



A DISTRIBUIÇÃO DO RANKING DE CLUBES BRASILEIROS REGIDO POR UMA LEI UNIVERSAL: UMA APLICAÇÃO A LEI DE ZIPF

Claudio Djissey Shikida¹
Rodrigo Nobre Fernandez²
André Carraro³

Resumo

Objetivo do estudo: O objetivo da pesquisa foi verificar o grau de adequação do *ranking* da CBF à lei de Zipf.

Metodologia/abordagem: Foram utilizadas regressões de corte transversal e em painel para os *rankings* no período 2013-2018.

Originalidade/Relevância: Um padrão como a lei de Zipf – verificado em diversos *rankings* em outras áreas – implica em forte concentração na amostra. No caso dos clubes brasileiros, isso significaria que poucos clubes teriam possibilidades de sucesso ao longo dos anos. A pesquisa contribui na discussão sobre a concentração do sucesso no futebol em poucos clubes.

Principais resultados: Os resultados sugerem que a concentração dos clubes brasileiros é um menos forte do que a sugerida pela Lei de Zipf.

Contribuições teóricas/metodológicas: a literatura nacional apresenta poucas tentativas de se verificar o grau de concentração entre clubes de futebol brasileiros. O uso da econometria – inclusive por meio das técnicas de dados em painel – são a principal contribuição desta pesquisa.

Palavras-chave: *Ranking* da CBF, Lei de Zipf, Economia dos Esportes.

Como referenciar em APA:

Shikida, C. D., Nobre Fernandez, R., & Carraro A. (2019). A distribuição do ranking de clubes brasileiros regido por uma lei universal: uma aplicação a Lei de Zipf. *PODIUM Sport, Leisure And Tourism Review*, 8(2), 230-240. <https://doi.org/10.5585/podium.v8i2.322>.

¹ Doutor em Economia. Professor do Programa de Pós-Graduação em Organizações e Mercados (PPGOM) da Universidade Federal de Pelota (UFPel). E-mail: claudio.shikida@ufpel.edu.br Orcid: <https://orcid.org/0000-0002-8305-3022>

² Doutor em Economia. Professor do Programa de Pós-Graduação em Organizações e Mercados (PPGOM) da Universidade Federal de Pelota (UFPel). E-mail: rodrigo.fernandez@ufpel.edu.br Orcid: <https://orcid.org/0000-0001-8596-2898>

³ Doutor em Economia. Professor do Programa de Pós-Graduação em Organizações e Mercados (PPGOM) da Universidade Federal de Pelota (UFPel). E-mail: andre.carraro@ufpel.edu.br Orcid: <https://orcid.org/0000-0001-5085-8185>



THE RANKING DISTRIBUTION OF BRAZILIAN CLUBS GOVERNED BY A UNIVERSAL LAW: AN APPLICATION TO THE ZIPF LAW

Abstract

Objective: The objective of the research was to verify the degree of adequacy of the CBF ranking to Zipf's law.

Methodology / approach: Cross-sectional and panel regressions were used for the rankings in the 2013-2018 period.

Originality / Relevance: A standard like Zipf's law - verified in various rankings in other areas - implies a strong concentration in the sample. In the case of Brazilian clubs, this would mean that few clubs would have a chance of success over the years. The research contributes to the discussion about the concentration of soccer success in a few clubs.

Main results: The results suggest that the concentration of Brazilian clubs is less strong than suggested by the Zipf Law.

Theoretical / methodological contributions: The national literature presents few attempts to verify the degree of concentration among Brazilian soccer clubs. The use of econometrics - including through panel data techniques - is the main contribution of this research.

Keywords: CBF's *Ranking* , Zipf 's Law, Sports Economics

LA DISTRIBUCIÓN DE CLASIFICACIÓN DE LOS CLUBES BRASILEÑOS REGIDOS POR UNA LEY UNIVERSAL: UNA APLICACIÓN A LA LEY ZIPF

Resumen

Objetivo del estudio: El objetivo de la investigación fue verificar el grado de adecuación de la clasificación CBF a la ley de Zipf.

Metodología / enfoque: se utilizaron regresiones de panel y de sección transversal para las clasificaciones en el período 2013-2018.

Originalidad / Relevancia: Un estándar como la ley de Zipf, verificado en varias clasificaciones en otras áreas, implica una fuerte concentración en la muestra. En el caso de los clubes brasileños, esto significaría que pocos clubes tendrían una posibilidad de éxito a lo largo de los años. La investigación contribuye a la discusión sobre la concentración del éxito futbolístico en algunos clubes.

Resultados principales: Los resultados sugieren que la concentración de clubes brasileños es menos fuerte que la sugerida por la Ley Zipf.

Contribuciones teóricas / metodológicas: la literatura nacional presenta pocos intentos de verificar el grado de concentración entre los clubes de fútbol brasileños. El uso de la econometría, incluso a través de técnicas de datos de panel, es la principal contribución de esta investigación.

Palabras clave: Ranking CBF, Ley Zipf, Economía del deporte.



1 Introdução

Os torcedores de futebol são aficionados pelas posições dos seus clubes em campeonatos. Os rankings de posicionamento dos clubes e das seleções nacionais, fornecem uma medida que permite a comparação quantitativa entre essas equipes com base nas suas conquistas de títulos e pontuação em campeonatos. O indicador mais antigo foi criado pela Federação Internacional de História e Estatísticas do Futebol (IFFHS) e teve o seu início em 1991. No contexto das equipes nacionais, a Federação Internacional de Futebol (FIFA), possui um indicador próprio que classifica as seleções mundiais desde 1993.

No âmbito nacional, a Confederação Brasileira de Futebol (CBF) possui um indicador que a partir de 2013 classifica as equipes brasileiras considerando o desempenho dos times nas competições nacionais tendo como base os últimos cinco anos. Esse indicador influencia o Ranking das Federações Nacionais, sendo que as organizações com o maior número de pontos se credenciam a vagas na Copa do Brasil e no Campeonato Brasileiro da Série D.

Dentro deste contexto, Deng, Li, Cai, Bulou & Wang (2012) destacam que *rankings* podem ser considerados como um fenômeno que se difundiu extensamente na sociedade. Contudo, há uma possibilidade menos conhecida pelos gestores esportivos e torcedores que sinaliza que esses indicadores possam seguir algum tipo de lei ou comportamento estatístico. Por exemplo, o primeiro clube do ranking poderia ter o dobro de pontos do segundo, três vezes o número de pontos do terceiro colocado e assim por diante.

Esse tipo de padrão foi descoberto por George Kingsley Zipf, e tal regularidade tornou-se conhecida como *lei de Zipf*. Conforme destacam Monasterio (2004) e Lestrade (2017) nem sempre há uma explicação teórica satisfatória para indicar esse padrão de comportamento nos indicadores. Encontrada originalmente por seu autor no contexto da frequência de palavras em um texto e sua posição no *ranking*, de acordo com Fairthorne (1969), Rosen & Resnick (1980) e Adamic & Huberman (2002) a relação também pode ser verificada em diversos outros estudos acerca da visita em *sites* da internet, economia regional e urbana, etc.

A lei de Zipf pode ser explicada por meio de um exemplo ligado à Economia Regional e Urbana. Para uma amostra de cidades em uma dada região, ordena-se as mesmas conforme sua população e, caso a lei seja respeitada, ela nos mostrará que a primeira delas será duas vezes mais populosa que a segunda, três vezes mais populosa que a terceira, etc.

Dentro deste escopo, este artigo busca analisar se tal regularidade pode ser encontrada no futebol profissional brasileiro. Para isto utilizou-se o *ranking* desenvolvido pela CBF entre 2013 a 2018, contemplando o desempenho dos clubes brasileiros nas quatro divisões dos campeonatos nacionais, na Copa do Brasil e nas competições continentais da América do Sul (Copa Sulamericana e Copa Libertadores da América). Foram adotadas duas estratégias: a estimação por corte transversal e a estimação para o painel de dados no total do período.

Desta forma, não foram encontradas evidências favoráveis que o ranking da CBF siga o padrão estatístico determinado pela lei de Zipf. Em média, os coeficientes estimados ficaram em torno de -0.6%. Ademais, os resultados indicam que o ranking segue uma lei de potência menos concentrada que a Zipf, ou seja, ao invés do segundo colocado ter 50% dos pontos do primeiro, ele tem cerca de 67% pontos deste.

Em suma, esse estudo está estruturado do seguinte modo, tendo sido iniciado por essa breve introdução. A seção seguinte apresenta a lei de Zipf no contexto do ranking da CBF e também mostra algumas estatísticas preliminares. A seguir, explica-se a metodologia de construção do *ranking* da CBF, além de se discutir características da amostra. A quarta seção apresenta e discute os principais resultados. Por fim, apresentam-se as considerações finais.

2 A Lei de Zipf: Uma Breve Apresentação

A Lei de Zipf foi formulada na década de 1940 por George Kingsley Zipf, linguístico da Universidade de Harvard que ao analisar a frequência com que as palavras aparecem em uma obra verificou uma regra de comportamento na qual a segunda palavra mais utilizada possui uma frequência que é metade da



primeira, a terceira palavra uma frequência de 1/3 e assim sucessivamente. A evidência empírica de Zipf implica que a partir de uma lista hierárquica a distribuição dos elementos desta lista será dada por alguma força da natureza de tal forma que a sua distribuição se aproxime de $P_n \sim \frac{1}{n^a}$, onde P_n representa a frequência de uma palavra ordenada na n -ésima posição e o expoente a é 1.

Em outras palavras, Zipf constatou que, enquanto poucas palavras são utilizadas com muita frequência, a maioria delas são raramente utilizadas. Outra forma de dizer isso é que há uma concentração de palavras. E, o mais importante, é que a relação entre elas é previsível. A partir daí diversos pesquisadores, como por exemplo, Fairthorne (1969), Rosen & Resnick (1980), Adamic & Huberman (2002), Naldi (2003), Monasterio (2004) e Lestrade (2017) exploraram a existência desta lei empírica para as mais diversas áreas do conhecimento.

Em especial para o futebol a evidência de existência da Lei de Zipf implicaria que poucos clubes poderiam ser “gigantes”, grandes equipes vencedoras, enquanto a maioria dos clubes raramente ganharia algum campeonato. Alterações nas posições dependeriam de eventos aleatórios, como por exemplo, o contrato inesperado com um novo patrocinador, que afetasse a capacidade de investimento dos clubes existentes.

No caso dos clubes de futebol, seja R a colocação (*rank*) do clube. A lei de Zipf é um caso particular da expressão (1), como segue:

$$R = AS^{-b} \quad (1)$$

Em que: $R = Rank$ do clube “ i ”, $A =$ constante, $S =$ pontos no *ranking* e $b =$ expoente de Pareto.

Quando $b = 1$ (o caso em que a lei de Zipf ocorre), tem-se que $RS = A$, ou seja, o produto da colocação do clube no *ranking* com sua pontuação seria uma constante. Uma outra forma de se verificar a lei é logaritizar (1). Assim, tem-se que:

$$\ln R = \ln A - b \ln S \quad (2)$$

Nota-se que, neste caso, o parâmetro “ b ” é a elasticidade-pontos do *ranking*, o que fornece uma outra forma de se enunciar a lei de Zipf que é por meio do conceito de elasticidade. Ou seja, caso S aumente em 1%, a posição no *ranking* diminui em $b\%$. Cabe ressaltar um aspecto que pode gerar confusão nesta leitura:

como um clube melhor no *ranking* está em posições numericamente menores (primeiro, segundo, etc), isso significa que a diminuição é sinônimo de melhora no *ranking*, pois o clube se aproxima das primeiras posições.

Quando a lei de Zipf é satisfeita, $b = 1$ e, portanto, ter-se-ia um caso de elasticidade unitária, isto é, aumentos percentuais no número de pontos corresponderiam a uma queda (melhora) percentual idêntica na posição do clube no *ranking*, uma observação, no mínimo, curiosa, caso seja verificada.

Salienta-se que Naldi (2003) refere-se à equação (1) como a lei de Zipf generalizada. Outra possível interpretação da lei, conforme Adamic & Huberman (2002), é a de que “eventos menores” têm maior frequência do que “eventos maiores” (por exemplo, é mais comum a ocorrência de pequenos terremotos do que grandes).

A ocorrência da lei de Zipf poderia ser relacionada à noção de balanço competitivo, no sentido de que quanto mais próxima da lei de Zipf a distribuição dos clubes no *ranking* nacional, menos balanceado seria o equilíbrio competitivo nacional. Em seu estudo de 2003, Naldi mostrou que os índices de Gini e Hirschman-Herfindahl (HHI), populares medidas de (des)balanço competitivo têm uma relação com a lei de Zipf. A elasticidade-Zipf do HHI seria de aproximadamente 2.86. Ou seja, um aumento de 1% no parâmetro de Zipf seria correspondente a 2.86% de aumento no HHI. Para o Gini, a elasticidade seria de 0.99, ou seja, a lei de Zipf teria uma relação de elasticidade unitária com o índice de Gini.

Desta forma, uma contribuição deste trabalho é oferecer mais uma medida de balanço competitivo, por meio da estimação dos parâmetros de (2) para os *rankings* anuais da CBF, em suas edições de 2013 a 2018. Assim, a próxima seção fornecerá alguns detalhes sobre a construção do *ranking* de clubes nacional, bem como, apresentará alguns dados referentes a esse indicador.

3 O *ranking* de clubes da CBF

Os campeonatos de futebol brasileiros são organizados pela Confederação Brasileira de Futebol (CBF), uma entidade privada sob a qual operam as federações estaduais de futebol que, por sua vez, gerenciam os campeonatos estaduais. A CBF deriva da “Federação



Brasileira de Sports”, fundada em 1914, posteriormente renomeada para “Confederação Brasileira de Desportos (CBD)”, em 1916. Até 1979, a CBD respondia por uma gama maior de esportes. Neste ano, seguindo determinação da FIFA, criou-se a CBF para cuidar exclusivamente do futebol.

Desta forma, a CBF é responsável pela organização, dentre outros, do Campeonato Brasileiro de futebol masculino (séries A, B, C e D), da Copa do Brasil, e pelos seguintes campeonatos regionais: (a) a Copa do Nordeste, que abrange os estados da região Nordeste; e (b) Copa Verde, que abrange o estado do Espírito Santo (da Região Sudeste) e os estados das regiões Norte e Centro-Oeste.

Historicamente, o *ranking* da CBF pode ser dividido em três períodos: (a) até 2002 era utilizado um *ranking* que apenas somava os pontos recebidos pelos times nos campeonatos brasileiros; (b) entre 2003 e 2012 houve uma alteração na metodologia de cálculo do *ranking*, buscando incluir todas as competições gerenciadas pela CBF, mas que foi seguido de várias críticas sobre distorções e, finalmente, (c) de 2003 em diante adotou-se nova metodologia, resultando na edição atual.

Dada a quebra estrutural nos critérios de classificação, este artigo trata apenas da edição mais recente do *ranking*, considerando os índices do período 2013-2018. Nesta edição, a convenção de pontos da CBF é dividida em três tabelas formuladas com o intuito de garantir uma maior quantidade de pontuação para os melhores desempenhos nos campeonatos nacionais e internacionais, um conjunto composto pela Copa do Brasil, pelo Campeonato Brasileiro (séries A, B, C, D), a Copa Sul Americana e a Copa Libertadores da América.

No caso do Campeonato Brasileiro (“Brasileirão”), a pontuação entre as séries é tal que o primeiro colocado da série A recebe 800 pontos. O segundo colocado recebe 80% do valor do primeiro colocado, o terceiro 75%, o quarto, 70% e, a partir daí o percentual segue a tendência linear (69%, 68%, etc) até o 20º colocado, que recebe 51% do valor do primeiro.

Daí em diante, todos os clubes recebem 50%. Esta escala é seguida para todas as séries (B, C, D), embora exista um redutor de forma que a pontuação da série B seja sempre metade daquela da série A, o mesmo valendo para as séries C e D.

No caso da Copa do Brasil, o primeiro colocado recebe 600 pontos e o segundo colocado, 80% deste valor (480 pontos). O restante da pontuação depende da participação do clube nas diversas fases do torneio. Clubes que *apenas* participam da Copa Libertadores e da Sul-Americana recebem pontuações fixas, respectivamente, de 400 e 280 pontos.

Vale notar que durante o período 2012-2017, no Brasileirão, não houve mudanças na contagem de pontos. Já na Copa do Brasil ocorreram alterações, já que inicialmente a competição continha 6 fases, depois passou para 7 fases, e por último aumentou de 7 fases para 8 fases. Essas alterações tiveram impacto na distribuição de pontos da convenção, visto que depois de 2012 para a Copa Libertadores e, após 2016 para a Copa Sul Americana, as equipes não receberam mais pontuações fixas por participações nesses campeonatos.

Adicionalmente, é importante destacar o aspecto *autocorrelacionado* (relacionado com passado) da construção do índice. A pontuação de um clube no Brasileirão recebe pesos relativos à sua participação nos últimos cinco anos, cadentes no tempo. É importante notar que o índice de um certo ano, diga-se, 2017, diz respeito a 2016. Assim, por exemplo, se o clube participou das edições 2012 a 2016 do Brasileirão, a equipe receberá 5 pontos pela participação em 2016, 4 pontos pela participação em 2015 e assim sucessivamente até receber 1 ponto pela participação em 2012.

Como visto, o *ranking* da CBF apresenta alguns clubes com a mesma pontuação e, portanto, há clubes com a mesma posição no *ranking*. Isso geralmente ocorre nas últimas posições, conforme se vê na Tabela 1 a seguir. A exceção é a edição de 2018, na qual dois clubes terminaram empatados em primeiro lugar.



Tabela 1 – Sobreposição de clubes no Ranking da CBF: 2013-2018

Frequências Repetidas											
2013		2014		2015		2016		2017		2018	
Pos.	Freq.	Pos.	Freq.	Pos.	Freq.	Pos.	Freq.	Pos.	Freq.	Pos.	Freq.
219	11	217	5	222	9	215	9	217	18	214	7
214	5	215	2	218	4	210	5	211	6	205	9
203	11	206	7	202	14	203	5	202	5	201	4
195	5	201	4	198	3	191	12	198	4	194	5
192	3	194	6	190	5	186	3	196	2	189	3
188	4	187	7	181	9	181	3	193	3	186	3
180	7	182	5	179	2	174	4	189	3	172	2
177	3	164	14	176	3	168	3	182	4	170	2
168	8	158	4	169	2	164	2	178	2	157	8
155	9	150	2	162	6	160	2	171	2	148	2
145	9	143	6	157	2	157	3	164	3	128	4
127	2	124	2	152	5	150	2	148	15	125	2
122	3	107	4	150	2	142	7	132	2	118	2
111	4	77	2	143	2	122	3	124	2	110	5
104	2			133	2					1	2
				117	2						
				68	2						
86		70		74		63		71		60	
229		221		230		223		234		220	

Fonte: CBF, cálculos dos autores. Pos.: posição do clube no ranking. Freq.: número de clubes nesta posição do ranking.

Nota: Na última linha, o total de clubes de cada edição do *ranking*.

Percebe-se que, em 2013, dos 229 clubes, 86 tinham a mesma colocação no ranking. A média dos percentuais de repetidos é de 31%. Nas últimas colocações há clubes com pontuações como “50” ou “25”. Como o índice é calculado para o ano anterior ao de sua divulgação (o índice CBF do ano “*t*” é sempre relativo ao do período “*t-1*”), e leva em conta uma soma do desempenho do clube em cinco anos (por exemplo, o índice de 2017 para o clube considera seu desempenho em 2013-2017, com pesos cadentes dados aos anos anteriores), estes clubes participaram, basicamente, apenas de uma fase de uma edição da Copa Brasil no período. Na sequência será exposta a metodologia que permitirá avaliar se o ranking dessa instituição segue o padrão descrito pela lei de Zipf.

4 Metodologia

Para testar se a Lei de Zipf ocorre para o caso dos clubes brasileiros foram realizadas estimações com dados de corte transversal e também na estrutura de painel. No contexto de dados de corte, fez-se a estimativa da equação (2) a cada ano do período amostral. A equação (3) apresenta a versão para dados de corte da equação (2):

$$\ln R_i = \beta_0 + \beta_1 \ln S_i + \varepsilon_i \quad (3)$$

Na estrutura de dados em painel tem-se que cada clube representa uma unidade *i* e *t* descreve a variabilidade temporal (2013 a 2018). A escolha desse período se deu pela mais recente fórmula de cálculo do índice de clubes da CBF. A equação (4) representa a especificação nessa forma:



$$\ln R_{it} = \alpha + \lambda_t + \ln S'_{it} \beta + A'_i \gamma + \varepsilon_{it} \quad (4)$$

De acordo com Wooldridge (2010), os modelos em dados de painel podem ser classificados da seguinte forma:

- Modelos “Agrupados”: A estimação é feita assumindo que os β são comuns para todos as indivíduos.
- Modelos com efeitos fixos (EF): A estimação é feita assumindo que a heterogeneidade dos indivíduos é mantida fixa no tempo. A mesma é captada na parte constante, que é diferente de indivíduo para indivíduo, captando as diferenças invariantes no tempo.
- Modelos com efeitos aleatórios (EA): A estimação é feita introduzindo a heterogeneidade dos indivíduos no termo de erro. Esse modelo considera a constante não como um parâmetro fixo, mas como um parâmetro aleatório não observável.

Formalmente, no modelo de efeitos fixos, a heterogeneidade do indivíduo é representada pela mudança paramétrica na equação (4). É como se um novo intercepto $\alpha_i = \alpha + A'_i \gamma$ que não varia em relação ao tempo fosse incluso na estimativa de mínimos quadrados ordinários. Em outras palavras, os fatores não-observáveis que poderiam estar correlacionados com a pontuação e a ordenação dos clubes são considerados constantes ao

longo do tempo. Desta forma a equação (4) toma a seguinte forma:

$$\ln R_{it} = \alpha_i + \lambda_t + \ln S'_{it} \beta + \varepsilon_{it} \quad (5)$$

É importante destacar que o termo representa λ_t uma tendência temporal, isto é, são dummies referentes a cada ano. Por outro lado, no caso de efeitos aleatórios, o termo de heterogeneidade individual é considerado como parte do termo de erro, $u_{it} = A'_i \gamma + \varepsilon_{it}$. A escolha entre os modelos é feita a partir do teste de Hausman e serão usados erros-padrão robustos para lidar com potenciais problemas de heterocedasticidade.

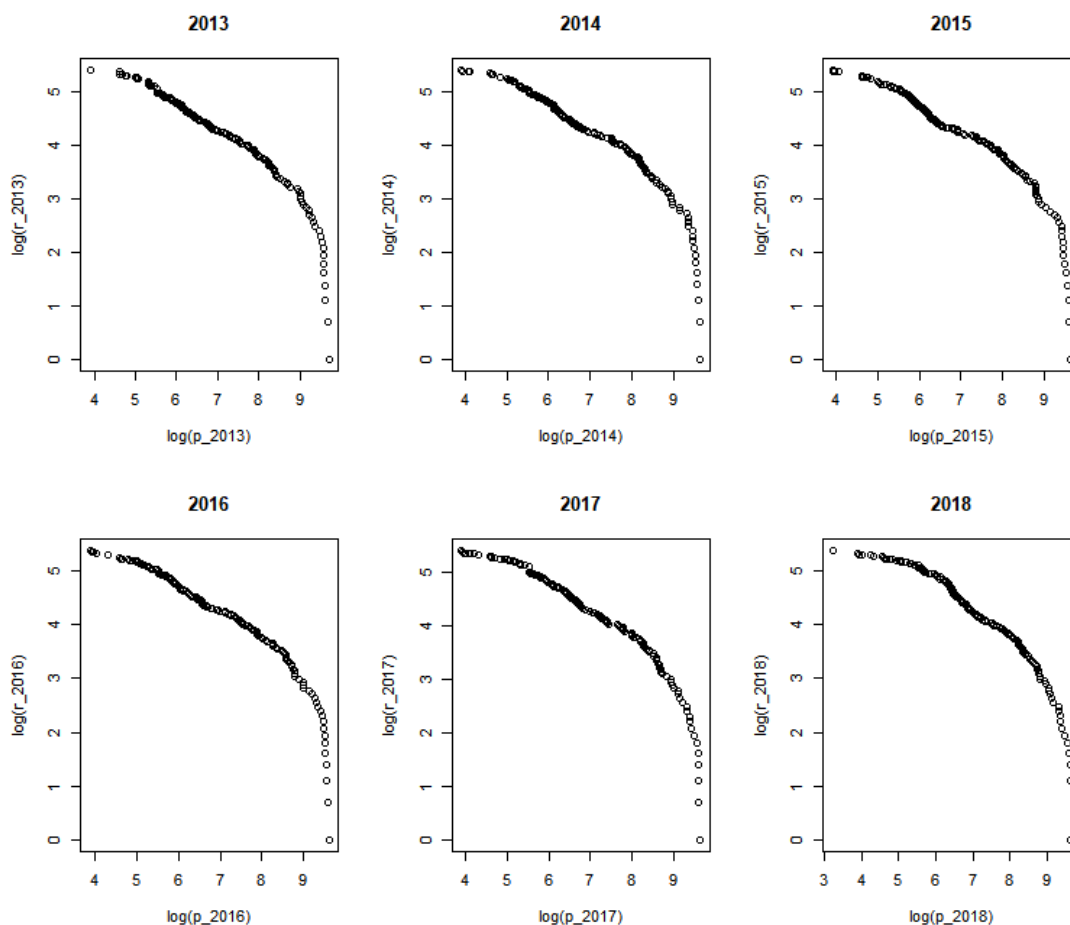
Com base nos métodos apresentados, destaca-se que todos os procedimentos estatísticos foram realizados no software R-3.6-1 (<https://cran.r-project.org/>). Na próxima seção serão apresentados os resultados das estimativas.

5 Resultados

Inicialmente, apresentam-se os diagramas de dispersão para a amostra completa (ou seja, com a ocorrência de mais de um clube com a mesma posição no *ranking*). Essas figuras mostram a relação linear (distanciamento linear) entre o logaritmo do posicionamento do clube no *ranking* da CBF e o número de pontos da mesma equipe no referido período medida na mesma escala.



Figura 1 – Diagramas de Dispersão



Fonte: autores

A figura acima confirma a relação negativa entre o posicionamento do clube e a pontuação do mesmo. Isto é, a equipe que possui um maior número de pontos está numa posição menor. Em outras palavras, o primeiro colocado no ranking possui uma pontuação bastante superior ao décimo, por exemplo, e assim sucessivamente. Além disso, é possível perceber que não há um grande distanciamento entre as pontuações, exceto na parte inferior do gráfico, em que o logaritmo do ranking está entre 1 e 2, ou seja as primeiras posições (entre 1 e 7). Para chegar-se a essa conclusão é

necessário tomar-se o antilogaritmo, ou seja, o exponencial do logaritmo de 1 e de 7.

Dando continuidade à análise procederam-se as estimativas da equação (3) ano a ano tendo como variável dependente o logaritmo do ranking de um clube e como independente o logaritmo da pontuação desta equipe. Ao todo foram estimadas seis equações, uma para cada ano do período amostral de 2013 a 2018. Salienta-se que, em todas as especificações estimadas foram utilizados erros-padrões robustos e que os parâmetros estimados foram estatisticamente diferentes de zero ao nível de 1%. Os resultados iniciais estão na Tabela 2.



Tabela 2 – Lei de Zipf – Estimacões anuais

	2013	2014	2015	2016	2017	2018
Constante	8.23	8.23	8.10	7.99	7.96	7.91
B	-0.59	-0.59	-0.58	-0.56	-0.56	-0.55
R ²	0.89	0.89	0.88	0.87	0.86	0.82
Observacões	229	221	230	223	234	220

Fonte: Autores. Nota: Todos os coeficientes foram estatisticamente diferentes de zero a 1%.

Os coeficientes estimados são menores, em valor absoluto, do que 1 (um) e, além disso, caem em pequena escala ao passar dos anos. Em outras palavras, para o ano de 2013, por exemplo, um aumento de 1% no número de pontos gera uma reducao de 0.59% na posicao. A interpretacao desse efeito evidencia a relacao negativa entre o número de pontos e posicionamento do ranking, ou seja, se um clube possui um número maior de pontos, a equipe estará numa posicao menor numericamente, o que indica estar mais bem classificado no ranking de times.

Outro ponto importante, é que as estimativas podem indicar que o índice de

clubes está mais desconcentrado ao passar do tempo, ou seja, há um maior distanciamento de pontos entre as equipes. Conforme destacam Santos, Silva, Ferla, Silva & Hein (2018) essa queda pode ter sido gerada pela mudanca no computo do cálculo do índice ou por alteracões na gestao esportiva das equipes.

A título de comparacao, segue-se Gabaix & Ibragimov (2011) em sua sugestao de estimacao da equacao (1) com correcao de vies para pequenas amostras. Basicamente, altera-se a variável dependente para $\ln(R-0.5)$. Os resultados estão na Tabela 3.

Tabela 3 - Lei de Zipf – Estimacões anuais com correcao de vies

	2013	2014	2015	2016	2017	2018
Constante	8.31	8.31	8.17	8.06	8.03	7.99
B	-0.60	-0.61	-0.59	-0.58	-0.57	-0.56
R ²	0.87	0.87	0.86	0.85	0.84	0.79
Observacões	229	221	230	223	234	220

Fonte: Cálculos dos autores. Nota: Todos os coeficientes foram estatisticamente diferentes de zero a 1%.

Observa-se que não há diferenca significativa entre os resultados. Em ambos os casos, além de não seguir a lei de Zipf, o coeficiente mantém-se relativamente estável, mas com uma queda de -0.04 desde 2013 até o final da amostra. Os resultados mostram que, em média, o aumento de 1% nos pontos gera um avanço de 0.6% no ranking. Vale destacar que não se verificou a aproximacao estimada por

(Naldi, 2003) para as estimativas de Zipf e os índices HHI ou Gini.

Como um teste adicional, foram também estimadas as regressões apenas para os clubes que aparecem em todas as edicoes do ranking. Os resultados, apresentados na tabela 4 a seguir, indicam a mesma tendncia geral de queda moderada do valor absoluto do coeficiente.

Tabela 4 – Lei de Zipf – estimacões para equipes que aparecem em todas as edicoes do ranking da CBF

	2013	2014	2015	2016	2017	2018
Constante	8.77	8.76	8.49	8.45	8.63	8.12
b	-0.67	-0.66	-0.63	-0.62	-0.65	-0.58
R ²	0.87	0.86	0.89	0.85	0.86	0.81
Observacões	135	135	135	135	135	135

Fonte: Cálculos dos autores. Nota: Todos os coeficientes foram estatisticamente diferentes de zero a 1%.

Pode-se notar que em termos absolutos o coeficiente é um pouco menor do que nas

estimativas das regressões anteriores. O modelo estimado para os clubes que sempre aparecem



na amostra no ano de 2018 se mostra mais próximo das estimativas anteriores. Com o objetivo de verificar se o efeito encontrado persiste, faz-se a estimativa para todo o período

de 2013 a 2018 utilizando a abordagem de dados em painel expressa nas equações (3), (4) e (5).

Tabela 5 – Lei de Zipf - estimativas em Dados em Painel

	(1)	(2)	(3)	(4)	(5)	(6)
Lpontos	-0.57*** (0.01)	-0.57*** (0.01)	-0.38*** (0.01)	-0.39*** (0.01)	-0.46*** (0.02)	-0.47*** (0.02)
constante	8.07*** (0.07)	8.11*** (0.07)	6.83*** (0.07)	6.93*** (0.07)	7.39*** (0.08)	7.50*** (0.09)
<i>Dummies</i>	não	Sim	não	sim	não	sim
<i>Temporais</i>						
<i>Modelo</i>	MQO	MQO	EF	EF	EA	EA
<i>Observações</i>	1344	1344	1344	1344	1344	1344
adj. R ²	0.866	0.866	0.611	0.616	0.611	0.616

Fonte: Elaborado pelos autores. Nota: Todos os coeficientes foram estatisticamente diferentes de zero a 1%.

O sentido do coeficiente estimado usando dados em painel não difere muito das estimativas anteriores. Embora a estimação dos modelos de efeito fixo (EF) seja mais adequada (conforme o teste de Hausman), mostram-se também as estimativas em MQO (agrupado) e com efeito aleatório (EA). Em resumo, ambos os modelos (1) e (2) informam que um aumento de 1% na pontuação faz com que o clube aumente de posição em 0.57%. Esse impacto cai em 0.2 pontos ao realizarem-se as regressões com efeitos aleatórios nos modelos (5) e (6). Já nas estimativas de efeito fixo, um aumento de 1% no número de pontos faz com que a equipe cresça em aproximadamente 0.4% posições.

Os modelos (3) e (4) consideram que os fatores não-observáveis que poderiam estar correlacionados com a pontuação e o ranking não estão no termo de erro e sim são fixos ao longo do tempo. Esses fatores, por exemplo, poderiam ser considerados como as diferenças regionais entre os estados brasileiros. Por sua vez, essas diferenças podem ser decisivas na decisão de investimento de um clube ou na migração de jogadores em virtude das categorias de base. Tal como nas estimações cross-section, os resultados das estimações em dados de painel também apontam para uma competitividade ligeiramente maior do que a prevista pela lei de Zipf.

Em síntese as estimativas que foram procedidas não estão em congruência com a Lei de Zipf. Os resultados para as estimativas em dados de corte e também em painel indicaram que há uma relação negativa e estatisticamente significativa entre o número de pontos e o

posicionamento da equipe no ranking de clubes da CBF. Esse relacionamento indica que um clube com melhor posicionamento possui um maior número de pontos.

6 Considerações Finais

A competitividade de campeonatos tem sido alvo de estudos, como por exemplo os de Coates & Humphreys, (2012) e Coates, Humphreys, & Zhou, (2014), na literatura de economia dos esportes. Ainda que não haja consenso sobre seu impacto sobre a demanda por ingressos, um dos maiores desafios segue sendo a sua mensuração mais adequada.

Este artigo teve por objetivo abordar este problema por meio da hipótese da lei de Zipf. Esta “lei” é uma forma alternativa de se visualizar a concentração no caso dos clubes de futebol profissionais brasileiros. Usando o ranking da CBF - no período 2013-2018 - não se encontram evidências favoráveis à lei de Zipf, pois os coeficientes estimados ficam em torno de -0.6. Em linhas gerais, encontrou-se evidências que há uma relação negativa e estatisticamente significativa entre a pontuação de um clube e sua colocação no ranking, isto é, equipes com um número maior de pontos estão melhor classificadas.

A evidência é de que o ranking da CBF segue uma lei de potência, mas não a de Zipf. Isso significa que a diferença entre a pontuação dos clubes no ranking é ligeiramente menos concentrada do que no caso de Zipf. Por exemplo, ao invés do segundo colocado ter 0.5



dos pontos do primeiro, ele tem cerca de 0.67 pontos deste.

Mesmo que a análise não permita identificar os determinantes desta suave queda no parâmetro estimado, pode-se especular que diversos choques tecnológicos e institucionais tenham atingido os clubes ao longo dos anos (e.g., técnicos que mudam de clube com alguma frequência levando consigo sua expertise e disseminando diferentes abordagens táticas).

Embora este estudo limite-se somente a análise da última versão do indicador da CBF espera-se que estes choques referidos não ocorram simetricamente nos clubes ao longo do ranking. Isto se daria porque esta classificação abarca clubes de todas as divisões do futebol profissional e a mobilidade de técnicos (e mesmo de jogadores) não é, geralmente, alta inter-divisões, mas sim intra-divisões, o que pode ser uma hipótese adicional para se explicar a queda moderada da concentração estimada.

Uma abordagem alternativa para uma nova agenda de pesquisa seria a inclusão de variáveis que captassem a situação econômico-financeira dos clubes. Esses indicadores poderiam auxiliar na identificação do padrão estatístico que segue o ranking da CBF, bem como, na caracterização da relação do posicionamento da equipe com sua gestão.

Referências

- Adamic, L. A., & Huberman, B. A. (2002). Zipf's law and the Internet. *Glottometrics*, 3, 143–150.
- Coates, D., & Humphreys, B. R. (2012). Game Attendance and Outcome Uncertainty in the National Hockey League. *Journal of Sports Economics*, 13(4), 364–377. <http://doi.org/10.1177/1527002512450260>
- Coates, D., Humphreys, B. R., & Zhou, L. (2014). Reference-dependent preferences, loss aversion, and live game attendance. *Economic Inquiry*, 52(3), 959–973. <http://doi.org/10.1111/ecin.12061>
- Deng, W., Li, W., Cai, X., Bulou, A., & Wang, Q. A. (2012). Universal scaling in sports ranking. *New Journal of Physics*, 14. <http://doi.org/10.1088/1367-2630/14/9/093038>
- Fairthorne, R. A. (1969). Empirical Hyperbolic Distributions (Bradford-Zipf-Mandelbrot) for Bibliometric Description and Prediction. *Journal of Documentation*, 25(4), 319–343. <http://doi.org/10.1108/eb026404>
- Gabaix, X., & Ibragimov, R. (2011). Rank-1/2: A simple way to improve the ols estimation of tail exponents. *Journal of Business and Economic Statistics*, 29(1), 24–39. <http://doi.org/10.1198/jbes.2009.06157>
- Lestrade, S. (2017). Unzipping Zipf's law. *Plos One*, 1–13. <http://doi.org/10.1038/474164a>
- Monasterio, L. M. (2004). A lei de Zipf: Rio Grande do Sul (1940-2000). *Redes*, 9(2), 181–190.
- Naldi, M. (2003). Concentration indices and Zipf's law. *Economics Letters*, 78(3), 329–334. [http://doi.org/10.1016/S0165-1765\(02\)00251-3](http://doi.org/10.1016/S0165-1765(02)00251-3)
- Rosen, K. T., & Resnick, M. (1980). The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics*, 8(2), 165–186. [http://doi.org/10.1016/0094-1190\(80\)90043-1](http://doi.org/10.1016/0094-1190(80)90043-1)
- Santos, C. A., Silva, T. B. J., Ferla, R., Silva, I. F. N. & Hein, N. (2018). O desempenho no ranking da CBF é sinônimo de solvência? Um estudo com clubes de futebol brasileiros. In: ADM 2018 - Congresso Internacional de Administração, 2018, Sucre - Bolívia. Anais do ADM 2018 - Congresso Internacional de Administração - Sucre, Bolívia. Website: https://admpg2018.com.br/anais/2018/arquivos/05142018_100529_5af99605780ad.pdf
- Wooldridge, J. M. (2010) *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass: MIT Press.