Estimating the Probability that FC Barcelona Win a Game Using a Logistic Model

Introduction

In this paper, we seek to implement a logistic regression in order to estimate the probability that FC Barcelona, winner of the 2022-2023 LaLiga, win a game given that a game is Home or Away, the amount of possession they held, who the captain was between one of Sergio Busquets, Marc-André ter Stegen, Sergi Roberto, and Gerard Piqué, as well as what day of the week the game was played.

Data

The data that this paper utilizes contains information all 38 games played by FC Barcelona in the 2022-2023 LaLiga season, including the result of the game (win, draw, or loss), what day of the week the game was played, the percentage of possession that FC Barcelona held throughout the game (between 0 and 100), as well as which of the aforementioned captains was in charge of each particular game.

We mutated the data such that a win registers a value of "1", and a draw or loss register a value of "0" for use in the regression.

Model

The particular model that this paper will utilize is logistic. The primary reason why I have decided to make use of a logistic model in particular, is due to the binary nature of the outcome variable that we are interested in - whether FC Barcelona will win or lose. Though there is the possibility of a draw, as this paper is primarily concerned with whether or not FC Barcelona wins, we treat a draw as a loss in this regard.

Table 1: Coefficient values derived from the logistic regression

Table 2: Predicted likelihood of FC Barcelona winning a game compared with the actual result.

Game	Estimated Probability	Result	Day	Venue	Possession	Captain
1	0.8789373	0	Sat	Home	67	Sergio Busquets
2	0.8256687	1	Sun	Away	58	Marc-André ter Stegen
3	0.8609706	1	Sun	Home	67	Sergio Busquets
4	0.9064148	1	Sat	Away	54	Sergio Busquets
5	0.6622977	1	Sat	Away	70	Sergio Busquets
6	0.7164586	1	Sat	Home	76	Marc-André ter Stegen

The independent variables which we are interested in examining with regards to their effect on the probability that FC Barcelona wins or loses a game are as follows: Day (one of Saturday, Sunday, Monday, Tuesday, Wednesday, or Thursday), Venue (Home or Away), Possession (integer value between 0-100), and one of the aforementioned captains.

Thus, the model is as follows:

$$y_i|\pi_i \sim \text{Bern}(\pi_i)$$
$$\text{logit}(\pi_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i}$$

Here,

$$y_i = 1$$

denotes a win, and each of the independent variables from 1 to 4 represent the day, venue, possession, and the captain respectively.

Results and Discussion

After running our regression,