient is slightly negative (-0.056), indicating a role of corners in goal-scoring that may not always align with increased goal counts.

Model Comparison

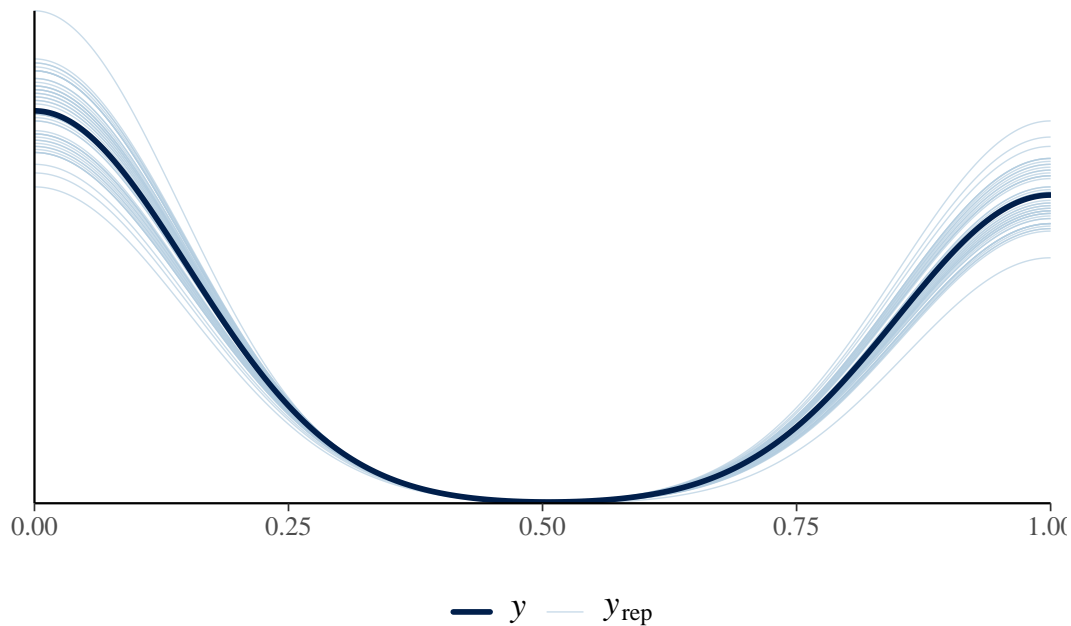
The comparison of the logistic and Poisson models using the provided data reveals contrasting performance. The logistic model's R2 value suggests a modest explanation of the variability

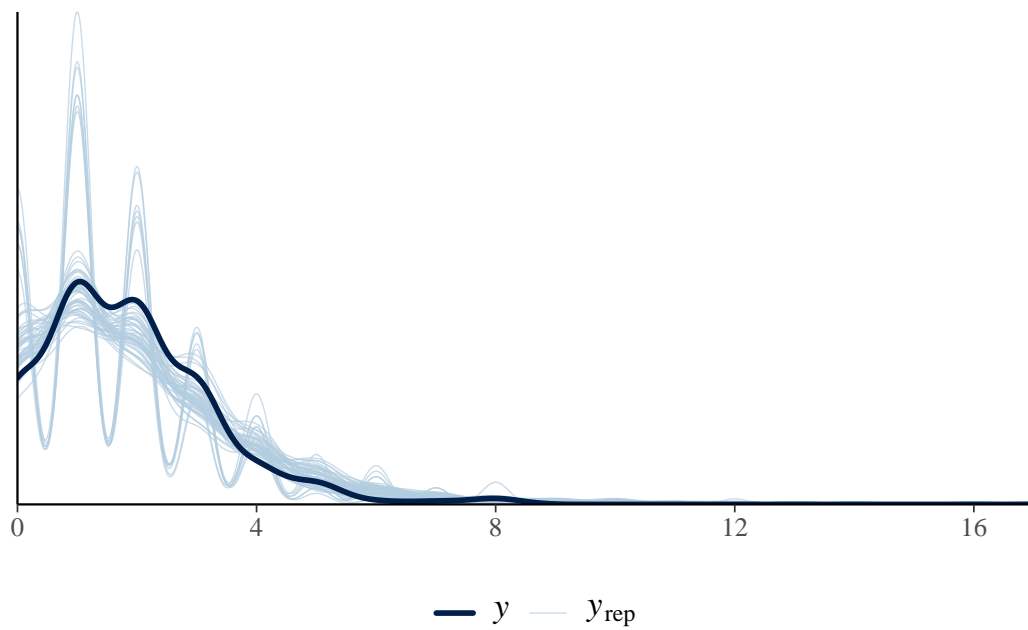
in home wins. However, the Poisson model's significantly higher RMSE indicates a less accurate fit for predicting the number of home goals. Moreover, the vastly higher LOOIC and WAIC values for the Poisson model point to poorer predictive quality compared to the logistic model.

Model diagnostics

The logistic regression diagnostic plot indicated a alignment between the observed binary outcomes of home team wins and the predicted probabilities, suggesting that the logistic model possesses adequate discriminative ability.

In contrast, the Poisson regression diagnostic plot revealed discrepancies between the observed and predicted counts of home goals, signaling potential overdispersion—an issue commonly rectified by employing a negative binomial regression.





References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Boca Raton: CRC Press. <https://tellingstorieswithdata.com/>.
- Football-Data.co.uk. 2024. "German Bundesliga 1 Data 2023/2024." Available online. <https://www.football-data.co.uk/germanym.php>.