

COMPARAÇÃO ESTATÍSTICA DE COMPETITIVIDADE DE CAMPEONATOS NACIONAIS DE FUTEBOLGustavo Fontoura de Souza¹
Luiz Ramiro Araújo²**RESUMO**

Introdução: Os Campeonatos de Futebol são disputados em quase todos os países do mundo e chamam a atenção pela disputa de títulos em cada um deles. O Futebol é o esporte mais popular do mundo e muito dinheiro é movimentado em torno dele, contudo não é igualmente competitivo em todos os países. Objetivos: Neste artigo, aborda-se a competitividade de campeonatos nacionais de 12 países visando encontrar similaridades/diferenças entre eles. Amostra: A análise é feita com base em 60 campeonatos realizados 12 países selecionados segundo critério de desempenho. Considera-se, para cada país e ano, o campeão e o vice-campeão. Métodos: Define-se indicadores para a comparação, seleciona-se os indicadores e utilizando a técnica de clustering encontra-se grupos de similaridades. Resultados: Observa-se que existem três grupos de campeonatos, com respeito a competitividade. Em um deles, o menos competitivo, apresenta grande concentração de títulos obtidos por apenas dois clubes. Conclusões: Através da análise de agrupamentos pode-se apresentar a formação dos grupos, na qual os resultados são similares a análise com o Top5.

Palavras-chaves: Equilíbrio. competitividade. futebol. Análise de agrupamentos. Clustering.

ABSTRACT

Statistical comparison comparison of national football championships

Introduction: Football championships are played in almost every country in the world and draw attention to the title challenge in each of them. Football is the most popular sport in the world and a lot of money is bustling around it, yet it is not equally competitive in all countries. Objectives: In this paper, the competitiveness of national championships in 12 countries is analyzed in order to find similarities/differences between them. Sample: The analysis is based on 60 championships held 12 countries selected according to performance criteria. It is considered, for each country and year, the champion and the runner-up. Methods: Indicators are defined for the comparison, the indicators are selected and using the clustering technique are groups of similarities. Results: It is observed that there are three groups of championships, with respect to competitiveness. In one of them, the less competitive, presents great concentration of titles obtained by only two clubs. Conclusions: Group analysis can show the group formation, in which the results are similar to the Top5 analysis.

Key words: Balance. competitiveness. football. Analysis of clusters. Clustering.

1-Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN), Natal-RN, Brasil.

2-Universidade Federal do Rio Grande do Norte (UFRN), Natal-RN, Brasil.

E-mails dos autores:
gfontoura@gmail.com
apontadores@gmail.com

INTRODUÇÃO

O futebol é, sem dúvida, um dos esportes mais populares do mundo. Segundo a FIFA (Fédération Internationale de Football Association) (Count, 2006), o futebol tem 265 milhões de jogadores no mundo.

Tal quantidade de adesão faz com que o futebol seja, não só o esporte mais praticado, como também o mais assistido. Muitas competições são organizadas pelas federações e disputadas a cada ano.

Além desses milhões de praticantes, existem outros milhões de expectadores, fazendo com que o futebol movimente muito dinheiro.

Os Campeonatos Nacionais de Futebol são disputados em todos os mais de 200 países filiados a FIFA. A maioria deles disputados sob a forma de "pontos corridos", um sistema de disputa na qual cada equipe joga com todas as outras duas vezes (jogos de ida e volta), ao final o que obtiver maior número de pontos é declarado campeão. Todos os anos diversos produtos relacionados ao futebol são comercializados, como diretos de imagens, álbuns de figurinhas, revistas, ingressos para jogos, sinais de tv fechada, apostas, materiais esportivos e etc. Além disso, as discussões sobre futebol, sobretudo no Brasil, levam horas e horas, tanto em canais de televisão com nas ruas. Os "chutes", análises e apostas sobre quem vai vencer os campeonatos são muitos e começam mesmo antes dos campeonatos se iniciarem.

Dois fatores parecem determinantes para a popularidade de um campeonato. Uma é o nível dos clubes e jogadores, fator muito correlacionado com poder aquisitivo dos clubes. Outro é o equilíbrio competitivo do campeonato, ou seja, a possibilidade de uma disputa efetiva entre a maior quantidade de clubes.

Neste trabalho abordaremos a competitividade dos campeonatos como característica principal do mesmo, buscando classificá-los segundo essa característica.

Trabalhos relacionados

Diversos pesquisadores têm utilizado o futebol como forma de pesquisa em diversas áreas do conhecimento. Almeida, Oliveira e Silva (2011) apresenta uma comparação entre jogar em seu próprio campo e jogar no campo

do adversário, para isso, utiliza os resultados do campeonato brasileiro de futebol masculino das séries A e B de 2003 a 2009. Seguindo a mesma linha de pesquisa, Silva e Moreira (2008) compara ligas de oito países (Alemanha, Argentina, Brasil, Espanha, França, Inglaterra, Itália e Portugal) segundo o mesmo critério de estar jogando no seu próprio estádio ou não. Os resultados de ambos os estudos são que as ligas apresentam diferenças nesse critério, na qual, as diferenças regionais e distância entre os locais dos jogos podem ser as principais responsáveis.

Já em uma outra linha de comparação Seligman (2011) utiliza cadeias de Markov para comparar o campeonato italiano e o campeonato brasileiro de futebol. Nessa comparação utiliza dados de 2003 a 2010 e utiliza como variável a diferença do número de gols marcado pelo clube que joga no seu estádio subtraído pelo número de gols do visitante.

Já Artuso (2017) apresenta uma série de testes estatísticos para mostrar que o desempenho dos clubes ao final do campeonato segue uma distribuição normal, sendo assim possível prever a pontuação necessária para obter determinado objetivo.

Com relação a competitividade, Drummond, Júnior e Shikida (2013) apresenta um balanço dos anos de 1971 a 2009 do campeonato brasileiro. Para isso, se utiliza de diversos índices como "Top4", "C4", Gini e etc. Através desses índices foi possível observar que o campeonato brasileiro tornou-se mais competitivo ao longo dos anos, sobretudo quando passou a ser disputado através dos pontos corridos.

Já em Júnior, Shikida e Monasterio (2003), o campeonato brasileiro é analisado por estado e verificado a influência de fatores relativos a renda dos estados e influência política sobre os resultados dos campeonatos de 1971 até 1998. Esses trabalhos apresentam resultados interessantes e alguns corroboram com a realidade que já conhecemos.

Contudo, neste artigo propomos comparar os campeonatos com um pouco mais de tempo (utilizaremos 60 campeonatos passados) e outros países não utilizados como Uruguai e Holanda, além de buscar identificar algum padrão nos dados.

Construção das Bases de dados

Os dados utilizados nestas análises foram definidos utilizando alguns critérios. A análise consiste em comparar campeonatos quanto a sua competitividade.

Existem aproximadamente 211 países filiados a FIFA e que apresentam algum tipo de campeonato de futebol. Destes, em função da disponibilidade de dados e da importância para o futebol mundial, foram selecionados 12 ($\approx 5\%$).

Países selecionados

A fim de eleger os Países que fariam parte da análise utilizou-se o critério de melhores desempenhos em torneios internacionais. Assim, o primeiro critério foi a Copa do Mundo de Futebol, realizado desde 1930. Isto é, inclui-se todos os países que já venceram a Copa do Mundo, que são oito países.

O segundo critério advém dos campeonatos continentais da Europa (Champions League) e da América do Sul (Libertadores da América). Será incluído o país que apresente pelo menos dois títulos continentais. A Tabela 1 apresenta os Países considerados o critério utilizado e o resultado da aplicação do critério.

A partir da seleção dos países foram eleitos os dados que deveriam ser colhidos de cada campeonato. Como o campeonato Brasileiro apresentam problemas de registro e validade para dados anteriores a 1.959, esse foi o ano escolhido para iniciar a pesquisa. Dessa forma, para cada país foram considerados os 60 últimos campeonatos nacionais realizados, e em cada um deles foi observado o campeão e o vice-campeão.

Apesar de simples, o critério adotado foi aperfeiçoado para contemplar as mais diversas realidades, uma vez que cada país teve a liberdade em cada ano de formular o regulamento do seu campeonato. No Brasil, por exemplo, em alguns anos a quantidade de participantes superou os 40 clubes e em outros apenas 16. Já na Argentina, em alguns anos teve-se dois campeonatos sendo os dois clubes declarados campeões nacionais. A tabela 2 mostra os campeonatos considerados em cada País. Os dados foram obtidos dos sites das confederações de futebol de cada país.

Índices de competitividade

A partir da definição do conjunto de dados apresentado na Tabela 2 foi construído nove índices com o objetivo de medir características dos campeonatos. Como para cada País tem-se 120 observações (Campeão e Vice-Campeão de cada torneio), elaborou-se o conjunto formado pelos índices mostrados na tabela 3. Esses índices referem-se a tabulação dos dados de campeão e vice-campeão de cada país.

Alguns desses índices estão correlacionados entre si, como por exemplo Clubes e Média, já que a média é calculada como sendo a quantidade de ocorrências (120 para todos) dividido por Clubes.

Tabela 1 - Países Selecionados para compor a Base de dados.

País	Critério	Resultado do Critério
Alemanha	Copa do Mundo	Campeão em 1954, 1974, 1990 e 2014
Argentina	Copa do Mundo	Campeão em 1978 e 1986
Brasil	Copa do Mundo	Campeão em 1958, 1962, 1970, 1994 e 2002
Colômbia	Libertadores da América	Campeão com Atlético Nacional (2) e Once Caldas (1)
Espanha	Copa do Mundo	Campeão em 2010
França	Copa do Mundo	Campeão em 1998
Holanda	Champions League	Campeão com Ajax(4), PSV (1) e Feyenoord (1)
Inglaterra	Copa do Mundo	Campeão em 1966
Itália	Copa do Mundo	Campeão em 1934, 1938, 1982 e 2006
Paraguai	Libertadores da América	Campeão com Olímpia (3)
Portugal	Champions League	Campeão com Benfica(2) e Porto (2)
Uruguai	Copa do Mundo	Campeão em 1930 e 1950

Legenda: * Número entre parêntese indica a quantidade de vezes que o clube foi campeão.

Tabela 2 - Campeonatos considerados em cada País.

País	Campeonatos	Quantidade	Observações
Alemanha	1956-2015	60	
Argentina	1983-2015	60	Em alguns anos tiveram 2 campeonatos por ano.
Brasil	1959-2015	60	Os anos de 1967, 1968 e 1987 houveram dois torneios ou campeões
Colômbia	1969-2015	60	A partir de 2002 passou a ter 2 torneios por ano.
Espanha	1956-2015	60	
França	1955-2015	60	O ano de 1993 não foi considerado na análise
Holanda	1956-2015	60	
Inglaterra	1956-2015	60	
Itália	1956-2015	60	
Paraguai	1964-2015	60	A partir de 2008 passou a ter 2 torneios por ano.
Portugal	1956-2015	60	
Uruguai	1956-2015	60	

Tabela 3 - Índices selecionados para compor a base de dados.

Ordem	Índice	Significado
1	Clubes	Quantidade de clubes diferentes no conjunto de dados
2	Campeões	Quantidade de clubes diferentes que foram campeões
3	Média	Quantidade média de ocorrência por clube
4	Desvio Padrão	Desvio padrão associado a quantidade de ocorrência por clube
5	Curtose	Curtose associada a quantidade de ocorrência por clube
6	Mediana	Mediana associada a quantidade de ocorrência por clube
7	Moda	Moda associada a quantidade de ocorrência por clube
8	Concentração 2	Quantidade de ocorrências dos 2 clubes com maiores ocorrências
9	Clubes 50	Menor quantidade de clubes cujas ocorrências somam 60

Após ser calculado os índices, realizou-se uma análise exploratória afim de observar os índices e verificar a necessidade de manter todos eles na base de dados. Algumas características são indesejáveis para o processamento, como baixa variância, alta correlação com outros índices, redundância de informações, etc. Após a realização dessa análise, encontrou-se o conjunto de sete índices para fazer parte da base de dados final. Foram retirados os índices "moda" e "média".

Uma vez obtidos os índices que serão considerados, foram realizada uma análise de agrupamentos com os doze países através do algoritmo de agrupamento hierárquico.

Algoritmo de Agrupamento Hierárquico

O Algoritmo de agrupamento Hierárquico é uma técnica que permite agrupar elementos a partir de alguma medida de similaridade. Essa técnica consiste em medir a similaridade entre os grupos e se os grupos apresentarem similaridade inferior a um limiar preestabelecido, os grupos serão agrupados. O Algoritmo apresentado no quadro 1 mostra o pseudo-código do método de clustering hierárquico.

Quadro 1 - Algoritmo de agrupamento hierárquico.

1. Faça um agrupamento para cada elemento
2. Encontre os pares de agrupamentos mais similares, de acordo com uma medida de distância escolhida
3. Funda-os em um agrupamento maior e recalcule a distância deste agrupamento todos os outros elementos
4. Repita os passos 2 e 3 até sobrar um único agrupamento.

Fonte: Linden (2009).

Aplicação do Algoritmo aos dados

O dendrograma é tipo de gráfico, em forma de árvore, que apresenta no eixo vertical uma escala de similaridade e no eixo horizontal as instâncias, a medida que a similaridade aumenta as instâncias são agrupadas. A figura 1 apresenta um exemplo de dendrograma. Para aplicação dos dados foi utilizado o Software R. O R auxiliou o processo de cálculo dos agrupamentos e a construção dos gráficos.

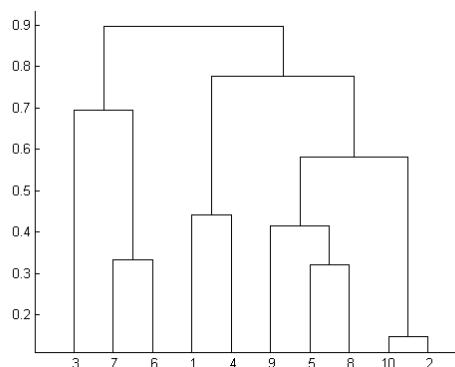


Figura 1 - Exemplo de Dendrograma.

RESULTADOS

Os índices foram construídos para evidenciar o equilíbrio do campeonato. Sendo o Equilíbrio competitivo, neste contexto, definido, segundo Michie e Oughton (2004), como sendo igualdade ou equidade existente entre quantidades. Utopicamente, campeonato perfeitamente equilibrado seria aquele em que todos os N clubes tem a mesma probabilidade (1/N) de vencer. Dessa forma, o campeonato A é mais equilibrado que o B se em A houver mais clubes com a probabilidades estatisticamente não nulas de vencê-lo do que em B.

O Campeonato Português, por exemplo, apresentou em 60 edições apenas quatro vencedores. Isso dá uma ideia de que a probabilidade de um clube, diferente dos quatro que já venceram, vencer o campeonato é muito menor que 1/N. Isso evidencia um campeonato pouco equilibrado. Já o Campeonato Brasileiro, nas mesmas 60 edições apresentaram 17 campeões diferentes, indicando um maior equilíbrio.

Considerando esses critérios, a aplicação do algoritmo de agrupamento Hierárquico fornece grupos mediante um nível a ser considerado. A figura 2 apresenta do dendrograma obtido a partir do algoritmo hierárquico. É possível observar os grupos com maiores similaridades. Considerando o número de grupos igual a três (equilibrado, intermediário e desequilibrado), encontra-se o limiar de 30 (linha pontilhada na figura 2) e os grupos formados são mostrados na tabela 4.

Observa-se que o grupo "equilibrado" é formado pelo Brasil e França, o grupo "desequilibrado" é o formado pelo Uruguai, Espanha, Portugal e Paraguai, enquanto que os demais países estão incluídos no grupo "intermediário".

É possível observar também, pelo dendrograma uma subdivisão dentro dos grupos. Nos desequilibrado, o grupo formado por Holanda e Paraguai diferem um pouco do formado por Espanha, Portugal e Uruguai. Da mesma forma acontece no grupo Intermediário, somando dois ou três subgrupos.

Tabela 4 - Grupos formados pela aplicação do limiar.

Grupo	Nome do Grupo	Países
1	Desequilibrado	Portugal, Espanha, Uruguai, Holanda e Paraguai
2	Equilibrado	Brasil e França
3	Intermediário	Argentina, Colômbia, Itália, Alemanha e Inglaterra

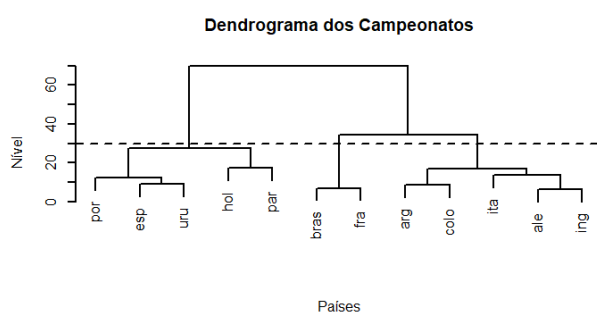


Figura 2 - Dendrograma obtido para os países.

Um fator importante que deve ser levado em consideração é que os países da América do Sul apresentaram ao longo de sua história alguns regulamentos diferentes para seus campeonatos, enquanto que os europeus usam o sistema de pontos corridos desde muito tempo. A mudança de regulamento pode ter favorecido a maior competitividade, apesar do Uruguai e Paraguai aparecerem no grupo mais desequilibrado.

Comparação com o Top5

Outra opção de Análise é a utilização do índice conhecido com "Top5". O uso de índices tipo "top4" ou "top5" permitem medir a variabilidade de equipes que chegaram ao final do campeonato entre os 4 ou 5 primeiros colocados em um determinado numero de anos. Em (6) é utilizado o "top4" por 5 anos, contudo o autor o utiliza quebrando o espaço de tempo em partes com tamanho 5, o que faz com tenha-se poucas medidas, por exemplo, uma análise de 20 anos teria apenas quatro medidas "top5". Para a utilização deste índice propomos um processo de calculo que calcula, para os mesmos 20 anos, 16 valores, pois o calcula "movendo-se" ano a ano ao longo do período. A equação XX apresenta o método de cálculo utilizado no "top5".

$$\text{Top5(ano)} = N_{\text{ano}-2}^{\text{ano}+2} \quad (1)$$

Na qual N_a^b é o número de equipes diferentes que apareceram entre as 5 primeiras posições entre os anos "a" até "b". Dessa forma, uma sequência de campeonatos de 2000 a 2015, por exemplo, teríamos o top5(2002) como o primeiro e a sequência top5(2003), top5(2004), ..., top5(2012) e top5(2013), totalizando 11 índices. A figura 3 apresenta o gráfico dos top5 dos países¹ utilizados na análise de agrupamentos.

Apesar de já ser perceptível no gráfico da figura 3 a diferença entre eles, se juntarmos os países dentro de cada um dos grupos indicados pelo algoritmo de agrupamentos observaremos uma separação mais evidente, conforme mostra a figura 4. É possível observar também que os grupos 2 e 3 não são tão dissimilares assim, como o dendrograma já induzia.

¹ Não foi encontrado dados sobre os campeonatos do Paraguai e da Colômbia

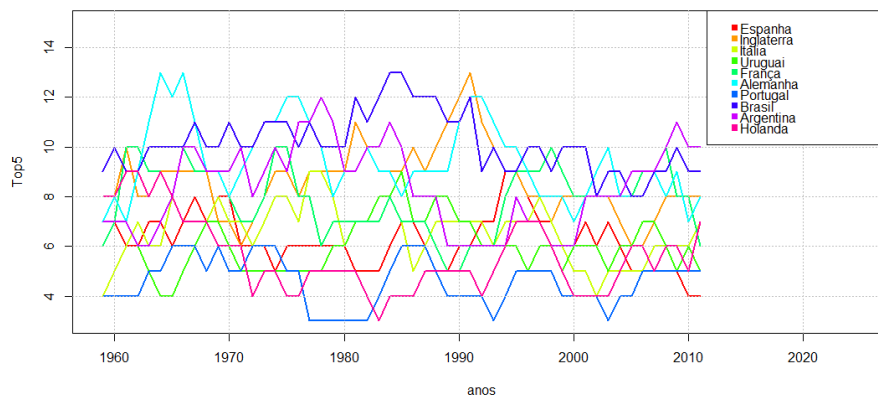


Figura 3 - Série do "top5" por país.

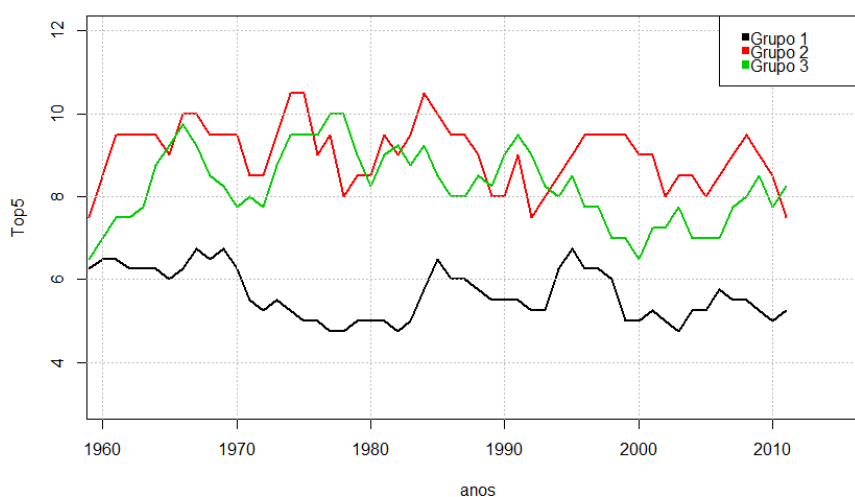


Figura 4 - Série do "top5" por grupo.

DISCUSSÃO E CONCLUSÃO

Os campeonatos de futebol são muito importantes para a população mundial, pois movimentam muito dinheiro e a atenção de milhares de pessoas.

A competitividade dos campeonatos é um fator que o caracteriza, contudo não é o único, pois campeonatos com pouca competitividade como o Espanhol apresenta uma grande média de público.

Os índices apresentados não são os únicos que indicam competitividade, contudo conseguiram captar essa característica e apresentar resultados que confirmam o que já se sabe no mundo do futebol.

O Algoritmo de agrupamento utilizado se mostrou útil para realizar a classificação e a forma de apresentação através do dendrograma é muito didático para explicitar esse comportamento.

Por fim observa-se que os resultados apresentados são semelhantes aos obtidos com o índice "top5", mesmo utilizando apenas dois clubes (campeão e vice), o que pode ser uma vantagem, pois utilizou-se menor quantidade de dados para obter conclusões similares.

Uma proposta de trabalho futuro é analisar dentro de cada campeonato o desequilíbrio com dados de outras fontes, como financeira, tamanho do elenco de

Revista Brasileira de Futsal e Futebol

ISSN 1984-4956 *versão eletrônica*

Periódico do Instituto Brasileiro de Pesquisa e Ensino em Fisiologia do Exercício

www.ibpex.com.br / www.rbff.com.br

jogadores, número médio de torcedores do estádio e etc. Ou ainda, considerar as diferenças territoriais e quantidade de clubes total por país.

Recebido para publicação em 28/02/2018
Aceito em 02/04/2018

REFERÊNCIAS

1-Almeida, L. G.; Oliveira, M. L.; Silva, C. D. da. Uma análise da vantagem de jogar em casa nas duas principais divisões do futebol profissional brasileiro. *Revista Brasileira de Educação Física e Esporte*. Vol. 25. Num. 1. p. 49-54. 2011.

2-Artuso, A. R. Análise do aproveitamento dos times no campeonato brasileiro a partir de uma distribuição normal. *Revista Brasileira de Biomecânica*, São Paulo. Vol. 25. Num. 4. p.49-63. 2007.

3-Count, F. B. 270 million people active in football. *FIFA Communications Division, Information Services*, v. 31. 2006.

4-Drummond, L.; Júnior, A. A.; Shikida, C. Campeonato brasileiro de futebol e balanço competitivo: uma análise do período 1971-2009. *Revista Brasileira de Futebol*. Vol. 3. Num. 2. p. 73-87. 2013.

5-Júnior, A. A.; Shikida, C.; Monasterio, L. Abrindo a "caixinha de surpresas": Uma análise econométrica do futebol brasileiro. *WorkingPaper (WP3)*. IBMEC/MG, 2003.

6-Linden, R. Técnicas de agrupamento. *Revista de Sistemas de Informação da FSMA*. Vol. 1. Num. 4. p. 18-36. 2009.

7-Michie, Jonathan, e Oughton, C. *Competitive balance in football: Trends and effects*. London: The sportsnexus, 2004.

8-Seligman, E. Comparação entre os campeonatos de futebol brasileiro e italiano utilizando cadeias de markov. Monografia. Universidade Federal do Rio Grande do Sul. 2011.

9-Silva, C. D. d.; Moreira, D. G. A vantagem em casa no futebol: comparação entre o campeonato brasileiro e as principais ligas nacionais do mundo. *Rev Bras Cineantropom Desempenho Hum*. Vol. 10. Num. 2. p. 184-188. 2008.