

# MODELAGEM ESTATÍSTICA PARA PREVISÃO ESPORTIVA: UMA APLICAÇÃO NO FUTEBOL

**Adriano Kamimura Suzuki**

Universidade de São Paulo

[suzuki@icmc.usp.br](mailto:suzuki@icmc.usp.br)

**Leandro Tavares**

Universidade de São Paulo

[leandrotz@usp.br](mailto:leandrotz@usp.br)

## RESUMO

Nos esportes, existe o interesse em obter a probabilidade de uma equipe vencer uma competição. Especificamente no futebol, encontramos diversos trabalhos publicados na literatura considerando previsão de futebol aplicados a ligas nacionais de futebol. Neste artigo nós aplicamos uma metodologia baseada no método Soma e Diferença (SD 0) proposto por [1], o qual assume que o número de gols marcados por um time em uma partida segue uma distribuição univariada de Poisson e considera modelos lineares que expressam a soma e a diferença de gols marcados em termos de quatro covariáveis: a média de gols em uma partida, a vantagem do time mandante, o poder ofensivo da equipe e o poder defensivo do adversário. O principal objetivo deste trabalho é calcular as probabilidades de interesses, tais como qual time será o campeão, quais serão os rebaixados, qual será o melhor time mandante, qual time será o melhor visitante (a equipe que marcar o maior número de pontos jogando fora de casa) etc. A metodologia apresentada foi aplicada no Campeonato Brasileiro 2013 série A de futebol e implementada no software R.

## ABSTRACT

In sports, there is interest in obtaining the probability of a team winning the competition. Specifically in football, several papers are found in literature considering football prediction applied to football national leagues. In this paper, we apply a methodology based on the Sum and Difference (SD 0) method proposed by [1], which assumes that the number of goals scored by a team in a match follows a univariate Poisson distribution and consider linear models that express the sum and the difference of goals scored in terms of four covariates: the goal average in a match, the home team advantage, the team's offensive power and the opponent team's defensive power. The main objective of this paper is to calculate the interest probabilities, such as which team will be the champion, which ones will be relegated, which team will be the best home team, which team will be the best away team (the team that scores the most points playing outside their hometown) etc. The presented methodology was applied to the 2013 Brazilian serie A football Championship and implemented in the R program.

**Palavras-chave:** Futebol; Previsão; Distribuição de Poisson; Simulação.

## 1 INTRODUÇÃO

Considerado uma paixão nacional, o futebol é um esporte mundialmente conhecido. Chegou ao Brasil por volta de 1894, trazido por Charles Miller. Com regras claras e objetivas começou a ser praticado apenas pela elite. O primeiro time a se formar no Brasil foi o SÃO PAULO ATHLETIC, fundado em 13 de maio de 1888. A partir de 1950, os brasileiros revelaram seus primeiros grandes craques, entre os quais destacamos Pelé e Garrincha. O Campeonato Brasileiro é o principal torneio entre clubes de futebol do Brasil. Iniciou-se em 1971, cujo o time campeão o Atlético Mineiro, treinado por Telê Santana e com Dário, o Dada Maravilha, no ataque. Antes da sua criação, foram organizados vários torneios como a Taça Brasil, o Torneio Roberto Gomes Pedrosa, o Torneio Rio-São Paulo e o Torneio Nacional de Clubes. Faltava, porém, integrar todo país em uma mesma competição. E foi com a criação do Brasileirão, como é popularmente conhecido pelos torcedores, que esse objetivo foi cumprido. Após seu início, teve edições com inúmeros nomes, tais como Taça de Prata, Campeonato Nacional de Clubes, Taça de Ouro, Copa Brasil, Copa União e a partir de 1989, como Campeonato Brasileiro de Futebol e em 2000, como Copa João Havelange. Um fato curioso do Brasileirão aconteceu em 1977, onde tanto o vice-campeão (Atlético Mineiro) quanto o 5º colocado (Botafogo) terminaram o campeonato sem derrotas, mas o campeão naquela temporada foi o São Paulo. A partir de 2003, foi adotado o sistema de pontos corridos, sendo o Cruzeiro Esporte Clube o primeiro campeão neste formato de disputa. O torneio é organizado pela Confederação Brasileira de Futebol (CBF) e dá acesso ao seu campeão, vice, terceiro e quarto colocados à Taça Libertadores da América e os quatro últimos colocados são rebaixados para chamada série B (Segunda divisão) do Campeonato Brasileiro.

Quando se leva em consideração apenas o período a partir de 1971, período em que o Campeonato Brasileiro surgiu na forma do Campeonato Nacional de Clubes, o São Paulo é o clube que mais vezes venceu a competição, seis vezes, levando em consideração que o título obtido pelo Flamengo em 1987 é oficialmente reconhecido pela CBF não como título brasileiro, mas como um dos módulos do campeonato daquele ano. Se forem considerados a Taça Brasil (primeira competição nacional entre clubes, disputada em sistema de copa entre 1959 e 1968) e o Robertão (competição nacional de futebol no Brasil disputada de 1967 a 1970) recentemente oficializados pela CBF como equivalentes ao Campeonato Brasileiro, o Santos e o Palmeiras são os maiores campeões, com oito títulos cada. Desde 2003, com a mudança no sistema de disputa, não existe mais o conceito de “jogo final” do campeonato. Uma das características do Campeonato Brasileiro foi a falta de uma padronização no sistema de disputa, que mudava a cada ano, assim como as regras e o número de participantes. Após ter sido aprovado no Congresso Nacional o “Código do Torcedor”, a CBF fez um planejamento que visava organizar o confuso calendário do futebol nacional. Reduziu-se o tempo disponível para as competições estaduais e adotou-se o sistema de turno e retorno como forma de disputa. Como esse sistema exige muito tempo do calendário, também foi reduzido o número de competidores em 2004 que eram de 24 times, para 22 times em 2005 e, para 20 times em 2006, tanto na Série A (Primeira Divisão) como na Série B (Segunda Divisão). Até 2011, apenas três clubes disputaram todas as edições da divisão principal: Cruzeiro, Internacional e Flamengo. Santos e São Paulo também nunca foram rebaixados, mas em 1979 se recusaram a participar do Brasileiro ao terem seus pedidos de entrarem apenas na fase final recusado depois de alegarem que a competição daquele ano tornava o calendário demasiado apertado. Em 2006, o Campeonato Brasileiro foi considerado como o 5º melhor campeonato nacional de futebol do mundo pela IFFHS.

Diante disso, o futebol sempre foi alvo de especulações sobre quem seria o campeão, o vice-campeão, quais os clubes seriam rebaixados, qual o time que marcaria mais gols, entre muitas outras perguntas relacionadas. Surgiram então os primeiros modelos estatísticos para tentar prever a resposta para essas perguntas que muitos torcedores e admiradores do esporte faziam.

[8] verificou que times levariam vantagem em jogar em casa por meio da porcentagem

de vitórias obtidas em jogos realizados com seu mando de campo e apoio da torcida. Em seu trabalho verificou que há vantagem em jogar em casa em qualquer que for a divisão do campeonato. Mas, as competições com menores expressões essa vantagem é minimizada. Alguns artigos encontrados na literatura referem-se a obter previsões para a copa do mundo, tais como [4], [11], [9], entre outros. [4] apresentam um modelo de regressão de Poisson log-linear que leva o ranking da FIFA como covariável. Apresentam alguns resultados sobre o poder preditivo do modelo e também os resultados da simulação para estimar probabilidades de ganhar a Copa do Mundo de 1998. [11] utiliza uma aproximação de processo de contagem, modelando o desenvolvimento de um jogo marcando dois processos aleatórios pontuais interagindo dependentes do tempo. A interação entre as equipes são modelados por meio de um modelo de regressão multiplicativa semi-paramétrica de intensidade. Como aplicação, realizou uma análise do desempenho das oito equipes que alcançaram as quartas-de-finais da Copa do Mundo de 2006. Já [9] propuseram uma metodologia bayesiana para prever resultados das partidas, utilizando opiniões de especialistas e o ranking da FIFA como informações *a priori*. O método é aplicado para calcular as probabilidades de vitória, empate e de perda para cada jogo como também para estimar as probabilidades de classificação na fase de grupos e as chances de cada equipe consagrar-se campeã da Copo da Mundo de 2006.

Aplicados às ligas e aos campeonatos nacionais, podem ser encontrados vários trabalhos direcionados para a predição de resultados de jogos de futebol. Por exemplo, [6] utiliza a distribuição de Poisson para modelar os números de gols marcados pela Inglaterra, Irlanda, Escócia e País de Gales, no Campeonato Internacional Britânico 1883 à 1980. [7] também considera a distribuição de Poisson em que as médias dependente de um efeito de jogar em casa, efeitos ofensivos e defensivos de cada equipe. [5] aplica a distribuição do Skellam para modelar os gols feitos pelas equipes jogando em casa e fora de casa. A análise bayesiana para predizer os resultados de correspondência para Liga Inglesa (2006-2007) é realizada por meio de uma função ligação log-linear e distribuições *a priori* não informativas para os parâmetros do modelo. Levando em consideração uma outra abordagem, [2] propôs para modelar diretamente as probabilidades de vitória, empate e derrota com aplicação de um modelo de regressão trinomial ao Campeonato Brasileiro de 2006 série A. Por meio de uma simulação, estimou para cada equipe o total de pontos, e assim obter a probabilidade de vencer o campeonato e a probabilidade de acabar a temporada entre os quatro primeiros colocados.

Neste trabalho, vamos aplicar uma metodologia para previsão de resultados do Campeonato Brasileiro de 2013, tais como a probabilidade de vitória, empate ou derrota de cada equipe, a probabilidade de um time ser campeão, classificação para a Copa Libertadores da América e ser rebaixado.

## 2 MODELO

Este modelo é baseado no método denominado SD 0 proposto por [1]. Para uma dada partida de futebol, considere  $X$  e  $Y$  o número de gols marcados pelos times mandante e visitante, respectivamente. Vamos assumir que  $X$  e  $Y$  são variáveis aleatórias independentes que seguem distribuição Poisson com médias  $\lambda_X$  e  $\lambda_Y$ , respectivamente.

Então, pela propriedade linear da esperança, temos que:

$$\lambda_X = E[X] = \frac{E[X + Y] + E[X - Y]}{2} \quad (1)$$

e

$$\lambda_Y = E[Y] = \frac{E[X + Y] - E[X - Y]}{2}, \quad (2)$$

em que  $\lambda_X$  e  $\lambda_Y$  podem ser expressadas como funções das médias das variáveis  $X+Y$  e  $X-Y$ . Baseado nas identidades (1) e (2), consideramos o vetor aleatório  $(U, V)$  definido por  $U=X+Y$  e  $V=X-Y$  como variável resposta ao invés do vetor  $(X, Y)$  de número de gols marcados por

cada time. Portanto, a estimação dos parâmetros  $\lambda_X$  e  $\lambda_Y$  podem ser obtidas baseada na estimação para as médias  $\mu=E[U]$  e  $\nu=E[V]$  usando  $\hat{\lambda}_X=(\hat{\mu}+\hat{\nu})/2$  e  $\hat{\lambda}_Y=(\hat{\mu}-\hat{\nu})/2$ . Para cada uma das futuras partidas da competição,  $\hat{\mu}$  e  $\hat{\nu}$  serão obtidas por meio de dois modelos lineares ajustados definidos a seguir.

Suponha que os times são categorizados como  $1, 2, \dots, T$  e que o conjunto de dados disponível consiste das  $n$  primeiras partidas da competição. Para a  $i$ -ésima partida entre o time mandante  $h[i]$  contra o time visitante  $a[i]$ , considere que  $U_i$  e  $V_i$  são a soma e a diferença de gols entre o time mandante e o visitante, respectivamente, em que  $h[i]$  e  $a[i] \in (1, 2, \dots, T)$ , para  $i = 1, 2, \dots, n$ . Os dois modelos lineares são definidos por:

$$U_i = \mu_i + \varepsilon_i \tag{3}$$

e

$$V_i = \nu_i + \varepsilon'_i, \tag{4}$$

em que  $\varepsilon_1, \dots, \varepsilon_n$  e  $\varepsilon'_1, \dots, \varepsilon'_n$  são erros independentes com média 0,  $i = 1, 2, \dots, n$ . As médias  $\mu_i$  e  $\nu_i$  de  $U_i$  e  $V_i$  são dadas pelas funções de ligação:

$$\mu_i = \alpha I_{casa[i]} + \beta_{h[i]} + \psi_{a[i]} \tag{5}$$

e

$$\nu_i = \alpha' I_{casa[i]} + \beta'_{h[i]} + \psi'_{a[i]} \tag{6}$$

em que  $\alpha$  e  $\alpha'$  se referem ao efeito mando de campo,  $\beta_{h[i]}$  e  $\beta'_{h[i]}$  o poder ofensivo do time  $h[i]$ ,  $\psi_{a[i]}$  e  $\psi'_{a[i]}$  o poder defensivo do time  $a[i]$ ,  $I_{casa[i]}$  é uma variável indicadora que assume 1 caso um dos times tenha o mando de campo e 0 caso contrário,  $i = 1, 2, \dots, n$ .

Os modelos definidos por (3) e (4) são ajustados separadamente via estimação por mínimos quadrados assumindo pesos iguais usando a matriz inversa generalizada de Moore-Penrose. Com os modelos ajustados obtemos as estimativas desejadas  $\hat{\mu}$  e  $\hat{\nu}$  para cada partida futura.

### 3 METODOLOGIA

Nesta seção vamos mostrar as expressões utilizadas para calcular as previsões de partidas simples. Além disso, descreveremos o procedimento de simulação para calcular diversas outras probabilidades de interesse, tais como: de ser campeão, rebaixamento, classificados para as competições europeias. Avaliamos a qualidade das previsões pela Medida de DeFinetti e as estimativas dos parâmetros.

#### 3.1 PREVISÃO DE PARTIDAS SIMPLES

Em uma partida de futebol, o vencedor de um jogo é sempre aquela equipe que marcar mais gols. Se o número de gols é o mesmo, chamamos, o resultado de empate. Para um dado jogo entre os times  $A$  e  $B$ , a probabilidade de vitória ( $P_V$ ), empate ( $P_E$ ) e derrota ( $P_D$ ) para o time  $A$  contra o time  $B$  é obtida utilizando as expressões:

$$P_V = P(X_{AB} > X_{BA}) = \sum_{i=1}^{\infty} \sum_{j=0}^{i-1} P(X_{AB} = i)P(X_{BA} = j), \tag{7}$$

$$P_E = P(X_{AB} = X_{BA}) = \sum_{i=0}^{\infty} P(X_{AB} = i)P(X_{BA} = i), \tag{8}$$

e

$$P_D = P(X_{AB} < X_{BA}) = \sum_{j=1}^{\infty} \sum_{i=0}^{j-1} P(X_{AB} = i)P(X_{BA} = j). \tag{9}$$

### 3.2 PREVISÃO PARA O CAMPEONATO INTEIRO

Para uma dada rodada fixa  $i$  do segundo turno,  $i = 20, \dots, 38.$ , temos a atual tabela de classificação dos times (jogos, número de pontos, número de vitórias, número de empate, número de derrota, número de gols marcados, número de gols sofridos e saldo de gols). A previsão para a classificação final foi realizada a partir de um algoritmo de simulação que envolve as seguintes etapas:

#### Algoritmo

- (1) Considere  $r$  o número de campeonatos;
- (2) Para o  $r$ -ésimo campeonato, obter as estimativas dos parâmetros do modelo e simular o placar  $(X_{AB}, X_{BA})$  para cada um dos  $n$  jogos entre os times mandante (A) e visitante (B) a serem jogados,  $n = [38 - (r - 1)] * 10$ ;
- (3) Para cada um dos  $n$  jogos preditos, verifica-se se houve vitória do mandante ( $X_{AB} > X_{BA}$ ), empate ( $X_{AB} = X_{BA}$ ) ou vitória do visitante ( $X_{AB} < X_{BA}$ ). Atribua 3 pontos para o time vitorioso e 1 ponto para ambos os times se ocorreu empate;
- (4) A partir da atual tabela de classificação, construir a tabela final de classificação com os resultados simulados para o  $r$ -ésimo campeonato.

Nesse trabalho consideramos  $r = 1000$  campeonatos simulados. A partir daí podemos calcular a chance de um determinado time ser campeão e de ser rebaixado da seguinte forma:

$$P_{\text{campeão}} = \frac{\#(\text{time terminou em primeiro lugar})}{r} \quad (10)$$

e

$$P_{\text{rebaixamento}} = \frac{\#(\text{time terminou entre os 4 últimos colocados})}{r}, \quad (11)$$

em que # refere-se ao número de vezes obtidos na simulação.

### 3.3 QUALIDADE DAS PREVISÕES

Um método de verificação da qualidade das previsões dos resultados de uma partida, que será utilizado é a medida DeFinetti definido por [3], que consiste na consideração de um simplex contido em  $\mathbb{R}^3$  como representação geométrica do conjunto de possíveis previsões probabilísticas. Assim, os vértices desse simplex correspondem às ocorrências dos resultados e os demais pontos a todas as outras possíveis previsões. Formalmente,

$$S = \{(PV, PE, PD) \in \mathbb{R}^3 : PV + PE + PD = 1, PV \geq 0, PE \geq 0, PD \geq 0\}. \quad (12)$$

A medida de distância de DeFinetti corresponde à distância euclidiana quadrática entre o ponto correspondente à (distribuição) probabilidade prevista e o vértice de uma previsão, um índice dado pela média aritmética das distâncias de DeFinetti, é chamado de medida DeFinetti. Nossa aplicação no futebol, associam-se os vértices  $(1, 0, 0)$ ,  $(0, 1, 0)$  e  $(0, 0, 1)$  à vitória da equipe mandante, ao empate, e à derrota da equipe mandante, respectivamente. Ao vetor de probabilidades atribuídas para uma determinada partida associa-se o ponto  $(PV, PE, PD) \in S$ .

Assim, a distância de DeFinetti será igual a:

$$(PV - 1)^2 + (PE - 0)^2 + (PD - 0)^2 \quad \text{se a equipe mandante vencer a partida;} \quad (13)$$

$$(PV - 0)^2 + (PE - 1)^2 + (PD - 0)^2 \quad \text{se a partida terminar empatada;} \quad (14)$$

$$(PV - 0)^2 + (PE - 0)^2 + (PD - 1)^2 \quad \text{se a equipe mandante perder a partida.} \quad (15)$$

Por exemplo, se a previsão for  $(0.3, 0.6, 0.1)$  e o resultado for um empate  $(0, 1, 0)$ , então a distância de DeFinetti é  $(0.3 - 0)^2 + (0.6 - 1)^2 + (0.1 - 0)^2 = 0.26$ . Nas previsões temos que um padrão comumente utilizado é a atribuição equiprovável de probabilidades ( $PV = PE = PD = 1/3$ ), ou seja, atribuir chances iguais a cada resultado em cada jogo. Para essa atribuição a medida de DeFinetti é igual  $(1/3 - 1)^2 + (1/3 - 0)^2 + (1/3 - 0)^2 = 2/3$ .

Assim, podemos considerar métodos de previsões de qualidade minimamente aceitável, aqueles que apresentam medidas de DeFinetti menores que  $2/3$  e de má-qualidade, aqueles que apresentam medidas maiores que  $2/3$ .

Também, utilizamos a quantidade de número de acertos, isto é, consideramos que um método acerta o resultado de um determinado jogo quando o evento com maior probabilidade estimada é observado, por exemplo, se em uma determinada partida o time visitante vencer e a probabilidade estimada para a vitória do time visitante for maior que as probabilidades de empate e vitória do time mandante, consideraremos um acerto para esta previsão, caso contrário um erro. Da mesma forma para a ocorrência de empate e vitória do time mandante. Entretanto, é importante ressaltar que não é correto associar a eventos com alta probabilidade uma certeza de ocorrência ou a eventos com baixa probabilidade uma certeza de não ocorrência. O nosso objetivo não é afirmar que uma dada previsão é correta ou não, e sim construir uma métrica para o conjunto das previsões.

## 4 RESULTADOS

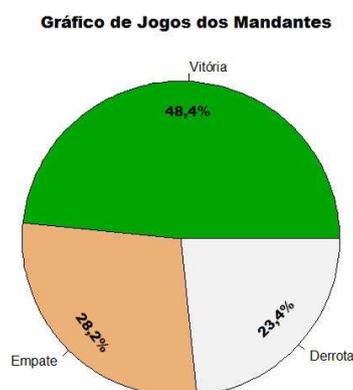
Nesta seção apresentamos inicialmente uma breve análise descritiva dos resultados dos jogos do Campeonato Brasileiro de 2013. Posteriormente, alguns resultados da aplicação do nosso modelo: as previsões de partidas simples e para o campeonato inteiro.

### 4.1 ANÁLISE DESCRITIVA

Nesta seção apresentamos uma análise prévia dos dados. A Tabela 1 indica o turno e quantos jogos cada equipe fizeram como mandante e visitante (os valores estão entre parentes) e o número de gols marcados. Podemos observar que o Cruzeiro tem o melhor ataque tanto como mandante como visitante.

A Tabela 2 refere-se ao total de todos os placares que ocorreram ao longo do campeonato, em que GEM significa os gols marcados pela equipe mandante e GEV os gols marcados pela equipe visitante. Verificou-se que os resultados mais ocorridos durante o campeonato foram  $(1 - 0)$ ,  $(1 - 1)$  e  $(2 - 1)$ , respectivamente.

Considerando todos os jogos do campeonato, observa-se na Figura 1 que os times mandantes apresentaram um maior percentual de vitórias. Portanto, o fator jogar em casa pode influenciar o resultado final de uma partida, seja por terem a maioria da torcida ou por conhecerem melhor o estádio.



**FIGURA 1:** Gráfico de Setores para os jogos realizados por times mandantes.

Times	1º Turno		2º Turno		Geral	
	(Mandante) Gols	(Visitante) Gols	(Mandante) Gols	(Visitante) Gols	(Mandante) Gols	(Visitante) Gols
Bahia	(9) 10	(10) 8	(10) 10	(9) 9	(19) 20	(19) 17
Botafogo	(9) 16	(10) 16	(10) 14	(9) 9	(19) 30	(19) 25
At. Mineiro	(9) 18	(10) 4	(10) 23	(9) 4	(19) 41	(19) 8
At. Paranaense	(9) 15	(10) 19	(10) 24	(9) 7	(19) 39	(19) 26
Coritiba	(10) 15	(9) 8	(9) 14	(10) 5	(19) 29	(19) 13
Corinthians	(10) 11	(9) 8	(9) 6	(10) 2	(19) 17	(19) 10
Criciúma	(10) 15	(9) 10	(9) 11	(10) 13	(19) 26	(19) 23
Cruzeiro	(10) 27	(9) 15	(9) 20	(10) 15	(19) 47	(19) 30
Flamengo	(9) 9	(10) 10	(10) 17	(9) 7	(19) 26	(19) 17
Fluminense	(10) 14	(9) 8	(9) 14	(10) 7	(19) 28	(19) 15
Goiás	(9) 11	(10) 8	(10) 16	(9) 13	(19) 27	(19) 21
Gremio	(10) 16	(9) 10	(9) 8	(10) 8	(19) 24	(19) 18
Internacional	(9) 16	(10) 16	(10) 14	(9) 5	(19) 30	(19) 21
Náutico	(9) 7	(10) 3	(10) 7	(9) 5	(19) 14	(19) 8
Ponte Preta	(10) 9	(9) 11	(9) 10	(10) 7	(19) 19	(19) 18
Portuguesa	(9) 16	(10) 10	(10) 16	(9) 8	(19) 32	(19) 18
Santos	(10) 14	(9) 9	(9) 14	(10) 14	(19) 28	(19) 23
São Paulo	(9) 10	(10) 7	(10) 13	(9) 9	(19) 23	(19) 16
Vasco	(10) 13	(9) 16	(9) 13	(10) 8	(19) 26	(19) 24
Vitória	(10) 17	(9) 6	(9) 15	(10) 21	(19) 32	(19) 27

**TABELA 1:** Número de gols marcados pelas equipes por cada turno.

		GEM				
		0	1	2	3	4+
GEV	0	37	51	29	18	8
	1	24	49	46	10	8
	2	15	21	20	7	2
	3	7	8	7	2	4
	4+	0	3	1	2	0

TABELA 2: Total de placares do campeonato.

#### 4.2 PREVISÃO DE PARTIDA SIMPLES

Como conjunto de dados, inicialmente consideramos os resultados dos jogos do primeiro turno do campeonato, pelo fato de termos uma observação do confronto dos times entre si. Assim, foram calculadas as previsões de todas as rodadas do segundo turno do campeonato, ou seja, a partir dos resultados das rodadas anteriores, foram calculados as probabilidades de uma vitória, empate e derrota de cada equipe para as rodadas seguintes.

A medida DeFinetti [3] associadas a essas previsões obteve uma média igual a 0.609 e o total de acertos foi 106 para os 200 jogos previstos.

Por exemplo, na Tabela 3 apresentamos a previsão para cada uma das partidas simples da rodada 28 juntamente com o placar observado, a medida de DeFinetti e se ocorreu acerto.

Analisando a Tabela 3 acima podemos observar que as chances de vitória do Cruzeiro eram maiores na partida por ser o líder do campeonato, mas o fator mandante pesou na partida e o Cruzeiro como visitante acabou perdendo. Outro caso foi o Internacional que tinha 71.6% de chances de ganhar a partida e realmente obteve a vitória. Dos dez jogos realizados na rodada 28 podemos observar que 7 tinham maiores chances de vitórias dos mandantes, em que os outros 3, um jogo era clássico estadual e tinha maior chance de empate (o que realmente aconteceu e a previsão acertou) e os outros 2 jogos com maior chance de vitória dos visitantes foram os jogos do primeiro e segundo colocado do campeonato.

#### 4.3 PREVISÃO PARA O CAMPEONATO INTEIRO

Para obter as outras previsões de interesses tais como a chance que cada equipe tem de ser campeã, de classificação para a Taça Libertadores da América e de rebaixamento, realizamos uma simulação do campeonato inteiro. Na Tabela 4, apresentamos as probabilidades, de cada um dos sete melhores times do primeiro turno de se tornar campeão da competição antes de cada rodada da segunda fase. Pode-se observar, que desde o começo do segundo turno o Cruzeiro tem mais de 50% de probabilidade de ser campeão.

Na Tabela 5, analisamos a probabilidade de cada uma das sete melhores equipes do primeiro turno de conquistar uma vaga para a Taça Libertadores da América. Foi observado que o Cruzeiro assegurou a vaga com bastante antecedência, já o Grêmio conseguiu a vaga na penúltima rodada do campeonato e, para a última rodada, haviam 4 times com chances de conquistar uma das 2 últimas vagas para a competição. Estavam na disputa os times At. Paranaense, Botafogo e Goiás, e no final do campeonato as vagas acabaram ficando com At. Paranaense que tinha 94% de chances de se classificar e a outra vaga acabou ficando com o Botafogo, apesar de ter menores chances de se classificar do que o Goiás. Juntamente com os quatro primeiros colocados do campeonato brasileiro que foram classificados para a fase de grupos ou repescagem da Taça Libertadores da América, estavam também classificados para a fase de grupos as equipes de Flamengo, campeão da Copa do Brasil de 2013 e o At. Mineiro campeão da Taça Libertadores da América de 2013.

Probabilidades							
Mandante	Visitante	Vitória	Empate	Derrota	Placar	Definetti	Acertou
Botafogo	Flamengo	0.558	0.242	0.201	2 - 1	0.294	Sim
Fluminense	Grêmio	0.333	0.323	0.344	1 - 1	0.687	Não
São Paulo	Corinthians	0.312	0.482	0.206	0 - 0	0.409	Sim
Vitória	Coritiba	0.498	0.241	0.262	2 - 1	0.378	Sim
Internacional	Náutico	0.716	0.185	0.099	4 - 1	0.125	Sim
Goiás	Bahia	0.476	0.293	0.230	3 - 1	0.413	Sim
Criciúma	Vasco	0.443	0.224	0.332	3 - 2	0.471	Sim
At. Mineiro	Cruzeiro	0.227	0.236	0.537	1 - 0	0.942	Não
At. Paranaense	Portuguesa	0.552	0.206	0.243	1 - 0	0.302	Sim
Santos	Ponte Preta	0.617	0.240	0.143	2 - 1	0.224	Sim

**TABELA 3:** Probabilidades de vitória, empate e derrota, medida de Definetti e acertos para a rodada 28.

Apresentamos na Tabela 6 as probabilidades de rebaixamento das oito piores equipes do primeiro turno. Ao término das 38 rodadas, os quatro últimos colocados são rebaixados para segunda divisão de campeonato brasileiro. Considerando os resultados dos jogos dentro de campo (deixaremos de fora a perda de pontos e a decisão do STJD) os times rebaixados foram Fluminense, Vasco, Ponte Preta e Náutico. O Náutico esteve no segundo turno inteiro com mais de 90% de probabilidade de ser rebaixado, o que acabou acontecendo na rodada 29.

A Figura 2 apresenta o Boxplot da pontuação das equipes antes da 32ª rodada. Depois de feito as simulações podemos observar que o Náutico está bem abaixo na pontuação das demais equipes e, o Cruzeiro mostrando seu favoritismo para ser campeão, nas posições intermediárias as equipes estão próximas.

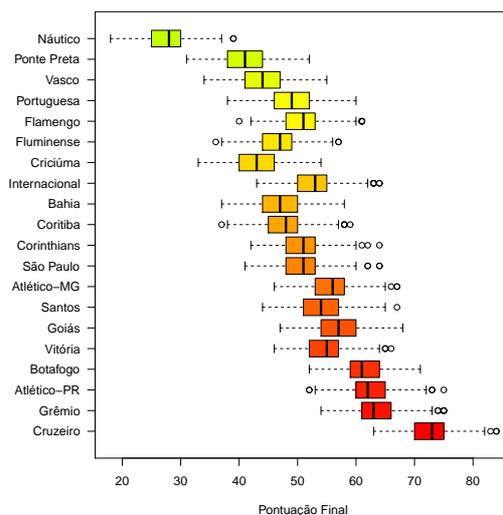


FIGURA 2: Gráfico boxplot da pontuação final prevista na rodada 31.

A Tabela 7 apresenta a previsão da classificação final com base nas 31 primeiras rodadas em comparação com as reais posições (resultado ocorrido apresentado entre parênteses). Para construir esta tabela, foram previstos 80 partidas, e observamos na simulação que o Cruzeiro seria campeão com uma diferença um pouco menor para o segundo colocado e também podemos perceber as equipes de Fluminense e Ponte Preta não seriam rebaixados, dando lugar aos times Coritiba e Criciúma que aliás, também estiveram muito próximos do rebaixamento ao final do Brasileirão de 2013.

Podemos também calcular as probabilidades de cada equipe terminar o campeonato em cada uma das 20 posições. Por exemplo, na Tabela 8 abaixo apresentamos os resultados obtidos para 31ª rodada. Observamos que, nesta rodada, a probabilidade do Cruzeiro tornar-se campeão era de 99.1% e o Grêmio, At. Paranaense e Botafogo estavam na luta para ser o segundo colocado, com 45.9%, 34.9% e 15.9% de probabilidade, respectivamente.

Rodada	Cruzeiro	Grêmio	At. Paranaense	Botafogo	Goiás	Corinthians	Internacional
19	0.440	0.093	0.184	0.157	0.013	0.025	0.033
20	0.506	0.090	0.124	0.197	0.002	0.010	0.031
21	0.565	0.109	0.077	0.197	0.003	0.007	0.018
22	0.634	0.058	0.031	0.243	0.002	0.000	0.026
23	0.736	0.047	0.045	0.148	0.000	0.000	0.018
24	0.752	0.038	0.089	0.101	0.002	0.000	0.004
25	0.863	0.044	0.032	0.053	0.001	0.000	0.000
26	0.915	0.042	0.009	0.027	0.000	0.001	0.000
27	0.960	0.031	0.005	0.002	0.000	0.000	0.000
28	0.947	0.033	0.009	0.010	0.000	0.000	0.000
29	0.914	0.034	0.022	0.028	0.000	0.000	0.000
30	0.938	0.036	0.024	0.002	0.000	0.000	0.000
31	0.950	0.035	0.008	0.006	0.001	0.000	0.000
32	0.985	0.008	0.002	0.005	0.000	0.000	0.000
33	1.000	0.000	0.000	0.000	0.000	0.000	0.000
34	1.000	0.000	0.000	0.000	0.000	0.000	0.000
35	1.000	0.000	0.000	0.000	0.000	0.000	0.000
36	1.000	0.000	0.000	0.000	0.000	0.000	0.000
37	1.000	0.000	0.000	0.000	0.000	0.000	0.000
38	1.000	0.000	0.000	0.000	0.000	0.000	0.000

**TABELA 4:** Probabilidades das equipes em conquistar o título para as rodadas 19 até 38.

Rodada	Cruzeiro	At. Paranaense	Botafogo	Gremio	Vitória	At. Mineiro	Goiás
19	0.843	0.619	0.602	0.471	0.051	0.113	0.164
20	0.901	0.588	0.709	0.506	0.047	0.059	0.088
21	0.943	0.547	0.764	0.628	0.034	0.123	0.058
22	0.969	0.448	0.864	0.598	0.052	0.193	0.090
23	0.990	0.625	0.843	0.554	0.082	0.125	0.095
24	0.993	0.706	0.745	0.526	0.097	0.184	0.122
25	0.996	0.604	0.685	0.656	0.133	0.294	0.090
26	1.000	0.557	0.671	0.786	0.199	0.263	0.045
27	1.000	0.684	0.577	0.886	0.115	0.278	0.046
28	1.000	0.709	0.750	0.870	0.077	0.171	0.051
29	1.000	0.762	0.821	0.814	0.073	0.207	0.075
30	1.000	0.867	0.686	0.891	0.120	0.110	0.149
31	1.000	0.746	0.663	0.883	0.109	0.184	0.231
32	1.000	0.704	0.807	0.749	0.164	0.080	0.376
33	1.000	0.837	0.636	0.688	0.088	0.143	0.522
34	1.000	0.967	0.506	0.604	0.269	0.123	0.496
35	1.000	0.892	0.404	0.761	0.079	0.171	0.664
36	1.000	0.663	0.541	0.865	0.095	0.030	0.806
37	1.000	0.940	0.509	0.846	0.040	0.044	0.621
38	1.000	0.938	0.340	1.000	0.194	0.074	0.454

**TABELA 5:** Probabilidades das equipes em conquistar uma vaga para a Taça Libertadores da América para as rodadas 19 até 38.

Rodada	Vasco	Bahia	Fluminense	Criciúma	Coritiba	Internacional	Ponte Preta	Náutico
19	0.139	0.183	0.389	0.141	0.085	0.045	0.627	0.932
20	0.144	0.185	0.233	0.194	0.043	0.025	0.680	0.938
21	0.199	0.185	0.265	0.204	0.069	0.014	0.800	0.974
22	0.300	0.223	0.185	0.254	0.065	0.004	0.783	0.981
23	0.398	0.158	0.130	0.377	0.074	0.014	0.732	0.997
24	0.460	0.076	0.113	0.466	0.079	0.021	0.809	0.998
25	0.496	0.093	0.048	0.616	0.126	0.025	0.735	0.995
26	0.370	0.118	0.048	0.623	0.172	0.052	0.810	0.978
27	0.357	0.149	0.083	0.611	0.247	0.022	0.845	0.996
28	0.323	0.069	0.180	0.619	0.189	0.062	0.815	0.997
29	0.407	0.129	0.146	0.423	0.285	0.019	0.901	1.000
30	0.505	0.204	0.234	0.529	0.348	0.013	0.796	1.000
31	0.575	0.264	0.231	0.688	0.179	0.011	0.823	1.000
32	0.691	0.266	0.323	0.796	0.066	0.020	0.696	1.000
33	0.471	0.264	0.463	0.821	0.099	0.015	0.710	1.000
34	0.499	0.253	0.573	0.670	0.088	0.001	0.814	1.000
35	0.654	0.402	0.359	0.404	0.126	0.002	0.943	1.000
36	0.832	0.248	0.205	0.298	0.353	0.018	0.994	1.000
37	0.677	0.098	0.404	0.321	0.423	0.004	1.000	1.000
38	0.637	0.000	0.810	0.221	0.331	0.001	1.000	1.000

**TABELA 6:** Chances de equipes com o risco do rebaixamento para as rodadas 19 até 38.

Equipes	Pontos	Jogos	Vitórias	Empates	Derrotas	Gols Pro	Gols Contra	Saldo
Cruzeiro	68 (76)	38	20 (23)	8 (7)	10 (8)	71 (77)	42 (37)	29 (40)
Grêmio	62 (65)	38	18 (18)	8 (11)	12 (9)	45 (42)	35 (35)	10 (7)
Atlético-PR	65 (64)	38	18 (18)	11 (10)	9 (10)	60 (65)	48 (49)	12 (16)
Botafogo	66 (61)	38	19 (17)	9 (10)	10 (11)	58 (55)	43 (41)	15 (14)
Vitória	55 (59)	38	15 (16)	10 (11)	13 (11)	51 (59)	52 (53)	-1 (6)
Goiás	63 (59)	38	17 (16)	12 (11)	9 (11)	51 (48)	41 (44)	10 (4)
Santos	52 (57)	38	13 (15)	13 (12)	12 (11)	47 (51)	41 (38)	6 (13)
Atlético-MG	52 (57)	38	14 (15)	10(12)	14(11)	41 (49)	40 (38)	1 (11)
São Paulo	48 (50)	38	13 (14)	9 (8)	16 (16)	39 (39)	40 (40)	-1 (-1)
Corinthians	47 (50)	38	10 (11)	17 (17)	11 (10)	25 (27)	22 (22)	3 (5)
Flamengo	46 (49)	38	11 (12)	13 (13)	14 (13)	43 (43)	48 (46)	-5 (-3)
Portuguesa	52 (48)	38	14 (12)	10 (12)	14 (14)	60 (50)	55(46)	5 (4)
Coritiba	42 (48)	38	10 (12)	12 (12)	16 (14)	44 (42)	53 (45)	-9 (-3)
Bahia	56 (48)	38	15 (12)	11 (12)	12 (14)	48 (37)	45 (45)	3 (-8)
Internacional	56 (48)	38	14 (11)	14 (15)	10 (12)	62 (51)	54 (52)	8 (-1)
Criciúma	36 (46)	38	10 (13)	6 (7)	22 (18)	45 (49)	69 (63)	-24 (-14)
Fluminense	53 (46)	38	14 (12)	11 (10)	13 (16)	44 (43)	45 (47)	-1 (-4)
Vasco	41 (44)	38	10 (11)	11 (11)	17 (16)	50 (50)	63 (61)	-13 (-11)
Ponte Preta	47 (37)	38	13 (9)	8 (10)	17 (19)	49 (37)	55 (55)	-6 (-18)
Náutico	28 (20)	38	7 (5)	7 (5)	24 (28)	28 (22)	70 (79)	-42 (-57)

TABELA 7: Classificação final prevista do campeonato na rodada 31.

Time	Média	D.P.	1º	2º	3º	4º	5º	6º	7º	8º	9º	10º	11º	12º	13º	14º	15º	16º	17º	18º	19º	20º
Cruzeiro	1.010	0.109	0.991	0.008	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Gremio	2.807	0.964	0.009	0.459	0.326	0.143	0.049	0.013	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
At. Paranaense	3.159	1.219	0.000	0.349	0.333	0.208	0.062	0.032	0.007	0.004	0.003	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Botafogo	4.051	1.504	0.000	0.159	0.214	0.298	0.181	0.085	0.030	0.026	0.004	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Vitória	8.773	2.633	0.000	0.001	0.008	0.030	0.069	0.104	0.132	0.135	0.136	0.111	0.115	0.069	0.054	0.019	0.013	0.004	0.000	0.000	0.000	0.000
Goiás	6.165	2.115	0.000	0.016	0.051	0.137	0.238	0.197	0.121	0.093	0.059	0.049	0.022	0.016	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Santos	8.090	2.381	0.000	0.001	0.013	0.035	0.093	0.126	0.178	0.131	0.145	0.111	0.080	0.047	0.029	0.008	0.002	0.001	0.000	0.000	0.000	0.000
At. Mineiro	6.342	1.959	0.000	0.007	0.045	0.118	0.189	0.213	0.178	0.113	0.061	0.049	0.017	0.009	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
São Paulo	11.038	2.656	0.000	0.000	0.001	0.005	0.012	0.033	0.038	0.085	0.116	0.127	0.141	0.137	0.122	0.090	0.054	0.023	0.010	0.006	0.000	0.000
Corinthians	9.039	2.400	0.000	0.000	0.004	0.011	0.043	0.085	0.138	0.162	0.144	0.142	0.113	0.070	0.054	0.023	0.008	0.000	0.001	0.002	0.000	0.000
Flamengo	12.160	2.554	0.000	0.000	0.000	0.001	0.007	0.015	0.019	0.042	0.068	0.098	0.143	0.139	0.127	0.146	0.118	0.051	0.021	0.004	0.001	0.000
Portuguesa	11.995	2.752	0.000	0.000	0.000	0.003	0.004	0.023	0.034	0.049	0.082	0.093	0.108	0.154	0.149	0.118	0.081	0.060	0.031	0.009	0.002	0.000
Internacional	9.402	2.653	0.000	0.000	0.004	0.011	0.052	0.072	0.113	0.145	0.135	0.128	0.117	0.098	0.056	0.036	0.019	0.010	0.004	0.000	0.000	0.000
Coritiba	14.344	2.375	0.000	0.000	0.000	0.000	0.001	0.000	0.007	0.005	0.021	0.034	0.051	0.095	0.117	0.160	0.163	0.157	0.108	0.073	0.008	0.000
Bahia	15.294	2.003	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.003	0.006	0.009	0.023	0.045	0.081	0.144	0.179	0.210	0.182	0.093	0.023	0.000
Criciúma	17.306	1.587	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.003	0.007	0.017	0.033	0.065	0.121	0.197	0.318	0.237	0.000
Fluminense	14.066	2.223	0.000	0.000	0.000	0.000	0.000	0.002	0.002	0.007	0.018	0.041	0.060	0.096	0.150	0.167	0.174	0.158	0.084	0.033	0.008	0.000
Vasco	16.715	1.753	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.007	0.016	0.037	0.046	0.102	0.161	0.246	0.254	0.128	0.000
Ponte Preta	18.244	1.183	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.007	0.008	0.022	0.044	0.116	0.208	0.593	0.000
Náutico	20.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

**TABELA 8:** Probabilidade de cada equipe terminar em cada uma das 20 posições.

## 5 CONCLUSÃO

Neste trabalho realizamos uma aplicação de uma metodologia simples para a previsão de resultado de jogos de futebol. Aplicando no Campeonato Brasileiro de 2013, a metodologia estudada apresentou bons resultados preditivos (considerando todos os jogos previstos foram 50% de acertos e a medida Definetti menor que  $2/3$ ). Com o procedimento de simulação foi possível prever com boa precisão o campeão (Cruzeiro), e obteve resultados satisfatórios para os classificados para a Copa Libertadores da América como também para os times rebaixados. Essa metodologia pode ser usada para outros campeonatos de pontos corridos, como por exemplo o Campeonato Espanhol (Liga BBVA), o Campeonato Italiano e o Campeonato Inglês.

É importante ressaltar que o modelo utilizado neste trabalho não leva em consideração vários fatores que podem influenciar o resultado de um determinado jogo, tais como a condição atmosférica, o esquema tático, suspensões, entre outros.

Todas as implementações computacionais foram todas realizadas utilizando o sistema R [10]. E, os resultados dos jogos do Campeonato Brasileiro de 2013 podem ser encontrados na página da Confederação Brasileira de Futebol (CBF): <http://www.cbf.com.br/>.

## REFERÊNCIAS

- [1] L. Arruda: *Poisson, Bayes, Futebol e DeFinetti*. Dissertação apresentada ao Instituto de Matemática e Estatística da Universidade de São Paulo para a obtenção do grau de mestre em Estatística, 2000.
- [2] D. R. Brillinger: *Modelling game outcomes of the Brazilian 2006 Series A Championship as Ordinal-Valued*. Brazilian Journal of Probability Statistics, 22(2):89–104, 2008.
- [3] B. DeFinetti: *Probability, Induction and Statistics*. John Wiley, London, 1972.
- [4] D. Dyte e S. R. Clarke: *A Ratings Based Poisson Model for World Cup Soccer Simulation*. The Journal of the Operational Research Society, 51(8):993–998, 2000.
- [5] D. Karlis e I. Ntzoufras: *Bayesian Modelling of Football Outcomes: Using the Skellam's Distribution for the Goal Difference*. IMA Journal of Management Mathematics, 20(2):133–145, 2009.
- [6] J. B. Keller: *A Characterization of the Poisson Distribution and the Probability of Winning a Game*. The American Statistician, 48(4):294–298, 1994.
- [7] A. Lee: *Modeling Scores in the Premier League: Is Manchester United Really the Best?* Chance, 10(1):15–19, 1997.
- [8] R. Pollard: *Home advantage in soccer: a retrospective analysis*. Journal of Sports Sciences, pp. 237 – 248, 1986.
- [9] A. Suzuki, L. Salazar, F. Louzada-Neto e J. Leite: *A bayesian approach for predicting match outcomes: The 2006 (Association) Football World Cup*. Journal of the Operational Research Society, 61:1530–1539 (October 2010), 2009.
- [10] R. D. C. Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2007. <http://www.R-project.org>, ISBN 3-900051-07-0.
- [11] P. Volf: *A random point process model for the score in sport matches*. IMA Journal of Management Mathematics, 20:121–131, 2009.