

Uma nota sobre a Lei de Zipf no futebol: o ranking da CBF (2013-2018)

Renan Makoto da Silva Kumagawa¹; Claudio Djissey Shikida²

¹Universidade Federal de Pelotas(UFPel) – rkumagawa@gmail.com

² UFPel – cdshikida@gmail.com

1. INTRODUÇÃO

Atualmente está em alta a aplicação de sistemas complexos em eventos mais próximos das ciências sociais, mesmo com origem na física, matemática e estatística. Já em pesquisas na área da ciência econômica, utilizam para padrões de especialização regional, para o mercado financeiro e, no nosso caso, para aplicação da Lei de Zipf em Economia do Futebol.

Este artigo utiliza a lei de Zipf, que estabelece um ranking de N clubes em ordem não decrescente conforme o seu tamanho sendo tal que o produto da potência da posição do item no ranking e seu tamanho é constante igual a um. Se tratando do ranking da CBF, a lei auxiliaria na demonstração que existem poucos grandes clubes e vários pequenos clubes.

Um fato interessante é que se a posição de um clube seguir uma regularidade estatística como lei de Zipf, então é plausível que outros fatores como a capacidade de gestão, contratação de jogadores bons, etc não sejam determinantes na posição de um clube no ranking ou, o que é provavelmente mais plausível, que a mesma possa ser explicada a partir de modelos que foram feitos de acordo com a teoria dos sistemas complexos.

2. METODOLOGIA

2.1. A Lei de Zipf:

Este artigo tenta verificar se o ranking da Confederação Brasileira de Futebol (CBF) apresenta o comportamento previsto da Zipf. Sabendo disso a lei de Zipf é um caso particular da expressão (1).

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

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

Quando o b=1 (evento em que a lei de Zipf ocorre), obteremos que RS = A, ou seja, o produto da colocação do clube no ranking com uma pontuação seria uma constante. Uma forma análoga de se verificar a lei é aplicar o logaritmo em (1). Neste caso, teríamos:

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

Dessa maneira, o parâmetro “b” é a elasticidade-pontos do ranking. Ou seja, quando S aumenta 1%, a posição no ranking diminui em b%. Quando a lei de Zipf vale, b = 1, teríamos um caso de elasticidade unitária, i.e. Acréscimos percentuais no número de pontos caracterizaria uma queda (melhora) percentual idêntica, na posição do clube no *ranking*. Uma observação, no mínimo, curiosa caso seja verificada.

A lei de Zipf poderia ser relacionada com o balanço competitivo, no sentido de quanto mais próxima da lei de Zipf for a distribuição dos clubes no ranking, menos balanceado seria o equilíbrio competitivo. Naldi (2003) concluiu que os índices de

Gini e Hirschman-Herfindahl (HHI), populares medidas de (des)balanço competitivo têm uma relação com a lei de Zipf.¹

Desta forma, este artigo oferece 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. A fim de se ter uma ideia do grau de desigualdade entre os clubes no *ranking* da CBF, observe o Quadro 1 a seguir.

Quadro 1 - HHI (medido em %)

	2013	2014	2015	2016	2017	2018
HHI	1.78	1.79	1.77	1.78	1.73	1.73
Min	Min: 0.43	Min: 0.45	0.43	0.45	0.43	0.45
HHI/(1/n)	409.42	396.00	406.50	396.34	404.57	380.22
Gini	0.65	0.66	0.67	0.67	0.66	0.66

Fonte: Cálculos dos autores

Nota: “Min” diz respeito ao mínimo valor do HHI em uma escala percentual para o número total de clubes da amostra, ou seja, o valor para o qual o ranking seria perfeitamente balanceado. O valor máximo é sempre 100% e significa o máximo de concentração. “n” diz respeito ao número de clubes.

Em qualquer medida, observa-se um elevado grau de desigualdade entre os clubes. No HHI sem ajuste para o número de clubes, o maior grau de competitividade é representado pelo valor mínimo, “Min”. Por exemplo, em 2018, esse corresponde a 0.45, bem distante do valor do índice, 1.73. Para o HHI ajustado pelo número de clubes, a competição perfeitamente balanceada assume sempre o valor “100”. Também nesta medida, nota-se uma forte desigualdade entre os clubes. Por fim, observando o Índice de Gini, novamente a desigualdade é marcante (os valores basicamente oscilam em torno de 0.7) – e estável – ao longo dos seis anos da amostra.

2.2. Como é elaborado o *ranking* da CBF

No Brasil os campeonatos de futebol são organizados pela Confederação Brasileira de Futebol (CBF), que por sua vez é responsável pela organização do Campeonato Brasileiro de futebol masculino das séries A, B, C e D; pela Copa do Brasil e pelos seguintes Campeonatos Regionais: (a) Copa do Nordeste, que abrange a região Nordeste, e (b) Copa Verde, que abrange o estado do Espírito Santo e os estados da região Centro-Oeste.

Este artigo trata apenas da edição do ranking que considera os períodos entre 2013-2018. Sendo que nela, a convenção de pontos da CBF é desmembrada em três tabelas. Formuladas para garantir maiores pontuações com os melhores desempenhos nos campeonatos nacionais e internacionais, os quais são formados pela Copa do Brasil, pelo Campeonato Brasileiro (séries A,B,C, D), pela Copa Sul Americana e pela Copa Libertadores da América.

No caso do Campeonato Brasileiro, a pontuação entre as séries é dividida de tal forma que o primeiro colocado da série A recebe 800 pontos. O segundo colocado pontua 80% do valor do primeiro colocado, o terceiro 75%, o quarto 70%, e a partir daí o percentual segue uma tendência linear (69%, 68%, etc) até o 27º colocado, que recebe 51% do valor do primeiro. Daí em diante todos os clubes recebem 50%.

¹ 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.

Esta mensuração é seguida para todas as séries (B, C, D), embora exista um redutor de forma que a pontuação da série B seja sempre a metade da série A, o mesmo valendo para as séries C e D.

De forma adicional, é válido destacar que a pontuação de um clube no Brasileirão recebe pesos relativos à sua participação nos últimos cinco anos, cadentes no tempo.

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. Os participantes da Copa Libertadores e da Sul-Americana recebem pontuações fixas, respectivamente, de 400 e 280 pontos².

Adicionalmente, é importante destacar o aspecto “autocorrelacionado” da construção do índice. A pontuação de um clube no Brasileirão recebe pesos relativos à sua participação nos últimos cinco anos, cadentes no tempo³.

3. RESULTADOS E DISCUSSÃO

Através do uso do Programa R, foram estimadas as regressões simples para a versão estocástica de (2), qual seja, (2').

$$\ln R_i = \beta_0 + \beta_1 \ln S_i + \varepsilon_i \quad (2')$$

A variável dependente diz respeito ao logaritmo do ranking do clube e a variável independente é o logaritmo da pontuação do mesmo. Vale ressaltar que, em todas as especificações estimadas nesta seção foram utilizados erros-padrões robustos à heterocedasticidade (White) e ambos os parâmetros estimados mostram-se estatisticamente diferentes de zero ao nível de 1%. Os resultados iniciais estão na tabela a seguir:

	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
N	229	221	230	223	234	220

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

Para comparação, seguimos Gabaix, Ibragimov (2011) em sua sugestão de estimação de (1) com correlação de viés para pequenas amostras. Basicamente, altera-se a variável dependente para $\ln(R - 0.5)$. Os resultados estão na tabela a seguir:

² 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.

³ É importante notar que o índice de um certo ano, digamos, 2017, diz respeito a 2016. Assim, por exemplo, se o clube participou das edições 2012 a 2016 do Brasileirão, ele receberá 5 pontos pela participação em 2016, 4 pontos pela participação em 2015 e assim sucessivamente até receber 1 ponto pela participação em 2012.

	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
N	229	221	230	223	234	220

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

Observa-se que não existe diferença significativa entre os resultados. Sendo que em ambos os casos, além de não seguir a lei de Zipf, o coeficiente permanece relativamente estável, mas com uma queda de segunda ordem (-0,04) desde 2013 até o final da amostra.

Os resultados apresentam que em média, o aumento de 1% nos pontos provoca um avanço de 0,6% no ranking. O resultado nos deixaria dizer que, segundo este critério, o futebol brasileiro seria um pouco desbalanceado.

4. CONCLUSÃO

A competitividade de campeonatos esportivos tem sido alvo de estudos há anos na literatura de economia dos esportes. Mesmo que não exista consenso sobre seu impacto sobre a demanda por ingressos, um dos maiores desafios ainda é a mensuração mais adequada.

Medidas como o índice de Gini e o índice de Hirschman-Herfindahl (HHI) possuem uma relação muito próxima com a lei de Zipf, como mostrou Naldi (2003). Curiosamente, o valor absoluto das estimações do parâmetro de Zipf resultaram valores próximos aos de Gini.

Durante o período estudado observou-se que a lei de Zipf não é respeitada pelos rankings da CBF, o que permite apontar que o futebol do Brasil é desbalanceado.

5. REFERÊNCIAS BIBLIOGRÁFICAS

NALDI, M. Concentration indices and Zipf's law. **Economics Letters**, v. 78, n. 3, p. 329–334, 2003.

MONASTERIO, L. M. A lei de Zipf: Rio Grande do Sul (1940-2000). **Redes**, v. 9, n. 2, p. 181–190, 2004.

GABAIX, X.; IBRAGIMOV, R. Rank- 1/2: A simple way to improve the ols estimation of tail exponents. **Journal of Business and Economic Statistics**, v. 29, n. 1, p. 24–39, 2011.