**Research Paper**

# Statistical Modeling for Sports Prediction - An application in the 2018 Brazilian Football Championship with the bivariate Poisson model

Magno Xavier de Araujo[1]
Anna Karla Barros da Trindade[2]
Francisco de Paula Santos de Araujo Junior[3]
*[1]FARCE, Mathematics Department, Camocim, Brazil*
*[2] Federal Institute of Piauí, Mathematics Department, Corrente, Brazil*
*[3] Federal University of Piauí, Education Department, Teresina, Brazil*
*Corresponding Author: pjhatata@hotmail.com*

***ABSTRACT:*** *The present work aims to approach the Holgate bivariate Poisson model in its application to soccer data, considering the importance that soccer has worldwide and as a very popular sport in Brazil. In a football match, two teams play for 90 minutes, plus extra time defined by the match referee according to the organizations' own criteria, with the objective of scoring as many goals as their opponent. It can be seen in some studies the statement that the final result of a soccer match is a bivariate random vector whose number of goals scored by each team can be represented by its parameters. In view of the above, from Holgate's bivariate Poisson model, the attack and defense capacity of each team will be estimated, based on the average of goals scored and conceded by the home team playing at home, and the visiting team playing away from home. Such a model allows the analysis of prediction of the probability of victory, draw and defeat of the teams in a certain future match, with defined parameters and a model that takes into account the goals scored and suffered by the teams until the previous round to the analysis of the model. With this, the focus is on the application of the model, to improve and perfect the ability to predict results, where the four initial rounds of the championship serve to define the initial parameters for the model. It is worth noting that the methodology presented will be applied with data from the 2018 Campeonato Brasileiro de Futebol da Série A, which is the main national championship in Brazil and involves the top twenty teams based on their rankings in the previous year's championship.*
***KEYWORDS:*** *Football, Forecast, Poisson Distribution, Simulation*

## I.    INTRODUCTION

Considering the importance that football has worldwide, statisticians from several countries began to build statistical models trying to predict the outcome of football matches. In a soccer match there is a complexity of factors that are decisive in the final result, such as: playing at home; lawn type; twisted; strategy; climatic and atmospheric conditions; arbitration; injury to a player during the match; the time when the match will be played, the situation of the teams in the championship, among many others.

In studies carried out by Pollard [10], it was verified that a team has an advantage when playing at home through the percentage of victories obtained by the home team in all games of a championship, considering a tournament where all teams play the same number of matches as home and away. The percentage can be verified through the ratio between the points conquered at home and the total points possible as home team, such percentage indicates the advantage of the teams playing at home. One of the factors that influence a team's advantage when playing at home is the crowd. According to the works of Clark and Norman [6] and Kuk [9], the importance of fans in a football match was concluded. Fans usually encourage the team in various ways, such as assembling a mosaic in the stadium, singing the team's anthem, shouting the players' names, pressuring the referee and booing the opponent.

---

Therefore, considering all the factors that influence a soccer match, it is almost impossible to predict the exact result. However, with some data collected from previous games of the teams, it is possible to determine the probability of each result to occur and with that, to have a good indicator for the most probable result of a future match.

According to Barbetta [3], "often, the conditions of the experiment do not allow to deduce the result, but only the chance (or probability) of possible outcomes". In this context, the present work intends to present a probabilistic model in order to assist in the prediction of soccer matches through the calculation of the probabilities of winning, drawing or losing the teams. in a given game. To this end, the average of goals scored and conceded by the home team playing at home and the visiting team playing away from home will be used. Such averages were applied in the Holgate bivariate Poisson class to predict the results of the 2018 Serie A Brazilian Football Championship games.

## II.   HISTORICAL CONTEXT

Considered one of the most difficult championships in the world, the "Brasileirão", as it is popularly known, is the main club tournament in Brazil. It began in 1971, when the Brazilian Sports Confederation held the first National Football Championship with 20 participating clubs, whose "Atlético Mineiro" team was crowned champion. Previously, several tournaments were organized, such as the Brazil Cup, the Roberto Gomes Pedrosa Tournament, the Rio-São Paulo Tournament and the National Club Tournament.

From 1979 onwards, with the emergence of the Brazilian Football Confederation, all states were included in the national tournament. One of the historic characteristics of the championship was the lack of standardization in the dispute system, which changed every year, as well as the number of participants. Among the different formats adopted are the elimination system, between 1959 and 1968, and mixed group systems, between 1967 and 2002. The form of dispute was standardized only in 2003, when the point system was adopted with all the teams facing each other in turn and return, with "Cruzeiro Esporte Clube" being the first champion in this format. On December 22, 2010, the CBF started to recognize the Brazil Cup and the Roberto Gomes Pedrosa Tournament as the Brazilian Championship.

The tournament is still organized by the CBF and the top six play in the Copa Libertadores, with the top four entering the group stage, while the fifth and sixth play in the preliminary stage. From seventh to twelfth place, they qualify for the Copa Sudamericana. The last four placed in the Brazilian championship from "Seria A" are relegated to Série B. In 2018, the Campeonato Brazilian was considered the third most difficult national football championship in the world by the IFFHS.

## III.   PROBABILISTIC MODEL

To calculate the likely outcome of a given match, take X as the number of goals scored by a team, and using the Poisson distribution given by the formula below:

$$P(X = x) = \frac{e^{-\delta_x}(\delta_x)^x}{x!}, \delta_x > 0, \qquad (1)$$

And making the necessary changes, we have:

$$P(A = a) = \frac{e^{-\delta_a}.(\delta_a)^a}{a!}, \delta_a > 0, \qquad (2)$$

$$P(B = b) = \frac{e^{-\delta_b}.(\delta_b)^b}{b!}, \delta_b > 0, \qquad (3)$$

Where, in a soccer match, A is the number of goals of the home team per game, $\delta_a$ is the average of goals of the home team per game, B is the number of goals of the away team per game, and $\delta_b$ e´ a me´ away team goals per game.

According to the theorem pointed out by Griffiths [8], "every infinitely visible bivariate Poisson distribution has a non-negative correlation". According to this theorem, for any two teams A and B, only one of the three situations occurs:

---

1. Team A is better than Team B - In this situation, the home team's victory is the most likely to occur, and the probability can be calculated by:

$$P(H) = \sum_{a>b} P(A = a, B = b). \qquad (4)$$

2. Team A and Team B are equivalent - In this situation, a tie is the most likely to occur, and the probability can be calculated by:

$$P(T) = \sum_{a=b} P(A = a, B = b). \qquad (5)$$

3. Team A is worse than Team B - In this situation, the away team's victory is the most likely to occur, and the probability can be calculated by:

$$P(V) = \sum_{a<b} P(A = a, B = b). \qquad (6)$$

Furthermore, it can be noted that:

$$P(H) + P(T) + P(V) = 1. \qquad (7)$$

**3.1. Calculation of Probabilities**

In this work, it will be assumed that the number of goals scored by the home team regardless of the number of goals scored by the visiting team in each match. Then, it is assumed that:

$$P(A = a, B = b) = P(A = a) \cdot P(B = b). \qquad (8)$$

Thus, substituting (8) in (4), (5) and (6), we have:
1. For the home team to win the game:

$$P(H) = \sum_{a>b} (P(A = a. B = b)). \qquad (9)$$

2. For the game to end in a tie:

$$P(T) = \sum_{a=b} (P(A = a. B = b)). \qquad (10)$$

3. For the visiting team to win the game:

$$P(H) = \sum_{a<b} (P(A = a. B = b)). \qquad (11)$$

**3.2. Parameters**
Clearly, one of the greatest difficulties adopted is the determination of the parameters, at the time, according to the average to be used.

Proposing an estimate for $\delta a$ as the average between scored by the team
home team playing at home and scoring goals for the visiting team playing away from home, in the same way, $\delta b$ is estimated as the average between goals scored by the visiting team playing away from home and teams playing for the home team playing at home, for the dimensioning of the parameters, we have:

$$\delta_a = \frac{(A_{scored} + B_{suffered})}{2} \tag{12}$$

Where $A_{scored}$ and the average of goals scored by the home team is $B_{suffered}$ and the average of goals conceded by the visiting team.

$$\delta_b = \frac{(B_{scored} + A_{suffered})}{2} \tag{13}$$

Where $B_{scored}$ is the average number of goals scored by the visiting team and $A_{suffered}$ is the average number of goals conceded by the home team.

**3.3. Goal Limit for Odds Calculation**
Analyzing the data from the 380 games of the 2017 Brazilian Championship, it is noticeable that there was a distribution of goals per match, as shown in the figure below.



| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 26 | 90 | 110 | 63 | 55 | 22 | 7 | 5 | 1 | 1 |

**Figure 1**. Goals per match of the Brazilian Championship 2017.

It can be observed that in 96, 31% of the data, the number of goals of the teams was at most five and in only 3, 69% of the data, the number of goals was superior to five. It can also be noted that the most repeated data was two goals per match, with 28.94%, followed by one goal per match, with 23.68%. Therefore, the number of five goals will be considered as the maximum number of goals per match, since this was presented in 96, 31% of the cases.

## IV.    DATA COLLECTION
As an example, consider a match valid for the penultimate round of the 2017 Serie A Brazilian Championship, played between "Palmeiras" and "Botafogo". The table below shows the average of goals scored and conceded that the two teams presented in the championship until the moment of the confrontation.
As mentioned before, the number of five goals will be considered the maximum number of goals per match. Therefore, for the calculation of probabilities, 0 to 5 goals occur, with the symbol F being the probability of 5 or more goals.

| Average goals scored | Home | Away |
|---|---|---|
| Palmeiras | 1,83 | 1,44 |
| Botafogo | 1,55 | 0,83 |

| Average goals conceded | Home | Away |
|---|---|---|
| Palmeiras | 0,88 | 1,44 |
| Botafogo | 1,27 | 0,83 |

**Figure 2.** Average goals to calculate parameters.

Therefore, to calculate the probability of each team scoring $n$ goals, take $\delta$ to be the average value between the goals scored when the home team plays at home and the goals conceded when the away team plays. Therefore, the average of goals of Palmeiras is:

$$\delta_a = \frac{1{,}83 + 0{,}83}{2} = \frac{2{,}66}{2} = 1{,}33$$

Substituting $\delta a = 1{,}33$ in (2) and checking the probability of Palmeiras playing as home team do:

1. For 0 (zero) goals, we have:

$$P(a = 0) = \frac{1{,}33^0 . e^{-1.33}}{0!} = 0{,}264$$

2. For 1 (one) goal, there is:

$$P(a = 1) = \frac{1{,}33^1 . e^{-1.33}}{1!} = 0{,}352$$

3. For 2 (two) goals, there are:

$$P(a = 2) = \frac{1{,}33^2 . e^{-1.33}}{2!} = 0{,}234$$

4. For 3 (three) goals, we have:

$$P(a = 3) = \frac{1{,}33^3 . e^{-1.33}}{3!} = 0{,}104$$

5. For 4 (four) goals, we have:

$$P(a = 4) = \frac{1{,}33^4 . e^{-1.33}}{4!} = 0{,}034$$

6. And for the probability of Palmeiras scoring more than four goals, it is given by:

$$P(a = F) = 1 - [P(0) + P(1) + P(2) + P(3) + P(4)]$$

$$P(a = F) = 1 - [0{,}264 + 0{,}352 + 0{,}234 + 0{,}104 + 0{,}034]$$

$$P(a = F) = 0{,}012$$

Later checking the same calculations for Botafogo, where $\delta$ is now considered the average value between the goals scored when the visiting team plays away from home and goals conceded by the home team when playing at home.

Therefore, Botafogo's average goals are:

$$\delta_b = \frac{0,83 + 0,88}{2} = \frac{1,71}{2} = 0,86$$

Substituting $\delta_b = 0,86$ in (3) and checking the probability of Botafogo playing as visitor do:
1. For 0 (zero) goals, we have:

$$P(b = 0) = \frac{0,86^0 . e^{-0.86}}{0!} = 0,425$$

2. For 1 (one) goal, there is:

$$P(b = 1) = \frac{0,86^1 . e^{-0,86}}{1!} = 0,364$$

3. For 2 (two) goals, there are:

$$P(b = 2) = \frac{0,86^2 . e^{-0,86}}{2!} = 0,155$$

4. For 3 (three) goals, we have:

$$P(b = 3) = \frac{0,86^3 . e^{-0,86}}{3!} = 0,044$$

5. For 4 (four) goals, we have:

$$P(b = 4) = \frac{0,86^4 . e^{-0,86}}{4!} = 0,009$$

6. And for the probability of botafogo scoring more than four goals, it is given by:

$$P(b = F) = 1 - [P(0) + P(1) + P(2) + P(3) + P(4)]$$

$$P(b = F) = 1 - [0425 + 0,364 + 0,155 + 0,044 + 0,009]$$

$$P(a = F) = 0,002$$

1. To obtain the probabilities of winning, drawing and losing the teams, the probability of each result will be calculated:

P(a > b) = P(a = 1, b = 0) + P(a = 2, b = 0) + P(a = 3, b = 0) + P(a = 4, b = 0)+ P(a = F, b = 0) + P(a = 2, b = 1) + P(a = 3, b = 1) + P(a = 4, b = 1)+ P(a = F, b = 1) + P(a = 3, b = 2) + P(a = 4, b = 2) + P(a = F, b = 2)+ P(a = 4, b = 3) + P(a = F, b = 3) + P(a = F, b = 4).

2. For the tie between the teams, consider, in order to calculate the probabilities, that if the two teams score more than four goals in the same game, the result is a tie:

P(a = b) = P(a = 0, b = 0) + P(a = 1, b = 1) + P(a = 2, b = 2)+P(a = 3, b = 3) + P(a = 4, b = 4) + P(a = F, b = F)

3. And for the victory of the visiting team, we have:

P(a < b) = P(a = 0, b = 1) + P(a = 0, b = 2) + P(a = 0, b = 3) + P(a = 0, b = 4)+ P(a = 0, b = F) + P(a = 1, b = 2) + P(a = 1, b = 3) + P(a = 1, b = 4)+ P(a = 1, b = F) + P(a = 2, b = 3) + P(a = 2, b = 4) + P(a = 2, b = F)+ P(a = 3, b = 4) + P(a = 3, b = F) + P(a = 4, b = F).

Where, taking as an example: $P(a = 1, b = 0)$ is the probability of the score of the match ending up 1x0, that is, this being the probability of $(P(a = 1) . P(b = 0))$ where $P(a = 1)$ refers to the probability of the home team scoring a goal and $P(b = 0)$ the probability of the visiting team scoring zero goals.

Therefore, for this match between "Palmeiras" and "Botafogo", we have:

Probability of victory of the home team:
P(H) = P(a > b) = 47.78%.

Probability of a tie:
P(T) = P(a = b) = 28.17%.

Probability of victory of the visiting team:
P(V) = P(a < b) = 24.05%.

The final result in the game was Palmeiras 2 x 0 Botafogo, that is, the home team won according to the trend that can be seen with greater probability in the calculations above.

Note that with this model, the probability of possible outcomes of a match can be found, and through these, analyze the probability of each team, taking into account the average of goals scored and conceded by the teams in previous games.

## V. EXECUTION OF THE ALGORITHM AND ITS RESPONSES

Checking the results obtained with the model presented above, and also considering the first four rounds of the 2018 Brazilian Championship, they were used to obtain the initial parameters, since, in the games referring to these rounds, the teams still did not have initial parameters for scoring goals. made and goals conceded, which means that the analyzes in these initial matches are not counted for hits or misses, but to generate the initial data. Where after the first four rounds all teams have two home games and two away games.

The data from the thirty-four championship rounds that were analyzed are constructed from tables 1 to 38, but for the purposes of data presentation, we suppressed some so as not to leave too extensive data, where the base and initial model and the data endings are highlighted. to as hold method. The software used and fed with the new data was Microsoft Excel, and as highlighted, it was fed with each new round, so the database of each of the teams was improved throughout the championship.

**Table 1.** Base Model.

| HOME TEAM | PVH | PT | PVV | VISITING TEAM | SCORE | HIT |
|-----------|-----|-----|-----|---------------|-------|-----|
| A1 | % | % | % | B1 | 0 - 0 | |
| A2 | % | % | % | B2 | 0 - 0 | |
| A3 | % | % | % | B3 | 0 - 0 | |
| A4 | % | % | % | B4 | 0 - 0 | |
| A5 | % | % | % | B5 | 0 - 0 | |
| A6 | % | % | % | B6 | 0 - 0 | |
| A7 | % | % | % | B7 | 0 - 0 | |
| A8 | % | % | % | B8 | 0 - 0 | |
| A9 | % | % | % | B9 | 0 - 0 | |
| A10 | % | % | % | B10 | 0 - 0 | |

Note that teams An and Bn are for completion only, I emphasize again that the first four tables are for obtaining the initial parameters.

The list of teams that make up the "series A" of the 2018 Brazilian championship are (Grêmio, Flamengo, Ceará, Sport, Fluminense, Bahia, Atletico MG, Chapecoense, Palmeiras, Paraná, Cruzeiro, Vitoriá, Santos, America MG, Corinthians, International, Vasco, Atletico PR, Botafogo and Sao Paulo). That is 20 (twenty) teams, so each round has 10(ten) games.

Where we highlight that PVH is the probability of victory of the home team, PT is the probability of a draw and PVV is the probability of victory of the visiting team, and all values are tabulated in percentage.

### 5.1. Initial Tables for Obtaining the Initial Parameters
The following are the four starting tables of the championship and their respective results

**Table 2.** First round.

| HOME TEAM | PVH | PT | PVV | VISITING TEAM | SCORE | HIT |
|---|---|---|---|---|---|---|
| Cruzeiro | % | % | % | Grêmio | 0 - 1 | |
| Vitória | % | % | % | Flamengo | 2 - 2 | |
| Santos | % | % | % | Ceará | 2 - 0 | |
| América MG | % | % | % | Sport | 3 - 0 | |
| Corinthians | % | % | % | Fluminense | 2 - 1 | |
| Internacional | % | % | % | Bahia | 2 - 0 | |
| Vasco | % | % | % | Atlético MG | 2 - 1 | |
| Athletico PR | % | % | % | Chapecoense | 5 - 1 | |
| Botafogo | % | % | % | Palmeiras | 1 - 1 | |
| São Paulo | % | % | % | Paraná´ | 1 - 0 | |

**Table 3.** Second round.

| HOME TEAM | PVH | PT | PVV | VISITING TEAM | SCORE | HIT |
|---|---|---|---|---|---|---|
| Bahia | % | % | % | Santos | 1 - 0 | |
| Flamengo | % | % | % | América MG | 2 - 0 | |
| Atlético MG | % | % | % | Vito´ria | 2 - 1 | |
| Ceará | % | % | % | São Paulo | 0 - 0 | |
| Chapecoense | % | % | % | Vasco | 1 - 1 | |
| Fluminense | % | % | % | Cruzeiro | 1 - 0 | |
| Palmeiras | % | % | % | Internacional | 1 - 0 | |
| Paraná | % | % | % | Corinthians | 0 - 4 | |
| Grêmio | % | % | % | Athletico PR | 0 - 0 | |
| Sport | % | % | % | Botafogo | 1 - 1 | |

**Table 4.** Third round.

| HOME TEAM | PVH | PT | PVV | VISITING TEAM | SCORE | HIT |
|---|---|---|---|---|---|---|
| Botafogo | % | % | % | Grêmio | 2 - 1 | |
| Atlético MG | % | % | % | Corinthians | 1 - 0 | |
| Bahia | % | % | % | Athletico PR | 0 - 0 | |
| Ceara´ | % | % | % | Flamengo | 0 - 3 | |
| Fluminense | % | % | % | São Paulo | 1 - 1 | |
| Paraná | % | % | % | Sport | 1 - 2 | |
| Internacional | % | % | % | Cruzeiro | 0 - 0 | |
| Palmeiras | % | % | % | Chapecoense | 0 - 0 | |
| Ame´rica MG | % | % | % | Vitória | 2 - 1 | |
| Santos | % | % | % | Vasco | 1 - 1 | |

**Table 5.** Fourth round.

| HOME TEAM | PVH | PT | PVV | VISITING TEAM | SCORE | HIT |
|---|---|---|---|---|---|---|
| Vasco | % | % | % | América MG | 4 - 1 | |
| São Paulo | % | % | % | Atlético MG | 2 - 2 | |
| Sport | % | % | % | Bahia | 2 - 0 | |
| Corinthians | % | % | % | Ceará | 1 - 1 | |
| Athletico PR | % | % | % | Palmeiras | 1 - 3 | |
| Cruzeiro | % | % | % | Botafogo | 1 - 0 | |
| Flamengo | % | % | % | Internacional | 2 - 0 | |
| Grêmio | % | % | % | Santos | 5 - 1 | |
| Vitória | % | % | % | Fluminense | 1 - 2 | |
| Chapecoense | % | % | % | Paraná | 1 - 1 | |

It is now possible to make predictions for the future rounds of the matchmaking and verify if the statistical model presents us with a higher probability of success compared to the verification with the real results, "yes" is the answer if the highest probability was verified compared to reality and "no" otherwise

**5.2. Tables with hit results**

**Table 6.** Fifth round.

| HOME TEAM | PVH | PT | PVV | VISITING TEAM | SCORE | HIT |
|---|---|---|---|---|---|---|
| Grêmio | 79,72% | 16,51% | 3,77% | Internacional | 0 - 0 | No |
| Corinthians | 31,64% | 25,32% | 43,04% | Palmeiras | 1 - 0 | No |
| Vasco | 70,03% | 16,84% | 13,13% | Vitória | 2 - 3 | No |
| Chapecoense | 21,48% | 23,54% | 54,98% | Flamengo | 3 - 2 | No |
| Atletico PR | 54,27% | 19,08% | 26,65% | Atlético MG | 1 - 2 | No |
| Bahia | 32,44% | 53,33% | 14,23% | São Paulo | 2 - 2 | yes |
| Santos | 69,59% | 24,53% | 5,88% | Paraná | 3 - 1 | yes |
| Cruzeiro | 48,17% | 29,52% | 22,31% | Sport | 2 - 0 | yes |
| Botafogo | 43,04% | 25,32% | 31,64% | Fluminense | 2 - 1 | yes |
| Ceará | 48,79% | 25,99% | 25,22% | América MG | 2 - 2 | No |

Note that in this first round that we can have predictions, only 40% of the games were correct. However, as we can see in the next round.

**Table 7.** Sixth round.

| HOME TEAM | PVH | PT | PVV | VISITING TEAM | SCORE | HIT |
|---|---|---|---|---|---|---|
| Atético MG | 54,57% | 36,27% | 9,16% | Cruzeiro | 1 - 0 | yes |
| Flamengo | 62,06% | 25,75% | 12,19% | Vasco | 1 - 1 | No |
| Palmeiras | 71,35% | 28,65% | 0,00% | Bahia | 3 - 0 | yes |
| Viória | 43,04% | 25,32% | 31,64% | Ceará | 2 - 1 | yes |
| América MG | 68,02% | 21,99% | 9,99% | Botafogo | 1 - 0 | yes |
| Paraná | 13,07% | 20,24% | 66,69% | Grêmio | 0 - 0 | No |
| São Paulo | 71,42% | 17,67% | 10,91% | Santos | 1 - 0 | yes |
| Sport | 29,56% | 28,45% | 41,99% | Corinthians | 1 - 1 | No |
| Fluminense | 32,44% | 53,33% | 14,23% | Athletico PR | 2 - 0 | No |
| Internacional | 75,16% | 20,13% | 4,71% | Chapecoense | 3 - 0 | yes |

Where in this round the model hit has already reached 80% of the games. We draw attention to the eighth round where the model hit 80% of the games as it can be seen in the following table.

**Table 8.** Eighth round.

| HOME TEAM | PVH | PT | PVV | VISITING TEAM | SCORE | HIT |
|---|---|---|---|---|---|---|
| São Paulo | 52,48% | 31,65% | 15,87% | Botafogo | 3 - 2 | yes |
| Sport | 44,82% | 25,58% | 29,60% | Atlético MG | 3 - 2 | yes |
| Vasco | 60,12% | 20,17% | 19,71% | Paraná | 1 - 0 | yes |
| Chapecoense | 53,07% | 24,32% | 22,61% | Ceará | 2 - 0 | yes |
| Vitória | 48,43% | 28,16% | 23,41% | Internacional | 2 - 3 | No |
| Cruzeiro | 38,42% | 32,39% | 29,19% | Palmeiras | 1 - 0 | yes |
| Grêmio | 57,32% | 24,19% | 18,49% | Fluminense | 0 - 0 | No |
| Flamengo | 82,01% | 15,59% | 2,40% | Bahia | 2 - 0 | yes |
| Corinthians | 64,57% | 20,62% | 14,81% | América MG | 1 - 0 | yes |
| Athletico PR | 63,35% | 18,90% | 17,75% | Santos | 2 - 0 | yes |

We emphasize that these adjustments had variations and were not always this way, so we will present the last table made at 38° where we suppress the others because presenting all these would occupy a large percentage of the work but if any reader wants them, just enter in this way. contact any of the authors who will make these available.

**Table 9.** 38th round.

| HOME TEAM | PVH | PT | PVV | VISITING TEAM | SCORE | HIT |
|---|---|---|---|---|---|---|
| Atlético MG | 57,15% | 23,96% | 18,89% | Botafogo | 1 - 0 | Sim |
| Flamengo | 63,24% | 26,84% | 9,92% | Atletico PR | 1 - 2 | No |
| Bahia | 53,58% | 29,01% | 17,41% | Cruzeiro | 0 - 0 | No |
| Ceará | 44,84% | 30,56% | 24,60% | Vasco | 0 - 0 | No |
| Chapecoense | 39,72% | 28,81% | 31,47% | São Paulo | 1 - 0 | Sim |
| Fluminense | 54,37% | 30,03% | 15,60% | América MG | 1 - 0 | Sim |
| Grêmio | 55,74% | 25,22% | 19,04% | Corinthians | 1 - 0 | Sim |
| Palmeiras | 71,11% | 18,69% | 10,20% | Vitória | 3 - 2 | Sim |
| Paraná | 27,26% | 30,96% | 41,78% | Internacional | 1 - 1 | No |

| Sport | 40,62% | 30,14% | 29,24% | Santos | 2 - 1 | Sim |

As can be seen in this last round, the hits still remain above 50%, as not all rounds were presented, let's make a general analysis of the right ones later.

The results obtained in the present work show that, as proposed by Pollard [10], there is an advantage for the home team when playing at home, since, of the 380 games played, 202 ended with a victory for the home team, 110 ended tied and 68 with the victory of the visiting team.

The percentage of the results cited above can be seen in the subsequent image.
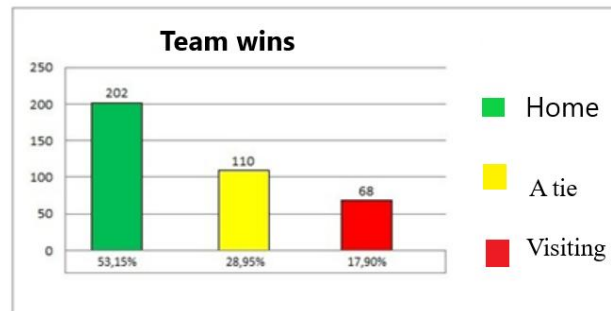


**Figure 3**. Final match results.

The model adopted for calculating the probabilities proved to be effective, since, of the 340 games analyzed, there were 184 hits and 156 errors in the predicted final results.

Now, checking the percentage of the final results correct using the model to calculate the probabilities, we have:
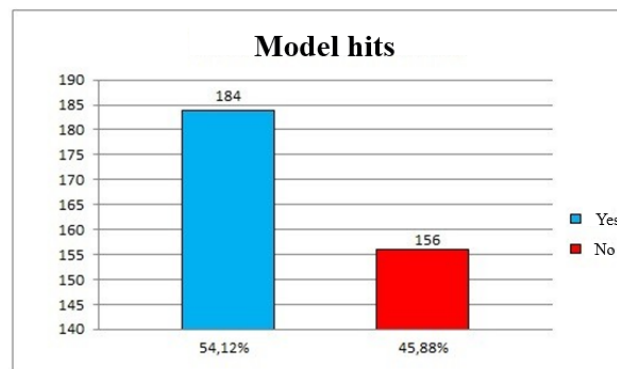


**Figure 3.** Model hits index.

Taking into account the complexity of factors that are decisive in a match, the reference parameter adopted for the calculation of probabilities proved to be satisfactory, as it indicated most of the correct results.

## VI.    FINAL CONSIDERATIONS

In the present work, an application of a simple method was carried out to calculate the probabilities of the possible outcomes of a soccer match. The method adopted was applied in the 2018 Brazilian Championship games, where it obtained a rate of 54.12% of correct results. As mentioned earlier, the model used does not take into account several factors that influence the final result of a soccer match. Such factors cause some forecasts to be inaccurate. But all this contributes to making this sport so fascinating and engaging for football lovers. Even so, the hit rate achieved demonstrates a good predictive quality.

The estimation of parameters was one of the difficulties encountered in this work. The decision taken was to adopt a method to estimate a parameter through the average of the goals scored and conceded by the teams playing as home and away. Although the results have been very satisfactory, it leaves a question, since it can be considered that there are more efficient methods if the existence of dependence on the number of goals of each team is considered.

The calculation of the prediction of the final result through previous results improves the analysis of the games, being able to choose the information that is most suitable for the choice of the prediction for the

match. The results of the 2018 Brazilian Championship games can be found on the Brazilian Football Confederation (CBF) page: https://www.cbf.com.br.

Regardless of the results, we believe that this research will contribute to the dissemination of the Poisson Method, and the presentation of the applicability of the use of discrete mathematics and the use of numerical methods in order to facilitate the understanding and comprehension of the use of the method, the discussion referring to the use of statistics in the solutions of daily problems faced and the possibility of future work in the area.

We emphasize that texts such as those of Araujo [1], Arruda [2] and Barreira [4], All serve as a theoretical basis for the construction of this research in order to provide an epistemological basis for the same. The texts by CEOLATO JUNIOR [5], SILVA [11], Silva [12] and Suzuki [13] are important in consultations and statistical analysis, as some of these are dissections that we recommend reading and helped in the same way in the construction of this research.

For future works, a desired approximation that can be made with variables not analyzed in the model in question can be as "Markov Chains" as can be seen in Amorim, Trindade and Araujo Junior [7].

However, we emphasize that such an approximation may not be possible and is not in the field of speculation, and that the correct results with the bivariate Poisson present satisfactory results as highlighted and presented above.

## REFERENCES

[1]. ARAUJO, Clodoaldo Tsuyoshi Pereira de; TAVARES, Leandro; ALVARES, Luis Gus- tavo; LOUZADA NETO, Francisco; SUZUKI, Adriano Kamimura. MODELAGEM ES- TATI´STICA PARA A PREVISA˜ O DE JOGOS DE FUTEBOL: UMA APLICAC¸ A˜ O NO CAMPEONATO BRASILEIRO DE FUTEBOL 2014. Revista da Estatística UFOP, Vol IV, 2015, ISSN 2237-8111.

[2]. ARRUDA, Marcelo Leme de. Poisson, Bayes, Futebol e DeFinetti. Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística da USP, São Paulo, 2000.

[3]. BARBETTA, Pedro Alberto. Estat´ıstica para cursos de engenharia e informa´tica. Atlas, São Paulo, 2004.

[4]. BARREIRA, Júlia. Vantagem de jogar em casa no futebol feminino: uma análise dos três principais campeonatos no Brasil. Revista Brasileira de Ciência e Movimento. 2018;26(3):83-87.

[5]. CEOLATO JUNIOR, Carlos Alberto. CHUTEGOL - CHance de Um Time Em GOLs: Um sistema para progno´sticos de jogos de futebol. Dissertação (Bacharelado em Sistemas de Informação) - Universidade Federal de Santa Catarina, 2007.

[6]. CLARKE, S.; NORMAN, J.M. Home ground advantage of individual clubs in English soccer. The Statistician, London, v.44, n.4, p.509-21, 1995.

[7]. AMORIM, A. M. do N. TRINDADE, A. K. B AND ARAUJO JUNIOR, F. P. S. Cadeias de Markov: Uma introdução apresentável ao ensino médio com aplicação ao soneto "Amor é fogo que arde sem se ver". Revista Eletrônica da Sociedade Brasileira de Matemática PMO. v.8, n.2, 2020 :236-248.

[8]. GRIFFITHS, R. C., MILNE, R. K & WOOD, R. (1979), Aspects of Correlation in Bivariate Poisson Distributions and Processes. Australian Journal of Statistics 21.

[9]. KUK, A.Y.C. Modelling paired comparison data with large numbers of draws and large variability of draw percentages among players. The Statistician, 44, 523 – 228, (1995).

[10]. POLLARD, R. Home advantage in soccer: a retrospective analysis. Journal of Sports Sciences, London, v. 4, n. 3, p. 237-48, 1986

[11]. SILVA, Cristiano Diniz da; MEDEIROS, N´ısio Cunha; SILVA, Ana Cristina Diniz da. Vantagem em casa no campeonato brasileiro de futebol: efeito do local do jogo e da qualidade dos times. Revista Brasileira de Cineantropometria e Desempenho Humano. 2010; 12 (2): 148-154.

[12]. SILVA, Wesley Bertoli da. DISTRIBUIÇÃO DE POISSON BIVARIADA APLICADA À PREVISÃO DE RESULTADOS ESPORTIVOS. Dissertação (Mestrado em Estatística) - Universidade Federal de Sa˜o Carlos - DEs/UFSCar, São Carlos, Abril de 2014.

[13]. SUZUKI, Adriano Kamimura. Modelagem Estatística para a Determinação de Resultados de Dados Esportivos. Dissertação (Mestrado em Estatística) - Universidade Federal de São Carlos - DEs/UFSCar, São Carlos, junho de 2007.