

# MODELAGEM ESTATÍSTICA PARA A PREVISÃO DE JOGOS DE FUTEBOL: UMA APLICAÇÃO NO CAMPEONATO BRASILEIRO DE FUTEBOL 2014

Clodoaldo Tsuyoshi Pereira de Araújo<sup>1</sup>, Leandro Tavares<sup>2</sup>,  
Luis Gustavo Alvares<sup>3</sup>, Francisco Louzada Neto<sup>4</sup>, Adriano Kamimura Suzuki<sup>5</sup>

**Resumo:** *Atualmente, o futebol é um dos esportes coletivos que mais atrai público, gerando grande interesse comercial e movimentando enormes cifras ao redor do mundo. No Brasil, um dos maiores eventos do esporte é o campeonato Brasileiro de Futebol. Neste artigo, modela-se o número de gols marcados pelos times em uma determinada partida por meio do modelo Poisson e aplica-se o método soma e diferença (SD 0) proposto por [1] para estimar os seus parâmetros. Por meio de uma estrutura de simulação pode-se prever toda a segunda fase do Campeonato Brasileiro de Futebol 2014 Série A e obteve-se, para cada time participante, as probabilidades de interesse tais como ser campeão, classificar para a Taça Libertadores 2015 e até a de ser rebaixado para a Série B. Para avaliar a qualidade preditiva do modelo, utilizou-se a medida de DeFinetti.*

## 1 Introdução

O futebol é um esporte mundialmente conhecido e no Brasil é uma paixão nacional, faz parte do dia-a-dia dos cidadãos, podendo ser considerado como um aspecto cultural do povo brasileiro.

O primeiro Campeonato Nacional de Futebol foi realizado em 1971 pela CBD e teve 20 clubes disputando o torneio. Com o surgimento da CBF em 1979, houve a inclusão de todos os estados no Campeonato Brasileiro, tornando o mais abrangente e competitivo.

O futebol tem grande influência na sociedade, tanto em aspectos econômicos quanto culturais, atraindo cada vez mais atenção dos intelectuais de diversas áreas do conhecimento. Nos últimos anos, a Estatística tem produzido trabalhos voltados para criação de modelos para predição de resultados futebolísticos.

Há vários estudos relacionados a campeonatos de pontos corridos. [7] verifica a existência da vantagem de se jogar dentro de casa por meio da porcentagem de vitórias obtida por mandantes e não mandantes, e que em campeonatos de menor expressão essa vantagem é minimizada.

[5] e [6] utilizam a distribuição de Poisson para modelar o número de gols marcados pelos times. Já [4] modelam por meio da distribuição de Skellam.

Sob uma outra abordagem, [2] modela diretamente as probabilidades de vitória, empate e derrota a partir de um modelo trinomial para o Campeonato Brasileiro de 2006. Aliado a um processo de simulação, obtém as previsões para os resultados do campeonato e o número de pontos, calculando a probabilidade de ser campeão e de ficar entre os 4 primeiros colocados.

[10] propuseram uma metodologia bayesiana para a predição dos jogos da Copa do Mundo FIFA 2006, no qual a distribuição *a priori* leva em consideração as opiniões de especialistas e o

---

<sup>1</sup>Instituto de Ciências Matemáticas e de Computação - USP. e-mail: [tsuyoshi@usp.br](mailto:tsuyoshi@usp.br)

<sup>2</sup>Instituto de Ciências Matemáticas e de Computação - USP. e-mail: [leandrotz@usp.br](mailto:leandrotz@usp.br)

<sup>3</sup>Instituto de Ciências Matemáticas e de Computação - USP. e-mail: [lgalvares@usp.br](mailto:lgalvares@usp.br)

<sup>4</sup>Instituto de Ciências Matemáticas e de Computação - USP. e-mail: [louzada@icmc.usp.br](mailto:louzada@icmc.usp.br)

<sup>5</sup>Instituto de Ciências Matemáticas e de Computação - USP. e-mail: [suzuki@icmc.usp.br](mailto:suzuki@icmc.usp.br)

ranking FIFA. Com essa nova metodologia, pode-se calcular as probabilidades de uma partida simples (probabilidade de vitória, empate e derrota de cada time em cada partida) além das probabilidades considerando o campeonato inteiro, tais como a probabilidade de classificação na fase de grupos, de ser campeão da competição, de chegar às oitavas de final, entre outras.

Outra abordagem para a previsão da Copa do Mundo FIFA 2006 pode ser vista em [11], em que a interação entre as equipes é modelada a partir de um modelo de regressão multiplicativa semi paramétrica de intensidade.

Neste trabalho modelou-se o número de gols marcados pelos times em uma determinada partida por meio do modelo Poisson e aplicou-se o método soma e diferença (SD 0) proposto por [1] para estimar os seus parâmetros. Por meio de uma estrutura de simulação pode-se prever os resultados dos jogos do Campeonato Brasileiro de Futebol 2014 Série A, obtendo, para cada time participante, as probabilidades de interesse tais como ser campeão, classificar para a Taça Libertadores 2015 e ser rebaixado para a Série B. A qualidade dessas previsões será dada pela medida DeFinetti ([3]). Todo o estudo de simulação foi realizado usando software [8].

## 2 Modelo

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  $\lambda_Y$ . O método SD 0 proposto por [1] relaciona as médias  $\lambda_X$  e  $\lambda_Y$  e os valores esperados  $E[X + Y]$  e  $E[X - Y]$  utilizando a propriedade linear da esperança, obtendo:

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

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:

$$U_i = \mu_i + \varepsilon_i \quad (3)$$

e

$$V_i = \nu_i + \varepsilon'_i, \quad (4)$$

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:

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

e

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

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

#### 3.1 Previsão de partidas simples

Com os parâmetros  $\hat{\mu}$  e  $\hat{\nu}$  obtidos anteriormente, obtém-se as estimativas,  $\hat{\lambda}_X$  e  $\hat{\lambda}_Y$ , para o número de gols marcados por cada equipe. Pode-se, então, calcular a probabilidade do número de gols de um determinado time ser maior que outro, assim determinando a vitória do primeiro sobre o segundo. A probabilidade de empate é definida como número iguais de gols dos times em uma partida. E por fim, a probabilidade de derrota de uma equipe é calculada assumindo que esta fez menos gols que o adversário. As probabilidades de cada caso é representada por:

$$P_{Vitória} = \sum_{i>j} P(X = i, Y = j) \quad (7)$$

$$P_{Empate} = \sum_{i=j} P(X = i, Y = j) \quad (8)$$

$$P_{Derrota} = \sum_{i<j} P(X = i, Y = j) \quad (9)$$

#### 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) Inicie  $r$  com o número de campeonatos a serem simulados;
- (2) Para o  $r$ -ésimo campeonato, estimar a média *a posteriori* dos parâmetros do modelo e simular o placar,  $(X, Y)$ , em cada um dos  $n$  jogos entre o time mandante e visitante.  $X$  e  $Y$  representam o número de gols do time mandante e visitante respectivamente. Para o campeonato brasileiro temos:  $n = [38 - (i - 1)] * 10$ ;
- (3) Para cada jogo simulado é atribuído o número de pontos para cada time time conforme as regras do campeonato, descritas na tabela 1.

Tabela 1: Tabela de atribuição de pontos aos times conforme resultado das simulações.

Simulação	Resultado	Pontos time mandante	Pontos time visitante
$X > Y$	Vitória	3	0
$X = Y$	Empate	1	1
$X < Y$	Derrota	0	3

- (4) A partir dos pontos atribuídos aos times segundo as simulações realizadas, construir a tabela de classificação do  $r$ -ésimo campeonato. A partir dos  $r$  campeonatos simulados pode-se calcular a chance de um determinado time, por exemplo, ser campeão ou de ser rebaixado segundo:

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

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

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

### 3.3 Medida de avaliação das Previsões

Para avaliar a qualidade das previsões feitas pelo modelo, utilizou-se a medida de DeFinetti, definida em [3]. Se representarmos as probabilidades das previsões dos possíveis resultados de um time em uma partida (vitória, empate, derrota) em um tetraedro retangular contido em  $\mathbb{R}^3$ , em que os vértices representam cada um dos resultados a ser efetivamente observado, escrevendo:

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

Então a distância euclidiana quadrática entre a previsão probabilística para o resultado e o vértice correspondente é definido como medida de DeFinetti, utilizada como métrica para avaliar a qualidade das previsões de modelos.

A cada vértice do tetraedro associa-se um possível resultado da partida para o time com mando de campo, como ilustrado na Figura 1.

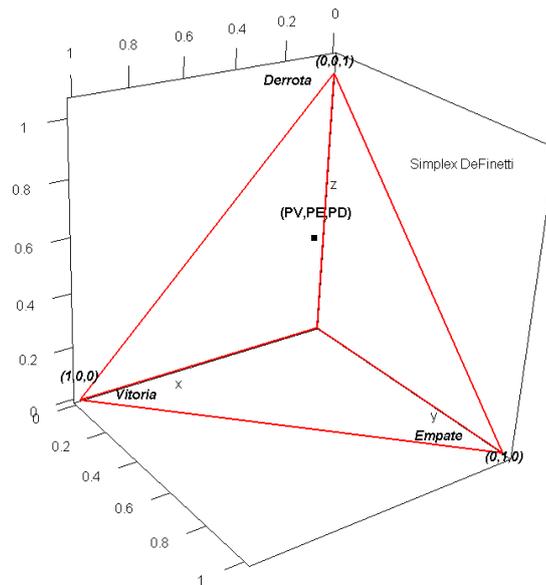


Figura 1: Gráfico ilustrando o simplex de DeFinetti.

Tabela 2: Correspondência entre os vértices e os eventos.

Vértice	Evento
(1,0,0)	Vitória do Time Mandante
(0,1,0)	Empate do Time Mandante
(0,0,1)	Derrota do Time Mandante

Ao vetor de probabilidades atribuídas para uma determinada partida, associa-se o ponto  $(PV, PE, PD) \in S$ .

A distância de DeFinetti é definida como:

$$(PV - 1)^2 + (PE - 0)^2 + (PD - 0)^2 \quad \text{se a equipe mandante vencer a partida;} \quad (12)$$

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

$$(PV - 0)^2 + (PE - 0)^2 + (PD - 1)^2 \quad \text{se a equipe mandante perder a partida.} \quad (14)$$

Para ilustrar o método suponha que as probabilidades previstas para cada um dos resultados possíveis para um time mandante são dadas por  $(0.6, 0.20, 0.20)$  e o resultado efetivo observado da partida é vitória da equipe mandante, que corresponde ao vértice  $(1, 0, 0)$ , então a distância de DeFinetti pode ser calculado pela equação 12 e temos:

$$(0.6 - 1)^2 + (0.20 - 0)^2 + (0.20 - 0)^2 = 0.24 \quad (15)$$

Para previsões no futebol têm se utilizado a atribuição equiprovável de probabilidades  $(PV = PE = PD = 1/3)$ , ou seja, atribui-se chances iguais aos resultados em cada jogo. Para essa atribuição a medida de DeFinetti é igual  $(1/3 - 1)^2 + (1/3 - 0)^2 + (1/3 - 0)^2 = 2/3$ .

Então considera-se métodos de previsão de boa qualidade aqueles que apresentam medidas de DeFinetti menores que  $2/3$ , enquanto os de má qualidade, aqueles que apresentam medidas maiores que  $2/3$ .

Uma outra forma de avaliar a qualidade preditiva do modelo é por meio da quantidade de acertos. Configura-se acerto na situação em que dentre os três possíveis resultados obtidos por um time em uma partida (vitória, empate ou derrota), aquele estimado com maior probabilidade de ocorrência pelo modelo coincide com resultado efetivo observado da partida. Em caso contrário, dizemos que ocorreu um erro de previsão.

## 4 Resultados

Apresenta-se aqui a análise descritiva dos dados utilizados bem como dos resultados obtidos por meio das simulações.

### 4.1 Análise descritiva

Antes de iniciar a simulação foi feito um tratamento inicial dos dados obtidos. Estes podem ser obtidos em qualquer site especializado em futebol. Para a simulação optou-se por usar apenas os dados da primeira fase, tendo assim um número maior de jogos previstos para mensurar a qualidade da predição do modelo, usando a medida DeFinetti. Considerando os jogos da primeira fase, nota-se no gráfico setorial na Figura 2 um número maior de vitórias quando o time analisado é o mandante da partida.



Figura 2: Gráfico Pizza de jogos realizados por times mandantes

Na tabela 3 temos os turnos e quantos jogos cada equipe realizou como mandante e visitante (indicados em parânteses) e o número de gols marcados.

Tabela 3: Número de gols marcados pelas equipes em cada turno

Times	1º Turno		2º Turno		Geral	
	(Mandante) Gols	(Visitante) Gols	(Mandante) Gols	(Visitante) Gols	(Mandante) Gols	(Visitante) Gols
Bahia	(10) 4	(9) 7	(9) 10	(10) 10	(19) 14	(19) 19
Botafogo	(8) 15	(11) 4	(11) 9	(8) 3	(19) 24	(19) 14
Atlético-MG	(10) 12	(9) 12	(9) 16	(10) 10	(19) 28	(19) 19
Atlético-PR	(10) 13	(9) 12	(9) 11	(10) 7	(19) 24	(19) 16
Coritiba	(9) 6	(10) 8	(10) 18	(9) 10	(19) 24	(19) 20
Chapecoense	(10) 8	(9) 5	(9) 16	(10) 10	(19) 24	(19) 19
Corinthians	(10) 16	(9) 8	(9) 16	(10) 9	(19) 32	(19) 18
Criciúma	(10) 8	(9) 1	(9) 11	(10) 8	(19) 19	(19) 17
Cruzeiro	(9) 25	(10) 16	(10) 18	(9) 8	(19) 43	(19) 18
Figueirense	(9) 8	(10) 9	(10) 16	(9) 4	(19) 24	(19) 14
Flamengo	(10) 11	(9) 5	(9) 18	(10) 12	(19) 29	(19) 21
Fluminense	(10) 23	(9) 8	(9) 17	(10) 13	(19) 40	(19) 22
Goiás	(9) 10	(10) 4	(10) 22	(9) 4	(19) 32	(19) 14
Grêmio	(9) 12	(10) 5	(10) 12	(9) 8	(19) 24	(19) 18
Internacional	(10) 17	(9) 8	(9) 20	(10) 8	(19) 37	(19) 17
Palmeiras	(9) 7	(10) 8	(10) 14	(9) 5	(19) 21	(19) 15
Santos	(10) 14	(9) 7	(9) 11	(10) 10	(19) 25	(19) 19
São Paulo	(9) 17	(9) 14	(9) 15	(10) 13	(19) 32	(19) 22
Sport	(9) 12	(10) 4	(10) 11	(9) 9	(19) 23	(19) 19
Vitória	(9) 7	(10) 10	(10) 16	(9) 4	(19) 23	(19) 14

## 4.2 Previsão de partida simples

Foram calculadas as previsões até a trigésima oitava rodada 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 associadas a essas previsões obteve uma média igual a 0.56 e o total de acertos foi 80 para os 151 jogos previstos.

Por exemplo, na Tabela 4 dispõem-se as previsões para cada uma das partidas simples da rodada 29, juntamente com o placar observado, a medida de DeFinetti e se ocorreu acerto da previsão.

Tabela 4: Probabilidades de vitória, empate e derrota, medida de Definetti e acertos para a rodada 29

Probabilidades							
Mandante	Visitante	Placar	Vitória	Empate	Derrota	Deffinetti	Acertou
Fluminense	Criciúma	4 - 2	0.781	0.177	0.042	0.081	Sim
São Paulo	Bahia	2 - 1	0.671	0.218	0.111	0.168	Sim
Goiás	Grêmio	2 - 1	0.414	0.407	0.179	0.54	Sim
Atlético -MG	Chapecoense	1 - 0	0.701	0.202	0.097	0.14	Sim
Palmeiras	Santos	1 - 3	0.368	0.287	0.345	0.647	Não
Atlético -PR	Flamengo	2 - 1	0.526	0.265	0.208	0.338	Sim
Internacional	Corinthians	1 - 2	0.471	0.307	0.222	0.922	Não
Figueirense	Coritiba	4 - 0	0.458	0.294	0.248	0.442	Sim
Botafogo	Sport	1 - 1	0.561	0.256	0.183	0.901	Não
Vitória	Cruzeiro	0 - 1	0.294	0.23	0.476	0.414	Sim

Pela tabela 4, vê-se que no modelo há uma predominância de maiores probabilidades de vitória para os times na condição de mandante, a não ser para o Cruzeiro, líder do campeonato, que nesta rodada jogou como visitante. Das 10 partidas analisadas, o modelo previu sete resultados corretamente. Nas partidas nas quais houveram erros de predição, nota-se uma sensível diminuição da probabilidade de vitória do mandante, mas não o suficiente para que indicasse a vitória do visitante, que ocorreu nos dois outros casos de erro na rodada analisada.

### 4.3 Previsão para o campeonato

Realizou-se a simulação de toda a segunda rodada do campeonato Brasileiro de Futebol 2014 para obter as previsões para o campeão, os classificados para Taça Libertadores da América (4 primeiros colocados) e quais seriam rebaixados. Na Tabela 5 dispõe-se as probabilidades dos sete primeiros classificados da primeira fase do campeonato de se tornarem campeões conforme o andamento da segunda fase. A simulação foi feita antes de cada rodada, considerando o desempenho de cada time nas rodadas anteriores, conforme o modelo proposto.

Tabela 5: Probabilidades das equipes em conquistar o título para as rodadas 19 até 38.

rodada	Cruzeiro	São Paulo	Grêmio	Atlético-MG	Corinthians	Internacional	Fluminense
19	0.661	0.097	0.014	0.014	0.055	0.097	0.046
20	0.657	0.134	0.019	0.011	0.050	0.075	0.036
21	0.718	0.146	0.018	0.012	0.046	0.028	0.020
22	0.612	0.222	0.025	0.012	0.036	0.053	0.030
23	0.762	0.117	0.008	0.015	0.034	0.040	0.019
24	0.682	0.105	0.028	0.021	0.067	0.069	0.017
25	0.746	0.080	0.013	0.028	0.025	0.094	0.007
26	0.722	0.042	0.018	0.040	0.012	0.150	0.014
27	0.840	0.047	0.006	0.02	0.017	0.059	0.007
28	0.782	0.099	0.021	0.014	0.035	0.038	0.006
29	0.777	0.075	0.012	0.026	0.016	0.088	0.002
30	0.825	0.078	0.001	0.033	0.019	0.034	0.004
31	0.807	0.074	0.007	0.031	0.052	0.016	0.012
32	0.762	0.117	0.001	0.044	0.034	0.026	0.016
33	0.837	0.116	0.001	0.010	0.004	0.020	0.012
34	0.865	0.119	0	0.004	0.007	0.003	0.002
35	0.907	0.091	0.001	0	0.001	0	0
36	0.986	0.014	0	0	0	0	0
37	1	0	0	0	0	0	0
38	1	0	0	0	0	0	0

Nota-se que o Cruzeiro apresenta mais de 60% de probabilidade de se tonar campeão já no início da segunda fase, mostrando ser o grande favorito para a competição. Na Tabela 6, apresenta-se a probabilidade de cada uma das sete melhores equipes do primeiro turno de conquistar uma vaga para a Taça Libertadores da América.

Tabela 6: Probabilidades de conquistar vaga para Taça Libertadores para as rodadas 19 até 38.

rodada	Cruzeiro	São Paulo	Grêmio	Atlético-MG	Corinthians	Internacional	Fluminense
19	0.957	0.574	0.187	0.190	0.467	0.606	0.402
20	0.951	0.688	0.294	0.255	0.442	0.505	0.386
21	0.982	0.774	0.349	0.164	0.568	0.42	0.322
22	0.958	0.828	0.314	0.159	0.441	0.505	0.416
23	0.991	0.826	0.323	0.244	0.456	0.514	0.324
24	0.968	0.681	0.354	0.280	0.53	0.547	0.277
25	0.990	0.655	0.327	0.441	0.382	0.746	0.250
26	0.979	0.500	0.433	0.491	0.269	0.822	0.315
27	0.994	0.682	0.292	0.381	0.423	0.734	0.294
28	0.990	0.747	0.436	0.316	0.505	0.579	0.216
29	0.990	0.708	0.314	0.509	0.434	0.729	0.143
30	0.996	0.776	0.174	0.567	0.510	0.601	0.181
31	0.998	0.729	0.292	0.548	0.641	0.412	0.264
32	0.995	0.814	0.155	0.622	0.562	0.468	0.341
33	0.999	0.910	0.245	0.374	0.403	0.607	0.459
34	1	0.958	0.337	0.540	0.577	0.350	0.238
35	1	0.986	0.455	0.286	0.625	0.411	0.237
36	1	0.997	0.182	0.567	0.862	0.335	0.057
37	1	1	0.034	0.266	0.994	0.676	0.030
38	1	1	0	0	1	1	0

Já na Tabela 7 estão dispostas as probabilidades de rebaixamento das oito equipes com as piores campanhas do primeiro turno. Ao término das 38 rodadas, os quatro últimos colocados são rebaixados para segunda divisão de campeonato brasileiro. Interessante notar que segundo a simulação para a última rodada a equipe do Palmeiras apresentava probabilidade de ser rebaixado, o que não aconteceu apenas por que a equipe do Santos venceu o Vitória, que foi rebaixado no final do campeonato.

Tabela 7: Chances de equipes com o risco do rebaixamento para as rodadas 19 até 38.

rodada	Palmeiras	Coritiba	Chapecoense	Vitória	Botafogo	Bahia	Criciúma	Figueirense
19	0.475	0.500	0.364	0.597	0.161	0.554	0.527	0.242
20	0.491	0.572	0.345	0.68	0.191	0.576	0.528	0.134
21	0.354	0.407	0.406	0.552	0.304	0.674	0.622	0.155
22	0.441	0.527	0.289	0.648	0.367	0.560	0.501	0.233
23	0.508	0.394	0.356	0.546	0.481	0.410	0.557	0.257
24	0.595	0.474	0.411	0.378	0.495	0.485	0.530	0.268
25	0.477	0.612	0.305	0.507	0.361	0.368	0.655	0.192
26	0.570	0.675	0.305	0.611	0.412	0.214	0.691	0.085
27	0.438	0.571	0.425	0.491	0.536	0.249	0.583	0.145
28	0.310	0.434	0.277	0.509	0.684	0.340	0.680	0.234
29	0.219	0.622	0.180	0.456	0.604	0.524	0.574	0.384
30	0.251	0.727	0.210	0.539	0.589	0.602	0.652	0.191
31	0.245	0.488	0.205	0.596	0.708	0.612	0.744	0.245
32	0.204	0.493	0.201	0.457	0.547	0.758	0.868	0.246
33	0.076	0.468	0.272	0.496	0.576	0.850	0.916	0.304
34	0.093	0.23	0.323	0.604	0.764	0.907	0.966	0.092
35	0.130	0.283	0.398	0.342	0.848	0.953	0.992	0.054
36	0.281	0.336	0.168	0.341	0.943	0.924	1	0.007
37	0.410	0.036	0.007	0.561	0.999	0.987	1	0
38	0.434	0	0	0.698	1	0.868	1	0

A Figura 3 apresenta o Boxplot da pontuação das equipes antes da 32ª rodada. Depois de feito as simulações, pode-se observar que as pontuações indicam o Cruzeiro como favorito ao título de campeão, enquanto que na outra ponta vemos como rebaixados o Criciúma e o Bahia.

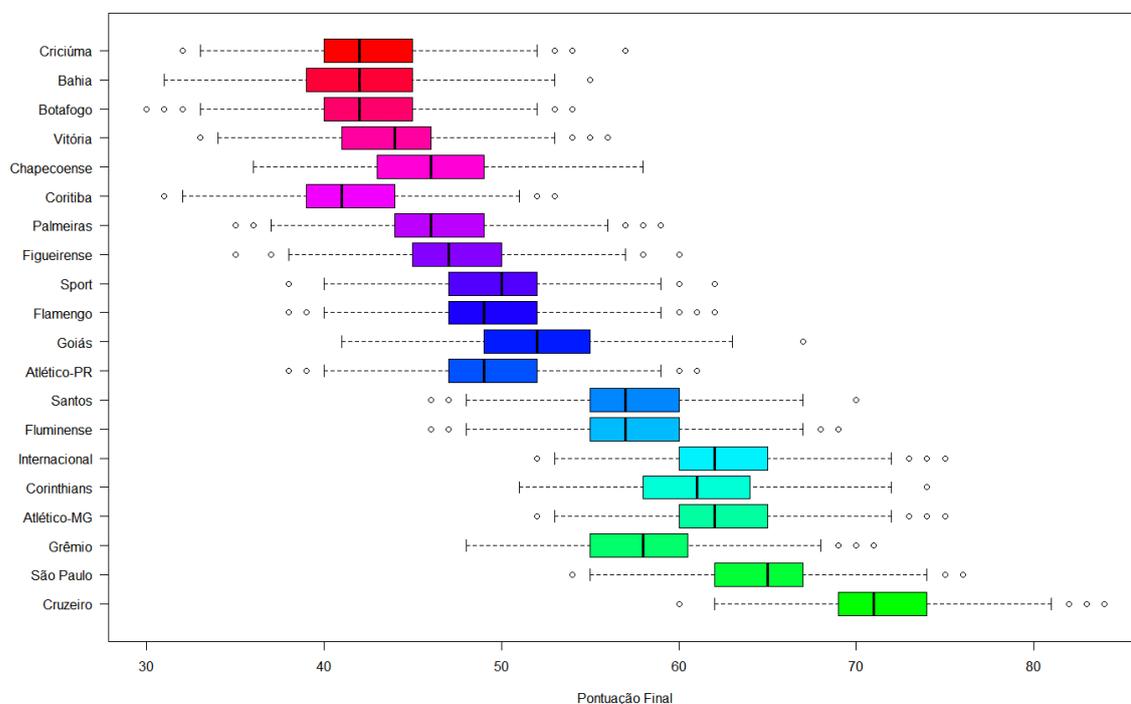


Figura 3: Gráfico boxplot da pontuação das equipes até a rodada 31

## 5 Conclusão

Aplicando a metodologia de Soma e Diferença (SD), proposto em [1], para as previsões dos resultados do Campeonato Brasileiro de 2014, foi obtida uma taxa próxima a 50% de resultados corretos e uma medida Definetti abaixo de  $2/3$ , indicando uma boa qualidade de previsão do modelo até a trigésima terceira rodada.

Como observou-se nos resultados, havia grandes chances do Cruzeiro ser campeão logo no início da segunda fase, assim como, do Criciúma jogar na segunda divisão do brasileiro no próximo ano.

Este tipo de modelo obteve boa qualidade preditivo quando aplicado a campeonatos de pontos corridos, que é o caso do campeonato Brasileiro. No entanto, pode ser aplicado em outros tipos de campeonatos tais como o mata-mata e o eliminatório. Além disso, há a possibilidade de incorporar outras covariáveis ao modelo, de forma a melhorar o poder preditivo deste, como foi feito por [9], que acrescentou a variável crise.

## Referências

- [1] ARRUDA, M. L. **Poisson, Bayes, Futebol e DeFinetti**. 2000. 123 p. Dissertação (Mestrado em Estatística), Instituto de Matemática e Estatística da Universidade de Sao Paulo, São Paulo.
- [2] BRILLINGER, D. R. Modelling game outcomes of the Brazilian 2006 Series A Championship as ordinal-valued. **Brazilian Journal of Probability and Statistics**. v. 22, p. 89-104, 2008.
- [3] DE FINETTI, B. *Probability, Induction and Statistics: The Art of Guessing*. J. Wiley, London, 1972.
- [4] KARLIS, D.; NTZOUFRAS, I. Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. **Ima Journal of Management Mathematics**. v. 20(2), p. 133-145, 2009.
- [5] KELLER, J. B. A Characterization of the Poisson-distribution and the Probability of Winning a Game. **American Statistician**. v.48(4), p.294-298, 1994.
- [6] LEE, A. Modeling Scores in the Premier League: Is Manchester United Really the Best? **Chance**. v.10(1), p.15-19, 1997.
- [7] POLLARD, R. Home advantage in soccer: A retrospective analysis. **Journal of Sports Sciences**. v.4(3), p.237-248, 1986.
- [8] R Development Core Team. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>, 2007.
- [9] SUZUKI, A. K. **Modelagem Estatística para determinação de Resultados de Dados Esportivos**. 2007. 70 p. Dissertação (Mestrado em Estatística) da Universidade Federal de São Carlos, São Carlos.
- [10] SUZUKI, A. K.; SALASAR, L. E. B.; LOUZADA-NETO, F.; LEITE, J. G. A bayesian approach for predicting match outcomes: The 2006 (Association) Football World Cup. **Journal of the Operational Research Society**, v.61, p.1530-1539, 2009.
- [11] VOLF, P. A random point process model for the score in sport matches. **Ima Journal of Management Mathematics**, v.20(2), p.121-131, 2009.