

Predição de resultados no campeonato brasileiro 2012 série A.

Prediction results in the Brazilian Championship 2012 series A.

MACEDO, PAP¹; SILVA, CD²

1. Pós graduando em Futebol pela Universidade Federal de Viçosa-MG.
2. Doutorando no Programa de Pós-Graduação em Ciências do Esporte Escola de Educação Física, Fisioterapia e Terapia Ocupacional Universidade Federal de Minas Gerais.

Resumo

Introdução: O bom rendimento de equipes e jogadores no futebol é dependente de várias áreas de conhecimento tais como o treinamento desportivo, a preparação física, a nutrição, a psicologia e a análise de desempenho. A partir dessa premissa, vários estudos ligados ao desempenho de equipes de futebol vêm surgindo com a finalidade de se descobrir tendências e padrões de várias ações ocorridas no jogo de futebol, assim como possibilidade de predição de rendimento.

Objetivo: Este estudo visou compreender fatores que podem influenciar a predição de vitória no Campeonato Brasileiro de futebol de 2012 (Série A).

Metodologia: Usou-se a regressão logística em 380 jogos devidamente separados com ocorrência de vitória utilizando-se para predição as variáveis ranking de clubes CBF do ano de 2012 (diferença de pontos), dimensão de gramado (maior, menor e igual), intervalo de recuperação (dias; um jogo para outro).

Resultados: Foi obtido um valor de 60,3% de acerto dos resultados, sendo significativa na predição a variável ranking de clubes, diferença de recuperação e dimensão de gramados (campo mandante maior que visitante). Apesar disso, a classificação do poder de predição (effect size) foi baixa (Cox & Snell R Square 0.073).

Conclusão: O ranking de clubes pode ser um bom parâmetro para comissões técnicas analisarem a real força de uma equipe sobre a outra, considerando que as outras variáveis utilizadas no presente estudo sofre forte influência por contextos momentâneos e pouco planejamento com referencia ao local de jogo (mudanças repentinas do mandante por questões econômicas ou punitivas) e incertezas a qualidade do processo recuperativo.

Palavras-chave: Predição. Futebol. Regressão logística.

ABSTRACT

pedromg18fut@hotmail.com

Introduction: The best team performance and players in soccer is dependent on various areas of knowledge such as sports training, nutrition, psychology and performance analysis. From this premise, several studies related to the performance of soccer teams are emerging in order to uncover trends and patterns of various actions taken in the match, as well as the possibility of predicting the yield.

Objective: This study aimed to understand factors that may influence the prediction of victory in the Brazilian League Soccer 2012 (Series A).

Methodology: It used logistic regression in 380 games properly separated with occurrence of victory using variables to predict the ranking of CBF clubs of the year 2012 (points difference), size of pitch (larger, smaller and equal), recovery interval (day, game to game).

Results: Value of 60.3% of correct results was obtained, being significant in predicting the ranking variable clubs, difference in recovery and size of pitch (principal field greater than visitor). Nevertheless, the classification of predictive power (effect size) was low (Cox & Snell R Square 0.073).

Conclusion: The ranking of clubs can be a good parameter to technical committees examine the real strength of a team over the other, whereas the other variables used in this study is strongly influenced by momentary Little planning contexts and with reference to the local game (sudden changes of the principal by economic or punitive) issues and uncertainties quality recuperative process.

Keywords: Prediction, soccer, logistic regression.

Introdução

O bom rendimento de equipes e jogadores no futebol é dependente de várias áreas de conhecimento tais como o treinamento desportivo, a preparação física, a nutrição, a psicologia e a análise de desempenho. A partir dessa premissa, vários estudos ligados ao desempenho de equipes de futebol vêm surgindo com a finalidade de se descobrir tendências e padrões de várias ações ocorridas no jogo de futebol, assim como possibilidade de predição de rendimento.

Estudos como o de Souza et al.^[1] apontam dados extremamente relevantes sobre a ocorrência de gol no campeonato brasileiro de futebol profissional da primeira divisão, indicando onde e quando são as maiores incidências de gols marcados nesse torneio, revelando um maior

aparecimento de gols nos quinze minutos finais da partida, além da origem ser em sua maioria gols feitos com a bola em movimento e dentro da grande área. Os estudos de Silva^[2], Farias^[3] e Pollard et al.^[4] mostram que fatores como viagens, acomodações, características dos locais de jogos, influência da torcida, pressão sobre os árbitros, dimensões do campo de jogo, podem alterar relativamente questões como a técnica, a velocidade, a força, a resistência aeróbia e anaeróbia dos atletas, reduzindo o desempenho do atleta de futebol. Parece por tanto, que a condição física do atleta de futebol pode estar associada a essa ocorrência de gol nos minutos finais de uma partida de futebol profissional.

A vantagem em casa (VC) de uma equipe de futebol também deve ser levada em consideração, considerando que estudos como o de

Silva e Silva^[5] mostram que essa variável é um fator de alta relevância no resultado de um jogo de futebol no Campeonato Brasileiro da Série A, comparando-o com outras ligas de futebol mundo afora. Dentro da perspectiva do mando de campo, surgem outras questões que podem interferir no resultado do jogo como diferença de dimensão do campo entre mandante e visitante, assim como as diferenças entre dias de recuperação e ranking, que pode demonstrar uma forte diferença de capacidade competitiva que se traduz em vitórias.

Portanto, o objetivo desse estudo foi verificar a capacidade explicativa desses fatores no resultado final do jogo no campeonato brasileiro série A de futebol no ano de 2012.

Metodologia

Amostra

Foram analisados 380 jogos do campeonato nacional da primeira divisão brasileira de futebol da temporada do ano de 2012.

O critério de atribuição de pontos para no campeonato foi o mesmo utilizando por outras competições, ou seja, três pontos em caso de vitória, um ponto para empate e zero ponto para derrota. Os times se enfrentavam duas vezes, em jogos de turno e retorno, com mando de campo sendo alternado. A tabela do campeonato analisado foi consultada via internet, no site www.soccerway.com (<http://br.soccerway.com/national/brazil/serie-a/2012/regular-season/r17449/>) o qual tem sido amplamente utilizado em estudos prévios dessa natureza.

Nessa tabela foram observados os dias de recuperação entre uma rodada e outra se apoiando também em jogos de competições em disputa paralela (Copa Sul Americana 2012, Copa do Brasil 2012, Taça Libertadores 2012 e Recopa Sul-Americana 2012). Também foi obtido as referências de tamanho de campo de jogo. No site da CBF – www.cbf.com.br, foi obtido o ranking dos clubes participantes da competição analisada correspondente ao ano 2012. Esse ranking é lançado sempre que termina cada campeonato nacional, e conta a pontuação dos clubes obtidos através do resultado alcançado nas competições durante a temporada.

Para verificar a possibilidade de predição, três variáveis foram tratadas como: i) recuperação em dias, ii) diferença do gramado em maior, menor e igual, iii) diferença de ranking em pontos. Para o desfecho preditivo, o resultado do jogo foi tratado como vitória sim e não. O empate foi considerado como desfecho negativo, ou seja, ausência de vitória para atendimento da regressão logística binária.

Análise estatística

Os dados contínuos foram apresentados como média \pm desvio padrão ou mediana, e os categóricos como número (frequência de resposta). Foi realizada análise de regressão logística binária por *backward stepwise* utilizando o pacote estatístico Statistical Package for the Social Sciences (SPSS® 15 for Windows, Chicago, IL, EUA). Considerou-se o valor de $P < 0,05$ para

nível de significância em todas as variáveis.

Resultados

Os resultados nos mostram algumas considerações importantes para que um time vença o outro dentro daquilo que foi objetivado nesse estudo. Com relação a dimensão de gramado, o modelo proposto de predição de resultado não pode levar essa variável em consideração, pois não se encontrou, a partir do tratamento estatístico feito pela regressão

logística, diferenças significativas entre uma equipe e outra. A única modalidade de dimensão de gramado que alcançou significação estatística foi a quando o mandante tem um campo de jogo maior ao campo de origem do visitante (DIF_GramadoEscala(2); $p=0.034$). Outra variável com predição significativa foi dias de recuperação ($p=0.016$). Apesar disso, a classificação do poder de predição (effect size) foi baixa (Cox & Snell R Square 0.073). A tabela 1 resume a análise estatística observada após a regressão logística.

Tabela 1. Os resultados da análise de regressão logística

Preditores	Coeficiente	P value	Exp(B)	95% C.I.for EXP(B)	
				Lower	Upper
Recuperação (Dias)	,171	,016	1,187	1,033	1,363
Gramado (Menor)		,095			
Gramado (Igual)	-,156	,558	,855	,508	1,441
Gramado (Maior)	,576	,035	1,779	1,043	3,036
Ranking	,000	,000	1,000	1,000	1,000
Constant	-,106	,331	,900		

Discussão

De acordo com os resultados obtidos, o modelo até explica muita coisa (60.3%) do resultado de vitória (são elas: diferença de recuperação em dias, diferença do gramado em maior, menor e igual, e diferença de ranking em pontos), porém, é claro que fatores como mudança de técnicos, venda dos principais jogadores, perda de mando de campo podem interferir no resultado de um jogo, além da questão da sorte, que segundo Anderson e Sally^[7], onde quantificaram em 50% a influência do fator sorte no resultado final de um jogo

de futebol entre as principais ligas do mundo e jogos da copa do mundo.

Como tivemos um resultado bastante expressivo com o uso do ranking da CBF, em relação às outras variáveis comparadas, podemos levar esse ranking como um forte dado de comparação entre uma equipe e outra, devido à porcentagem de acerto (60,3%) dos resultados dos jogos. No entanto, a classificação do poder de predição (effect size) foi baixa (Cox & Snell R Square 0.073).

Por isso, temos que levar em conta o fator sorte num resultado de um jogo, pois como já foi citado nesse estudo, ela contribui com 50%, uma

porcentagem relativamente alta em se tratando de um esporte como o futebol, onde esse fator pelo seu alto índice condiciona a resolver metade dos resultados dos jogos.

As variáveis que foram analisadas junto com o ranking (dias de recuperação e diferença de gramado), apresentadas na Tabela 1, nos revelam que são fatores que interferem muito pouco para a obtenção da vitória de uma equipe sobre a outra no campeonato brasileiro série A 2012, sendo questões comuns a grande maioria dos times que disputam o certame. Limitações do presente estudo foram à falta de controle sobre a mudança de campo de jogo do mandante por motivos punitivos ou por interesses comerciais. Assim, juntamente com disputas simultaneas, podem fazer com que os resultados tenham um poder de generalização diminuídos.

Já o ranking da CBF nos leva a acreditar que, quanto melhor um clube está colocado, melhor desempenho ele terá sobre os outros que estão abaixo no mesmo, devido a alta taxa de acerto de resultados durante o certame analisado, além de dar a ideia de que as equipes mais bem posicionadas no ranking serem mais fortes que as de menor qualificadas no mesmo. O ranking pode ser um bom parâmetro para a comissão técnica analisar se a equipe adversária realmente possui força suficiente para vencer sua equipe ou não, podendo também fazer uma correlação com os fatores que por nós encontrados (dias de recuperação, ranking CBF e dimensão de gramado) podem realmente fazer a diferença no

resultado final de uma partida do campeonato brasileiro.

Outros modelos preditivos ou descritivos são utilizados no futebol. Por exemplo, o estudo de Anderson e Sally^[7] nos mostra o quão é importante a marcação de um gol ou mais, para a obtenção de vitória num jogo de futebol. Neste estudo de Anderson e Sally^[7], foi observado que se um time da Premier League fizer dois gols, a porcentagem de vitória de uma equipe vencer a outra chega em 50%, mostrando que detalhes competitivos e contextos do jogo podem ser determinantes, independentes de fatores calculados a priori.

Mais um modelo, usando os dados do intervalo de dias entre uma partida e outra, independente da competição em disputa, também não se encontrou nesse quesito, diferenças significativas entre uma equipe e outra. O estudo de Vigneet al.^[8] mostrou, usando dados de distância percorrida por atletas em diferentes posições de jogo, que seu desempenho não sofreu quedas acentuadas de rendimento, mesmo havendo em média, de duas a três partidas por semana sendo disputadas no campeonato em que foi feito esse estudo. Essa informação baliza nosso modelo de predição, confirmando que a vitória no campeonato brasileiro Série A não é interferida pelo intervalo de dias entre um jogo e o outro. Em virtude da alta taxa de lesões, campeonatos simultaneos, o que consequentemente leva a rodízios de atletas, podem fazer falhar a expectativa teórica sobre a recuperação e sua influência sobre desempenho.

O uso da técnica de regressão logística vem começando a ganhar espaço na análise de algumas variáveis do jogo de futebol e também na predição de resultados dos jogos em algumas competições. Rosa et al.^[9] usando a regressão logística para prever os finalistas da Copa do Mundo nas competições realizadas entre 1994 e 2010, encontrou um acerto de seis seleções finalistas durante a fase eliminatória da competição.

Outro estudo (Pollard et al.^[10]), usando a regressão logística como estatística, preditiva, com um caráter voltado para a análise de desempenho, correlacionou a eficiência e a distância de onde foram feitas as finalizações a gols usando dados entre as Copas do Mundo de 1986 a 2002.

Algumas questões do estudo podem ser colocadas como limitações do estudo, além da mudança do local de jogo como mencionado, como por exemplo o número de torcedores presentes, densidade de estádio ou mesmo a ausência de torcedores no estádio; o próprio ranking e seus critérios. Desta forma, futuros estudos poderão controlar e incluir outras variáveis de interesse. No presente estudo, tal abordagem foi considerada em função das variáveis citadas terem forte apelo de sensu comum por influenciarem resultados no futebol. Além disso, retornar resultados práticos e não meramente modelos matemáticos,

Conclui-se através desse estudo, que o ranking de clubes pode ser uma variável interessante no que diz respeito a predição de resultados de partidas de futebol do campeonato brasileiro.

Parece-nos, portanto que fatores como intervalo de dias de recuperação entre uma partida e outra e dimensão do campo de jogo serem fatores comuns a todas as equipes que disputam o campeonato brasileiro de futebol serie A 2012, não influenciando diretamente no resultado final de uma partida. A partir deste estudo, fica a sugestão de outras pesquisas sobre o assunto, para que se possa entender melhor os fatores que condicionam a vitória a priori de um time de futebol.

Referencias

1. Souza ELND, Farah BQF, Dias RMR. **Tempo De Incidência Dos Gols No Campeonato Brasileiro De Futebol 2008**. Rev. Bras. Ciênc. Esporte, 2012,34(2):421-431. [Internet] 2012 [acesso em 14 abr. 2013]. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-32892012000200012&lng=pt&nrm=iso&tlng=en
2. Silva CD. **Fadiga: Evidências nas Ocorrências de Gols no Futebol Internacional de Elite**. Rev Efdeportes -Buenos Aires, n.97 2006 jun. [Internet] 2006 [Acesso em 22 fev. 2013] Disponível em: <http://www.efdeportes.com/efd97/gols.htm>
3. Farias FF. **Análise e previsão de Resultados de Partidas de Futebol**. [Dissertação]. Rio de

Rev Bras Futebol 2014 jul-dez 07(2): 35-41

- Janeiro: Universidade Federal do Rio de Janeiro; 2008.
4. Pollard R, Silva CD, Medeiros NC. **Home advantage in football in Brazil: differences between teams and the effects of distance traveled.** Braz J Soccer Sci 2008 Jan-Jul; 01(1): 03-10
 5. Silva CD da, Medeiros NC, Silva ACD da. **Vantagem em casa no campeonato brasileiro de futebol: efeito do local do jogo e da qualidade dos times.** Rev bras Cineantropom Desempenho Hum, 2010, 12(2):148-154. [Internet] 2010 [acesso em 13 jul. 2013]. Disponível em: <http://www.scielo.br/pdf/rbcdh/v12n2/a10v12n2>
 6. Pollard R, Gómez MA. **Home advantage in Football in South-West Europe: Long-term trends, regional variation, and team differences.** Eur J Sport Sci 2009, 9(6):341-352. [Internet] 2009 [Acesso em 20 mar. 2013] Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/17461390903009133>
 7. Anderson C, Sally D. **The Numbers Game: Why Everything You Know About Soccer Is Wrong.** P.57-70. Estados Unidos da América. Editora Schwarcz S.A, 2013.
 8. Vigne G¹, Dellal A, Gaudino C, Chamari K, Rogowski I, Alloatti G, Wong PD, Owen A, Hautier C. **Physical outcome in a successful Italian Serie A soccer team over three consecutive seasons.** J Strength Cond Res. 2013 May;27(5):1400-6. Acesso em 10 jan. 2014. doi: 10.1519/JSC.0b013e3182679382.
 9. Rosa ER, Neto FMV, Bernardes RS. **Uso da regressão logística para estimar a participação ou não de uma equipe na final de uma Copa do Mundo de Futebol.** In: Rev Mat Est Foco - ISSN: 2318-0552. Anais do XII Encontro Mineiro de Estatística - MGEST 2013. 05 e 06 de setembro de 2013. Uberlândia, 2013.
 10. Pollard, R., Ensum, J., & Taylor, S. (2004). **Estimating the probability of a shot resulting in a goal: The effects of distance, angle and space.** Int J Soccer Sci, 2(1), 50-55.