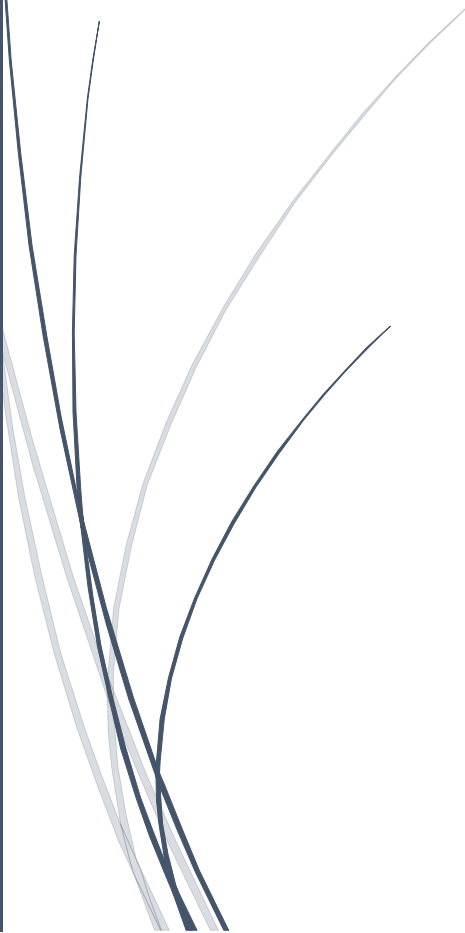


A dark blue vertical bar runs down the left side of the page. A blue arrow points to the right from this bar, containing the text 'Maio de 2020'.

Maio de 2020

Estudo de variáveis associadas à ocorrência de “cleansheets” para mandantes no Campeonato Brasileiro

Several thin, curved lines in dark blue and light grey originate from the bottom left and sweep upwards and to the right.

Lucas Rocha Rodrigues
E-mail: lucasrochex@gmail.com

Objetivo

Encontrar uma estratégia para maximizar o acerto da previsão de mandantes com *cleansheet* no campeonato brasileiro. Para tal, foram testados o efeito de seis variáveis, disponíveis antes do jogo começar, na chance de um mandante terminar o jogo com *cleansheet*, ou seja, sem ter sofrido gol.

As variáveis testadas foram:

- (1) a média de gols sofrida pelo mandante em jogos em casa;
- (2) a média de gols feitas pelo visitante em jogos fora de casa;
- (3) a soma das médias levantadas em (1) e (2);
- (4) a soma da média de gols sofridos pelo mandante em casa e fora com a média de gols feitos pelo visitante em casa e fora;
- (5) a soma da média de gols sofridos pelo mandante em casa e fora nos últimos 3 jogos com a média de gols feitas pelo visitante em casa e fora, também nos últimos 3 jogos; e
- (6) a diferença entre os pontos por partida acumulados entre o mandante e o visitante.

Metodologia

Levantou-se os resultados dos campeonatos brasileiros de 2014, 2015, 2016 e 2019, e a partir de suas tabelas calculou-se o valor de cada uma das variáveis citadas acima. Em seguida, contabilizou-se o percentual total de *cleansheets* que ocorreram em 3 intervalos dessas variáveis, estabelecidos de forma a ter aproximadamente o mesmo número de confrontos em cada um. Também, para cada ano testado, calculou-se o percentual total de mandantes que terminaram o jogo com *cleansheet* de forma a comparar esse valor com os obtidos para cada intervalo das variáveis.

Obs1: O valor de cada variável considerado foi aquele acumulado pelo time antes da partida começar. Por exemplo, para o cálculo da variável (3) para a rodada 5 foi somada a quantidade total de gols sofridos pelo mandante das rodadas 1 a 4, com a quantidade de gols feitas pelo visitante das rodadas 1 a 4.

Obs2: O estudo começou a ser feito, para todas as variáveis, a partir da quarta rodada. Isso para permitir que as respectivas médias testadas fossem representativas da realidade das equipes.

Obs3: A manipulação das tabelas dos campeonatos o levantamento das variáveis rodada a rodada e obtenção dos percentuais de *cleansheet* foram feitas utilizando a ferramenta *Jupyter Notebook*, empregando a linguagem Python. Os arquivos contendo as análises, encontram em anexo.

Obs4: Para a variável (6), o estudo foi conduzido apenas para o ano de 2019.

Resultados

As tabelas a seguir contêm o efeito de cada variável na ocorrência de *cleansheets* ano a ano. A escala de cores verde-amarelo-vermelho, foi utilizada para melhor observar possíveis relações.

Tabela 1: Percentual de "*cleansheets*" em função do histórico de gols/partida (x) sofrido como mandante

2014	< 0,7	0,7 < x < 0,9	x > 0,9	Global
	47,2	41,1	43	42,6
2015	< 0,7	0,7 < x < 0,9	x > 0,9	Global
	41,3	47,9	41,2	41,1
2016	< 0,8	0,8 < x < 1,2	x > 1,2	Global
	44,7	40,9	38,4	40,8
2019	< 0,6	0,6 < x < 1,0	x > 1,0	Global
	40	31,2	40,6	37,9

Tabela 2: Percentual de "cleansheets" em do histórico de gols/partida feitos pelo visitante (x).

2014	< 0,7	0,7 < x < 1,0	x > 1,0	Global
	50,8	39,2	38,8	42,6
2015	< 0,7	0,7 < x < 0,9	x > 0,9	Global
	46,7	42,6	41,4	41,1
2016	< 0,8	0,8 < x < 1,1	x > 1,1	Global
	42,9	41,8	38,5	40,8
2019	< 0,8	0,8 < x < 1,0	x > 1,0	Global
	41,5	36,5	35,9	37,9

Tabela 3: Percentual de "cleansheets" em função da soma (x) da média de gols sofridos em casa pelo mandante e feitos fora de casa pelo visitante

2014	< 1,5	1,5 < x < 1,9	x > 1,9	Global
	48,6	39,4	41,9	42,6
2015	< 1,5	1,5 < x < 1,9	x > 1,9	Global
	48,6	53,2	34	41,1
2016	< 1,8	1,8 < x < 2,2	x > 2,2	Global
	48,6	39,8	34,9	40,8
2019	< 1,6	1,6 < x < 2,0	x > 2,0	Global
	43,4	33,3	37,7	37,9

Tabela 4: Percentual de "cleansheets" em função da soma da média de gols sofridos pelo mandante e feitos pelo visitante (x)

2014	< 2,0	2,0 < x < 2,5	x > 2,5	Global
	50,4	36,5	39,5	42,6
2015	< 2,0	2,0 < x < 2,5	x > 2,5	Global
	49,3	44,9	38,2	41,1
2016	< 2,3	2,3 < x < 2,7	x > 2,7	Global
	50,4	39,3	33,9	40,8
2019	< 2,1	2,1 < x < 2,5	x > 2,5	Global
	51,3	33	30,3	37,9

Tabela 5: Percentual de "cleansheets" em função da soma da média de gols sofridos pelo mandante e feitos pelo visitante nos últimos 3 jogos (x)

2014	< 2,0	2,0 < x < 2,7	x > 2,7	Global
	47,1	45,2	36,1	42,6
2015	< 2,0	2,0 < x < 2,7	x > 2,7	Global
	47,2	40	43	41,1
2016	< 2,2	2,2 < x < 3,0	x > 3,0	Global
	41,5	42,6	39,8	40,8
2019	< 2,2	2,2 < x < 3,0	x > 3,0	Global
	42,2	38,6	32,8	37,9

Tabela 6: Influência da diferença de aproveitamento no campeonato - pontos/ partida (x)

2019	< -0,3	-0,3 < x < 0,25	x > 0,25	Global
	33,6	35,8	44,8	37,9

Discussão

Os resultados sugerem que existe sim relação entre o percentual de *cleansheets* e as variáveis estudadas. Caso não houvesse, ou seja, caso o percentual de *cleansheets* fosse indiferente às variáveis estudadas, o esperado seria não haver diferença expressiva entre os percentuais em cada faixa e o percentual global, o que não ocorreu.

Os resultados também evidenciam que, dentre as estratégias estudadas, a mais efetiva no sentido de aumentar a assertividade da previsão de um *cleansheet* para o time mandante é a que se baseia na soma da média de gols sofridas pelo mandante com a média de gols feitas pelo visitante como variável intervalar. Em relação à ocorrência “natural” ou global de *cleansheets*, essa estratégia, em sua pior performance (ano de 2014), captou 7,8% a mais de *cleansheets*. Já em sua melhor performance (ano de 2019), a estratégia filtrou 51,3% dos *cleansheets*, enquanto que a ocorrência global foi de 37,9%. Não cabe a esse estudo explicar o porquê dessa variável ter sido a mais influente, mas especula-se que seja pelo fato dela combinar as duas forças opostas do confronto, e também por ela calcular os poderios ofensivos e defensivos baseados em todos jogos até então dos clubes, o que reflete melhor a realidade do time do que quando esses poderios são calculados levando em consideração o mando de campo.

A última variável, a diferença de aproveitamento no campeonato entre as equipes, também se mostrou relacionada com a variação no percentual de *cleansheets*, ficando claro que quanto maior a diferença de pontos entre o mandante e o visitante, maior a chance de o mandante terminar a partida sem ter sofrido gol. Embora a amplitude da variação dos percentuais de *cleansheet* entre os intervalos não tenha sido tão pronunciada, essa variável é relevante por fugir da abordagem de gols feitos/sofridos.

Vale destacar que não apenas a previsão do *cleansheet* é relevante, como também a previsão da sua não ocorrência, sendo que a probabilidade de uma zaga ser vazada está associada com a probabilidade do ataque adversário ser bem-sucedido. Isso torna ainda mais relevante que se encontre uma variável que possua uma significativa amplitude do percentual de ocorrência de *cleansheets* em seus intervalos.

Por fim, cabe mencionar que, como pôde ser visto na apresentação dos resultados, os intervalos de cada variável variam a cada ano. Isso por que, para esse estudo, procurou-se manter um número próximo de confrontos em cada intervalo, ao invés de testar intervalos fixos que potencialmente conteriam quantidades muito diferentes de amostras. Para a utilização das variáveis em previsões futuras, sugere-se escolher como delimitadores dos intervalos valores intermediários entre aqueles obtidos a cada ano.

Conclusões

Fica evidenciado que a ocorrência global de *cleansheets* para mandantes do futebol brasileiro possui relação com as variáveis estudadas, sendo a relação mais clara, e potencialmente mais proveitosa dentre as estudadas, a soma da média de gols sofridos pelo mandante com a média de gols feitos pelo visitante no campeonato. Nos anos abordados no estudo, em média houve um aumento de 9,75% na ocorrência de *cleansheets* nos intervalos com menores valores dessa variável em relação à ocorrência global. Em outras palavras isso significa dizer que, em média, nos últimos anos a adoção da estratégia de prever a ocorrência de *cleansheets* baseado no critério explicado acima tenderia a ter uma taxa de acerto próxima a 50%, enquanto que a previsão isenta de critério renderia uma taxa de acerto próxima de 41%.

Sugestões de estudos futuros

- Efeitos das mesmas variáveis estudadas desse estudo, porém na previsão de *cleansheets* para visitantes.
- Efeito da combinação entre o poderio defensivo do mandante e ofensivo do visitante com a diferença de aproveitamento.
- Estender o estudo do efeito do aproveitamento na ocorrência de *cleansheets* para os anos não contemplados nesse estudo.