



SPOLM 2008

ISSN 2175-6295

Rio de Janeiro- Brasil, 05 e 06 de agosto de 2008.

UMA PROPOSTA DE PREVISÃO DE RESULTADOS PARA O CAMPEONATO BRASILEIRO ATRAVÉS DO MODELO LOGITO

Alessandro Martins Alves

Universidade Federal Fluminense
Rua Passo da Pátria 156, São domingos, Niterói
alessandro.martinsalves@yahoo.com.br

Thiago Graça Ramos

Universidade Federal Fluminense
Rua Passo da Pátria 156, São domingos, Niterói
tgramos@globo.com

João Carlos C.B Soares de Mello

Universidade Federal Fluminense
Rua Passo da Pátria 156, São domingos, Niterói
jcsmello@pesquisador.cnpq.br

Annibal Parracho Sant'Anna

Universidade Federal Fluminense
Rua Passo da Pátria 156, São domingos, Niterói
tppaps@vm.uff.br

RESUMO

O presente trabalho utiliza o modelo logit ordinal para calcular as probabilidades de cada um dos resultados possíveis (vitória, empate e derrota) para cada jogo não realizado no campeonato Brasileiro de 2007, com base nos jogos realizados. Foi estimado um modelo a cada rodada, considerando os dados até a rodada e prevendo as rodadas seguintes. Com base nas probabilidades derivadas de cada modelo, foram simulados 10.000 resultados possíveis e calculado o percentual de vezes que cada time aparecia entre os quatro primeiros colocados (classificados para a copa Libertadores da América do ano seguinte) e entre os 4 últimos colocados (rebaixados para a série B). O modelo logit utilizado considera como variáveis preditivas a vantagem de se jogar em casa (Home Advantage) e a força individual de cada time.

Palavras-Chaves: Home Advantage, Logit ordinal, Previsões em futebol

ABSTRACT

The following paper uses the logit model to figure out the probabilities of each possible result (win, draw or lose) for each game not realized in the Brazilian Soccer Championship 2007, using the realized ones previously. It was estimated one model for each turn, taking into consideration the data until the current one and forecasting the following ones. Based on the results of each model, it was simulated 10.000 possible results and calculated the percentage of times that each team was displayed between the first fours places in the ranking (classified for the Libertadores da América Championship) and the last four places (which were relegated to a lower division). The used logit model considered as independent variables the home advantage and the individual team strength.

Keywords:

Home Advantage, Ordinal Logit, Football Forecasts;

1. INTRODUÇÃO

A previsão dos resultados no futebol possui um papel fundamental no planejamento e na tomada de decisões sobre a postura e evolução do time ao longo do campeonato. Nos últimos anos, muito se dedicou a previsão de resultados esportivos para o desenvolvimento de estratégias de apostas, como, por exemplo, Craig e Hall (1994); Lee (1997) que estudou o futebol inglês e Stefani e Clarke (1992), que estudaram o futebol australiano.

Scarf e Shi (2008) avaliaram a importância dos jogos na liga inglesa de 2001 a 2004; com base nas previsões, os times poderiam determinar o esforço que deveria ser investido nos jogos, pois, em campeonatos do tipo *round robin*, cada jogo é importante para o resultado final do campeonato e, dependendo da importância desse jogo no resultado final do campeonato para o time, talvez fosse melhor poupar determinados jogadores chave para outros campeonatos que o time estivesse disputando. A previsão da importância de cada jogo é fundamental também para as televisões escolherem quais jogos transmitir.

A previsão dos resultados consiste na tentativa de extrapolar o comportamento futuro a partir dos dados históricos já observados. Para efeito de previsão, é de suma importância identificar que variáveis ou informações serão relevantes e fazem-se necessárias nos modelos preditivos.

Nos últimos anos, diversos trabalhos foram desenvolvidos para verificar a existência do *home advantage*, ou vantagem de jogar em casa em diversos esportes. Essa vantagem impacta no resultado final dos jogos e é defendida por diversos autores, sendo aplicada tanto no futebol como no beisebol, voleibol, futebol americano, entre outros esportes (Courneya e Carron, 1992; Pollard, 2002).

Em estudos anteriores, como em Pollard (1986), observou-se que há um grande predomínio das equipes mandantes, o que pode ser comprovado através de um aproveitamento de cerca de 60% a 67% de pontos obtidos em casa.

Acrescenta-se ainda o aspecto de familiaridade com este campo e estádio, sendo afirmado por Courneya e Carron (1992), Pollard (2002) que há uma maior consciência do atleta da "casa", o que lhe permite uma orientação mais eficaz nas ações decorrentes e exigidas na partida.

Segundo Pollard (2006), a grande familiaridade de um time com o seu estádio faz com que o mesmo leve vantagem, quando enfrenta um adversário em seus domínios, logo, espera-se um desempenho melhor em casa do que fora.

Com a mudança do modelo do campeonato Brasileiro para um torneio disputado através de pontos corridos com formato *Round Robin* (todos jogando contra todos), para maiores informações sobre o desenho do campeonato brasileiro, ver Sant'Anna et al (2007) e conseqüente aproveitamento da tabela final do campeonato na classificação para outras competições, assim como o rebaixamento dos piores times; despertou-se um interesse por compreender melhor o *home advantage* na competição e as chances de cada time nos seguintes objetivos: salvar-se do rebaixamento, estar entre os 4 primeiros colocados (copa libertadores da América) e ser campeão.

Em campeonatos do tipo *round robin*, o resultado de todos os jogos é de suma importância, pois é necessário acumular o maior número de pontos possíveis, pois o time que consegue ser mais eficiente nesse quesito, será o campeão. Logo a equipe necessita ser tão bom fora, quanto dentro de casa.

Enquanto isso, nos campeonatos mistos, o resultado da primeira fase só serve como balizador para o confronto direto entre as equipes. Pois a equipe que tiver melhor desempenho terá algumas vantagens como decidir em casa e jogar por resultados iguais. Com isso uma vitória em casa por uma boa diferença de gols tem uma importância muito grande.

Diante desses fatos, tentou-se modelar o campeonato Brasileiro de 2007 objetivando acompanhar o desempenho de cada time ao longo do campeonato, considerando a importância do mando de campo (*Home Advantage*) e a força de cada time nos resultados dos jogos e calcular chances de rebaixamento e classificação para a Copa Libertadores da América para

os clubes.

O modelo proposto é um modelo Logito Ordinal, também utilizado por Bayo H. et al (2007) na modelagem com campeonato Inglês (EPL) nos anos de 2002 e 2003.

Bayo H. et al (2007) trabalha com um modelo bastante simples, porém, eficaz, na previsão do EPL (english premier league). O modelo utilizar a vantagem de jogar em casa e as forças individuais de cada time como variáveis preditivas.

Pollard (2002) analisa a redução da vantagem de uma equipe em jogar em casa quando esta se muda de uma cidade para outra, algo muito comum em esportes onde os times são franquias particulares ou em casos em que o time perde o mando de campo, por algum motivo.

No caso do Brasileirão em particular, temos os casos onde os clássicos entre os times de uma mesma cidade são “obrigados” a ser realizado num estádio, que pode ser considerado campo neutro, como, por exemplo, o Maracanã no Rio de Janeiro, onde jogam Flamengo, Vasco, Botafogo e Fluminense; assim como o Morumbi em São Paulo, onde são jogados os jogos de São Paulo, Corinthians e Palmeiras e por fim o Mineirão onde são realizados os clássicos entre Cruzeiro e Atlético-MG.

2. OBJETIVO

O objetivo principal do presente trabalho é avaliar a capacidade preditiva do modelo *logito* ordinal estimado ao final de cada rodada, as probabilidades de um time estar classificado para a copa libertadores da América (encerrar o campeonato entre as 4 maiores pontuações), assim como ser rebaixado (4 piores pontuações). Isso será feito com a utilização de um modelo Logito Ordinal e com o auxílio de simulações baseadas nos resultados desse modelo.

3. DADOS

Para o presente trabalho foram utilizados os dados do campeonato Brasileiro do ano de 2007, onde havia 20 times e um total de 380 jogos. A pontuação a cada jogo é dada por: vitória (3 pontos), empate (1 ponto) e derrota (0 ponto).

Ao final da competição, o time com maior pontuação será o campeão, os quatro times melhor colocados garantem vaga para a Copa Libertadores da América do ano seguinte e os quatro times com menor pontuação são automaticamente rebaixados para a série B.

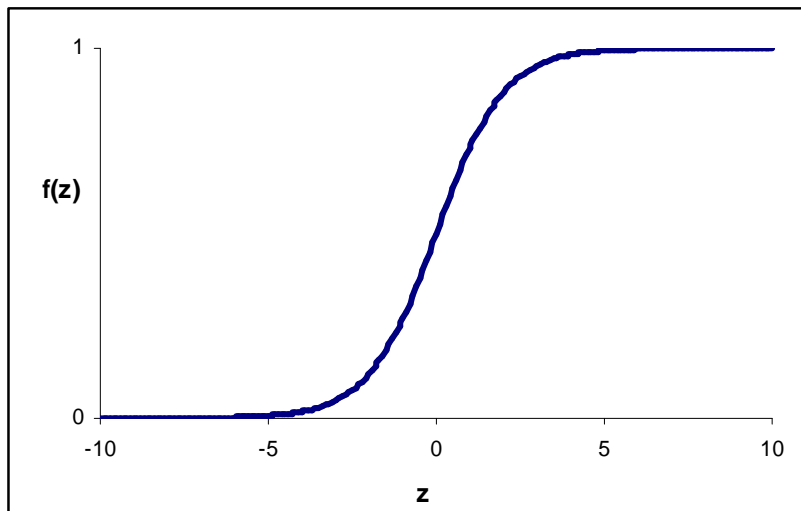
Como mandante de campo foram considerados times identificados desta forma pela tabela da CBF.

4. TRANSFORMAÇÃO LOGÍSTICA

Apesar de parecer complicado de visualização, o modelo é relativamente simples. A idéia é descrever a relação entre diversas variáveis independentes e uma variável dependente dicotômica (0,1); esse processo é diferente do modelo de regressão linear, que possui uma variável contínua como resposta. Esse processo é feito com base na regressão logística, que é baseada na função sigmóide $f(z)$, dada por:

$$f(z) = \frac{e^z}{1 + e^z}$$

Pode-se observar através do gráfico abaixo que a função logística varia entre 0 e 1 e é esse o principal motivo pelo qual utiliza-se essa função para descrever uma probabilidade dicotômica (ocorrência ou não ocorrência de um determinado fato).



Através dessa transformação, faz-se com que a variável resposta, antes dicotômica, assumam valores de uma variável métrica, portanto, o procedimento que calcula os coeficientes logísticos compara a probabilidade de um evento ocorrer com a probabilidade dele não ocorrer. A partir dessa função logística $f(z)$, escreve-se um modelo descrevendo z como uma soma linear de variáveis independentes, ou seja:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Desta forma, a função logística passa a ser trabalhada como uma soma de variáveis independentes X e seus parâmetros β . Para transformar essa função em um modelo logístico, deve-se substituir $f(z)$ por $\pi(x)$, da seguinte maneira:

$$E(Y/x) = \pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}$$

Onde $E(Y/x)$ é a esperança de Y dado x .

Como a relação em $\pi(x)$ é não linear, faz-se necessário a transformação logito, para que exista uma relação linear entre as variáveis explicativas e a variável transformada da variável resposta, assumindo $g = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$

E resolvendo a equação para $g(x)$, tem-se:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right]$$

Essa função logito $g(x)$ é linear nos seus parâmetros.

Após estimar os valores para os coeficientes β , basta substituí-los nas funções descritas anteriormente para encontrar o valor estimado de $\pi(x)$, que será um valor entre 0 e 1 e denotará a classificação do indivíduo na variável resposta dicotômica

5. MODELO LOGITO ORDINAL

Um modelo amplamente difundido e utilizado na previsão de resultados é o modelo Logito Binário, como proposto por Lawal (2002). Esse modelo é descrito como:

$$\ln \left(\frac{P_{ij}}{1 - P_{ij}} \right) = \alpha_i + \sum_{i=1}^{20} \phi_i X_i$$

Onde:

P_{ij} é a probabilidade do time “i” vencer o time “j” jogando em casa

α_i é o Home Advantage

$$X_i = \begin{cases} 1 & (\text{time está em casa}) \\ -1 & (\text{time é visitante}) \\ 0 & (\text{se não está jogando}) \end{cases}$$

Esse modelo possui uma restrição em sua variável resposta, que precisa ser Binária.

Com o objetivo de utilizar melhor a informação do resultado dos jogos, optou-se por trabalhar com um modelo *Logito* Ordinal proposto por Lawal (2002), onde é feito um

desmembramento dos dados em dois modelos, da seguinte forma – já adaptado para 20 times:

$$\ln \left[\frac{P(\text{Vitória}/X_1 - X_{19}, \delta)}{P(\text{Empate ou Derrota} / X_1 - X_{19}, \delta)} \right] = \alpha_1 + \sum_{i=1}^{20} \phi_i X_i$$

$$\ln \left[\frac{P(\text{Vitória ou Empate} / X_1 - X_{19}, \delta)}{P(\text{Derrota} / X_1 - X_{19}, \delta)} \right] = \alpha_2 + \sum_{i=1}^{20} \phi'_i X_i$$

Esse modelo está sujeito a:

$$\phi'_i = \phi_i \quad \forall i = 1, 2, \dots, 20$$

Ou seja, a força do time “i” jogando contra o time “j” precisa ser a mesma, independente se a variável resposta é apenas a vitória ou uma vitória ou empate em casa. Somente se esta condição for satisfeita é que se pode fazer uso desse modelo.

Considerando a hipótese satisfeita, o modelo final fica:

$$\ln \left[\frac{P(\text{Vitória}/X_1 - X_{19}, \delta)}{P(\text{Empate ou Derrota} / X_1 - X_{19}, \delta)} \right] = \alpha_1 + \sum_{i=1}^{20} \phi_i X_i$$

$$\ln \left[\frac{P(\text{Vitória ou Empate} / X_1 - X_{19}, \delta)}{P(\text{Derrota} / X_1 - X_{19}, \delta)} \right] = \alpha_2 + \sum_{i=1}^{20} \phi_i X_i$$

Através desse modelo, é possível calcular as probabilidades de vitória, empate ou derrota para um time jogando em casa, isso é feito com base na fórmula:

$$P(R \leq j) = \left[\exp(\alpha_j) + \sum_{i=1}^{20} \phi'_i X_i \right] / \left[\exp(1 + \alpha_j) + \sum_{i=1}^{20} \phi'_i X_i \right] \quad (\text{Equação 1})$$

Com base nesse modelo, é possível estimar a força do time em um confronto direto contra os demais times, criando, desta forma, um *ranking* dos times, isso é feito com base na fórmula abaixo:

$$P_i = \left[\exp(\phi'_i) \right] / \left[\sum_{i=1}^{20} \exp(\phi'_i) \right] \quad (\text{Equação 2})$$

Para tornar mais simples a compreensão do modelo, segue abaixo análise detalhada feita com base nos coeficientes da rodada 34.

Coeficientes da Rodada 34	
$\alpha_{\text{Vitória}}$	0.01
$\alpha_{\text{Vitória+Empate}}$	1.15
América	-1.77
Atlético_MG	-0.05
Atlético_PR	0.10
Botafogo	0.20
Corinthians	-0.18
Cruzeiro	0.34
Figueirense	-0.04
Flamengo	0.40
Fluminense	0.26

Goiás	-0.38
Grêmio	0.31
Internacional	0.00
Juventude	-0.64
Náutico	-0.20
Palmeiras	0.45
Paraná	-0.37
Santos	0.44
São_Paulo	1.26
Sport	-0.20
Vasco	0.06

Com base na equação 1, é possível, por exemplo, estimar a probabilidade do América, jogando em casa, vencer o Botafogo, da seguinte forma:

$$P_{ij} = \exp(0.01 + (1) * (-1.77) + (-1) * 0.20) / (1 + \exp(0.01 + (1) * (-1.77) + (-1) * 0.20)) = 0.1236$$

E a probabilidade do América, jogando em casa, vencer ou empatar com o Botafogo, é dada por:

$$P_{ij} = \exp(1.15 + (1) * (-1.77) + (-1) * 0.20) / (1 + \exp(1.15 + (1) * (-1.77) + (-1) * 0.20)) = 0.3061$$

Ou seja, a probabilidade do resultado ser empate é dada por: 0.3061 – 0.1236 = 0.1825; restando 0.6939 (1-0.3061), que é a probabilidade do América perder, jogando em casa, para o Botafogo.

Aplicando a fórmula (2), é possível criar um ranking dos times, ficando o resultado como segue abaixo:

	Ranking
América	0.0074
Atlético_MG	0.0414
Atlético_PR	0.0480
Botafogo	0.0531
Corinthians	0.0361
Cruzeiro	0.0610
Figueirense	0.0417
Flamengo	0.0644
Fluminense	0.0563
Goiás	0.0298
Grêmio	0.0589
Internacional	0.0432
Juventude	0.0229
Náutico	0.0356
Palmeiras	0.0683
Paraná	0.0299
Santos	0.0675
São_Paulo	0.1528
Sport	0.0355
Vasco	0.0461

Esse modelo, apesar de simples, por não considerar outros fatores externos, além da força do time e da vantagem de jogar em casa, foi utilizado com sucesso por Bayo H. Lawal e Richard A. Sundheim (2007) na modelagem com campeonato Inglês (EPL) nos anos de 2002 e 2003.

6. SIMULAÇÕES

Já faz alguns anos que o recurso a técnicas de simulação a utilização de números aleatórios é utilizado como alternativa aos procedimentos mais convencionalmente utilizados em estatística para testar hipóteses, Waton e Blackstone (1989). Com o aumento da velocidade de processamento observado nos computadores hoje em dia, este procedimento ganha cada vez mais importância, uma vez que a lentidão, que era o principal problema observado no passado, se reduz ao mínimo.

A simulação trabalhada no presente trabalho será calculada com base nas probabilidades estimadas para os jogos não realizados do campeonato derivadas dos modelos estimados com os jogos realizados até a rodada, isto é, a cada rodada será estimado um modelo e calculada a probabilidade de cada resultado possível para todos os jogos não realizados do campeonato.

A cada jogo não realizado, será atribuído um número aleatório entre 0 e 1 e, com base na probabilidade derivada do modelo e desse número aleatório, será atribuído o resultado ao jogo. Com base nos resultados estimados e nos resultados dos jogos já realizados, serão calculados os pontos estimados no final do campeonato e, através dessa pontuação estimada, analisados os times que figuram entre os resultados de interesse (times rebaixados e classificados para a Copa Libertadores da América) considerando todas as simulações realizadas.

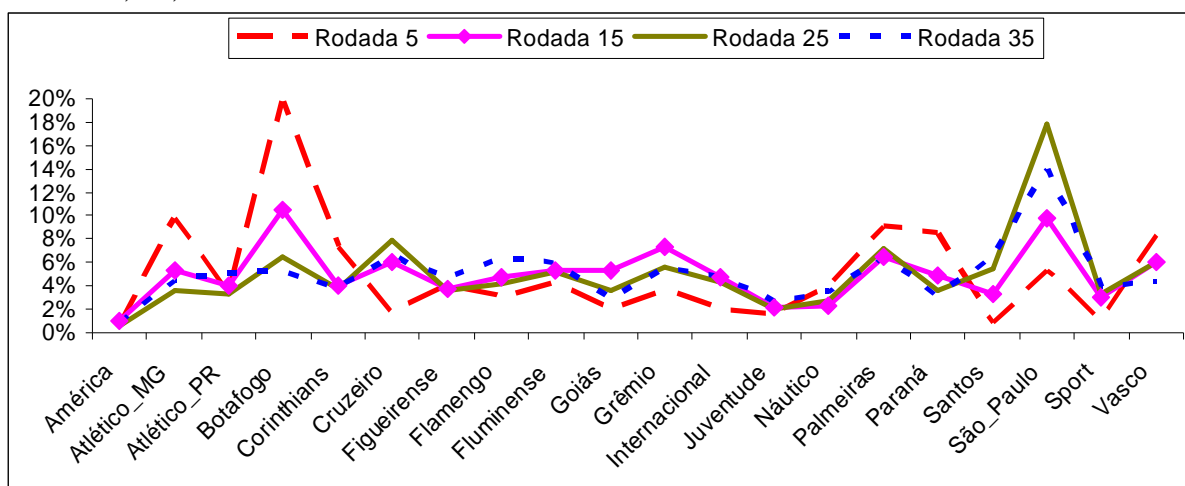
7. ANÁLISE DOS RESULTADOS

O *Home Advantage* observado no campeonato Brasileiro de 2007 até a rodada 35 (total de 350 jogos realizados) foi de 73,43%, ou seja, em 257 de todas as partidas disputadas,

o time da casa pontuou (vitória ou empate). Ao analisar apenas a vitória do time da casa, esse percentual passa para 49,43%, ou seja, em 173 dos 350 jogos, o time da casa sai vencedor.

Com relação aos modelos, no total, foram feitos 31 modelos, considerando os dados até uma determinada rodada k; esses modelos foram estimados a partir da rodada 5 até a rodada 35. Em outras palavras, existe um modelo com os dados até a rodada 5, outro modelo considerando os dados das rodadas 1 até a 6, ..., até um modelo considerando todos os dados desde a rodada 1 até a rodada 35. Em todos os modelos a restrição $\theta_t^k = \theta_t \forall t = 1, 2, \dots, 20$ foi satisfeita, tornando possível a utilização e análise dos mesmos.

A partir dos coeficientes θ_t de cada modelo, foi possível calcular a força de cada time – conforme equação 2. Encontra-se abaixo um gráfico ilustrando as forças dos times nas rodadas 5, 15, 25 e 35.



Analisando o gráfico, percebe-se a ascensão do São Paulo ao longo das rodadas 5, 15, 25 e 35, porém, ocorre uma leve queda no modelo ilustrado para a rodada 35 pelo fato do time já haver sido consagrado campeão por antecipação e conseqüente perda de interesse pelo campeonato.

Uma perda de rendimento ao longo do campeonato para os clubes: Botafogo, Atlético Mineiro, Corinthians e Paraná; sendo que os dois últimos permanecem brigando contra o rebaixamento.

Outra abordagem visava à previsão dos times rebaixados e classificados para a copa libertadores da América; para tal, foram estimados modelos utilizando os dados até a rodada k, e com base nesse modelo simulava-se 10.000 resultados das rodadas subseqüentes e calculava-se o percentual de vezes que time figurava entre os quatro primeiros colocados.

A tabela abaixo ilustra os percentuais que cada time figurava entre os 4 times com maior pontuação encontrados após 10.000 simulações de resultados considerando as probabilidades de resultado derivadas dos modelos de cada rodada (rodada 5, 11, 17, 23, 29 e 35). As rodadas foram selecionadas de maneira equidistante (6 rodadas de diferença entre elas). A tabela encontra-se ordenada pela pontuação final real do campeonato.

Percentual de vezes que o time está entre os 4 de maior pontuação ao final das 10.000 simulações com base nos resultados de cada rodada							
	Pontuação Final	Rodada 5	Rodada 11	Rodada 17	Rodada 23	Rodada 29	Rodada 35
São Paulo	77	7.2%	80.0%(*)	99.1%(*)	100.0%(*)	100.0%(*)	100.0%(*)
Santos	62	0.0%	0.2%	5.7%	8.2%	49.3%	84.9%(*)
Flamengo	61	0.1%	40.0%	17.4%	8.5%	18.5%	46.8%(*)

<i>Fluminense</i>	61	1.2%	1.5%	3.0%	1.7%	9.1%	27.0%
Cruzeiro	60	0.0%	6.0%	30.7%	92.0%(*)	92.1%(*)	88.9%(*)
Grêmio	58	0.6%	17.7%	24.5%	21.6%	62.3%(*)	13.7%
Palmeiras	58	61.6%(*)	73.4%(*)	35.0%(*)	45.6%(*)	56.8%(*)	38.3%
Atlético MG	55	77.9%(*)	1.1%	12.9%	0.6%	0.0%	0.0%
Botafogo	55	99.8%(*)	98.0%(*)	96.2%(*)	71.4%(*)	9.6%	0.2%
Atlético PR	54	0.3%	0.7%	0.1%	0.0%	0.2%	0.3%
Internacional	54	0.0%	0.2%	1.0%	1.1%	0.3%	0.0%
Vasco	54	53.0%	4.5%	35.4%(*)	43.6%	0.7%	0.0%
Figueirense	53	1.7%	12.4%	5.0%	0.0%	0.4%	0.0%
Sport	51	0.0%	0.0%	4.5%	1.3%	0.8%	0.0%
Náutico	49	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%
Goiás	45	0.0%	47.8%(*)	21.2%	3.8%	0.0%	0.0%
<i>Corinthians</i>	44	39.9%	14.2%	5.2%	0.4%	0.0%	0.0%
<i>Juventude</i>	41	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>Paraná</i>	41	56.3%(*)	2.4%	3.3%	0.1%	0.0%	0.0%
<i>América</i>	17	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Os quatro primeiros times (em negrito e itálico), foram os times que se classificaram para a copa libertadores da América e os quatro últimos times (negrito e itálico) foram os times rebaixados para a série B do campeonato. Assinalados com (*) encontram-se os 4 times com maiores chances de serem classificados segundo o modelo que considera os dados até a rodada.

O Botafogo começou o campeonato de maneira primorosa, o que fez com que na rodada 5 o time fosse o que apresentava maior chance de estar entre os times classificados para a Libertadores da América – em 99.8% das 10.000 simulações, o time aparecia entre os 4 de maior pontuação estimada para o final do campeonato. Ao longo do campeonato o time veio sofrendo derrotas e apresentando um desempenho inferior ao apresentado no início do campeonato, por isso, as chances do time terminar o campeonato entre as quatro maiores pontuações foram reduzindo ao longo das rodadas.

O inverso ocorreu com o São Paulo e com o Santos, que foram aumentando suas chances de estar entre 4 times de maior pontuação no campeonato ao longo do campeonato.

Considerando o modelo estimado com base nos jogos realizados até a rodada 35, o Fluminense tinha menos chance que o Cruzeiro e o Palmeiras, porém, conseguiu emplacar 2 vitórias muito importantes nas últimas duas rodadas, ao contrário de Cruzeiro e Palmeiras, que tropeçaram nas últimas rodadas, conseguindo somente uma vitória, perdendo a vaga para o Fluminense.

Assim como na análise para os times que poderiam se classificar para a copa libertadores da América, foram calculados o percentual de vezes que cada time aparecia entre os quatro de menor pontuação (que seriam rebaixados para a série B do campeonato). A tabela abaixo ilustra o percentual de vezes que o time fica entre os quatro times de menor pontuação no final do campeonato.

Percentual de vezes que o time está entre os 4 de maior pontuação ao final das 10.000 simulações com base nos resultados de cada rodada							
	Pontuação Final	Rodada 5	Rodada 11	Rodada 17	Rodada 23	Rodada 29	Rodada 35
<i>São Paulo</i>	77	77	0.0%	0.0%	0.0%	0.0%	0.0%
<i>Santos</i>	62	62	97.9%(*)	20.5%	5.8%	0.2%	0.0%
<i>Flamengo</i>	61	61	1.6%	0.1%	1.3%	0.6%	0.0%
<i>Fluminense</i>	61	61	0.1%	5.0%	8.8%	2.2%	0.1%

Cruzeiro	60	60	36.2%	1.3%	0.3%	0.0%	0.0%
Grêmio	58	58	0.3%	0.2%	0.8%	0.0%	0.0%
Palmeiras	58	58	0.0%	0.0%	0.4%	0.0%	0.0%
Atlético MG	55	55	0.0%	7.3%	2.0%	4.4%	21.8%
Botafogo	55	55	0.0%	0.0%	0.0%	0.0%	0.0%
Atlético PR	54	54	0.7%	10.2%	51.3%(*)	46.0%(*)	4.6%
Internacional	54	54	13.5%	18.0%	17.3%	3.0%	4.9%
Vasco	54	54	0.0%	1.8%	0.3%	0.0%	2.9%
Figueirense	53	53	0.1%	0.3%	5.8%	23.3%	6.2%
Sport	51	51	83.0%(*)	43.3%(*)	7.0%	3.7%	1.7%
Náutico	49	49	0.4%	96.5%(*)	84.6%(*)	93.8%(*)	18.5%
Goiás	45	45	17.4%	0.0%	1.0%	0.9%	20.8%
Corinthians	44	44	0.0%	0.2%	6.5%	7.7%	55.7%(*)
Juventude	41	41	49.1%(*)	92.2%(*)	98.6%(*)	98.4%(*)	97.5%(*)
Paraná	41	41	0.0%	3.3%	8.3%	15.9%	65.4%(*)
América	17	17	99.9%(*)	100.0%(*)	100.0%(*)	100.0%(*)	100.0%(*)

Fica evidente a melhora no desempenho do time do Santos ao longo do campeonato, terminando como segundo melhor pontuação; se o Santos mantivesse o mesmo desempenho que estava tendo no início do campeonato, o time tinha grandes chances de ser rebaixado. Assinalados com (*) encontram-se os 4 times com maiores chances de serem rebaixados segundo o modelo que considera os dados até a rodada.

Apesar do Corinthians apresentar uma melhor chance de escapar do rebaixamento que o Goiás (considerando sua história até a rodada 35), ele acaba sendo rebaixado por tropeçar com 2 empates e uma derrota nas últimas rodadas do campeonato, enquanto o Goiás consegue uma vitória na última rodada, escapando do rebaixamento.

Um dos problemas observados no modelo é a demora para que sejam captadas as mudanças de comportamento dos times. No presente modelo, todos os jogos contribuem com a mesma importância no modelo, ou seja, não existe uma maior importância para os jogos mais recentes do time; toda a história do time no campeonato até a rodada foi utilizada e cada jogo figurava com a mesma importância.

Essa demora em captar a mudança de comportamento do time ao longo do campeonato pode ser observada com o Flamengo, que, ao longo do campeonato não vinha apresentando bons resultados, porém, conseguiu fazer uma arrancada nas últimas 15 rodadas que o modelo não teve tempo hábil de captar. Das 17 vitórias que o Flamengo teve no campeonato, 8 delas ocorreram entre as rodadas 23 e 35, porém, como o modelo considerava toda a história do Flamengo no campeonato, acaba fazendo com que o Flamengo figure menos vezes entre os 4 times de maior pontuação nas simulações.

Da mesma maneira, o São Paulo apresentou um desempenho tão superior a maioria dos times que, se o time mantivesse o mesmo desempenho que vinha tendo até a rodada 17, estaria classificado para a Libertadores da América, supondo que o desempenho dos demais times não sofresse grandes alterações ao longo das rodadas seguintes. Em função disso, foram feitas novas simulações considerando que o São Paulo não tivesse mais vitórias a partir da rodada em estudo, ou seja, utilizava-se o modelo com os dados reais até a rodada N, porém, independente da probabilidade encontrada através das simulações para as rodadas seguintes, assumia-se que o São Paulo sempre perderia os jogos seguintes.

Percentual de vezes que o time está entre os 4 de maior pontuação ao final das 10.000 simulações com base nos resultados de cada rodada, porém, considerando que o São Paulo não vença nenhum jogo a partir dessa rodada					
	Pontuação Final	Rodada 17	Rodada 23	Rodada 29	Rodada 35

São Paulo	77	0.00%	0.00%	67.39%(*)	100.00%(*)
Santos	62	11.16%	21.34%	45.83%	78.78%(*)
Flamengo	61	17.65%	9.61%	14.83%	36.09%
Fluminense	61	5.17%	5.18%	15.27%	19.89%
Cruzeiro	60	45.91%(*)	97.91%(*)	97.78%(*)	80.69%(*)
Grêmio	58	39.90%	45.33%	82.26%(*)	55.93%(*)
Palmeiras	58	46.59%(*)	49.91%(*)	52.06%(*)	27.51%
Atlético MG	55	21.40%	2.50%	0.01%	0.00%
Botafogo	55	99.15%(*)	89.15%(*)	22.17%	0.66%
Atlético PR	54	1.18%	0.07%	0.32%	0.45%
Internacional	54	1.83%	3.03%	0.08%	0.00%
Vasco	54	43.64%(*)	64.50%(*)	0.44%	0.00%
Figueirense	53	12.03%	0.37%	0.19%	0.00%
Sport	51	7.05%	4.30%	1.37%	0.00%
Náutico	49	0.05%	0.00%	0.00%	0.00%
Goiás	45	30.32%	4.85%	0.00%	0.00%
Corinthians	44	9.53%	1.80%	0.00%	0.00%
Juventude	41	0.00%	0.00%	0.00%	0.00%
Paraná	41	7.44%	0.15%	0.00%	0.00%
América	17	0.00%	0.00%	0.00%	0.00%

Vale ressaltar que, mesmo supondo que o São Paulo não tenha outra vitória após a rodada 29, ele continua aparecendo entre os 4 times com maiores chances de serem classificados para a copa Libertadores da América.

8. CONCLUSÕES E TRABALHOS FUTUROS

O modelo em questão apresentou boas previsões para o resultado do campeonato, tanto para os times que, se não modificassem seu desempenho no campeonato, tinham as maiores chances de serem rebaixados, como para os times que, se mantivessem o mesmo desempenho observado até uma determinada rodada, tinham as maiores chances de serem classificados para a Libertadores da América. Contudo, uma melhora no desempenho do time nas últimas rodadas demora a ser percebida pelo modelo, pois o mesmo trata todos os jogos realizados pelo time com a mesma importância. Uma alternativa para tratar essa deficiência seria acrescentar mais variáveis no modelo, além do *home advantage* e da força do time, utilizar, por exemplo, o saldo de gols do time nas últimas n rodadas ou mesmo acrescentar uma variável *dummy* identificando se houve mudança de técnico recente nesse time.

9. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] **Courneya, K. S.; Carron, A. V.** (1992) The home advantage in sport competitions: a literature review. *Journal of Sport and Exercise Psychology*; 14:13-27.
- [2] **Craig, L. A.; Hall, A. R.** (1994). Trying out for the team: Do exhibitions matter? Evidence from the National Football League. *Journal of the American Statistical Association*, 89, 1091-1099
- [3] **Lawal, H. Bayo; Sundheim, A. Richard** (2007). Modeling 2002-2003 English Premier League Results. *IMA Sport* 2007, 115-124.
- [4] **Lawal, H. Bayo** (2002). Modelling the 1984-1993 American League Baseball Results as dependent categorical data. *Math. Scientist*, 27, 53-66.

- [5] **Lee, A. J.** (1997). Modeling scores in the premier league: Is Manchester United really the best?. *Chance*, 10, 15-19
- [6] **Pollard, R.** (1986) Home advantage in soccer: a retrospective analysis. *Journal of Sports Sciences*, 4, 237-248.
- [7] **Pollard, R.** (2002). Evidence of a reduced home advantage when a team moves to a new stadium. *Journal of Sports Sciences*, 20 (12), pp. 969-973.
- [8] **Pollard, R.** (2006). Worldwide regional variations in home advantage in association football. *Journal of Sports Sciences*, 24 (3), pp. 231-240.
- [9] **Sant'Anna**, Annibal Parracho; **Barboza**, Eduardo Uchoa; **Soares de Mello**, João Carlos Correa Baptista (2007). A framework to explain the Brazilian soccer championchips classifications. *IMA Sport 2007*, 180-185.
- [10] **Scarff**, Philip A.; **Shi Xin** (2008). The importance of a match in a tournament. *Computers & Operations Research* 35 (2008) 2406-2418
- [11] **Stefani, R. T. e Clarke, S. R.** (1992), Predictions and home advantage for Australian Rules football, *Journal of Applied Statistics*, 9, 251-261
- [12] **Watson, H.J. Blackstone Jr, J. H.** (1989). *Computer simulation*, 2nd ed. New York: John Wiley & Sons