

DISTRIBUIÇÃO POISSON: uma previsão para Copa América 2019

**Lucas Silva do Amaral
Dâmocles Aurélio Nascimento da Silva Alves**

RESUMO

Esse estudo teve como propósito a previsão da Copa América de 2019, o mais importante torneio futebolístico do continente americano. Para criar a modelagem, teve-se a necessidade inicial de analisar e coletar os dados de gols marcados e sofridos pelas 12(doze) seleções participantes do certame, o espaço observado foram os jogos amistosos Pós-Copa do Mundo FIFA 2018 onde obteve-se a média ponderada dos gols, analisando o poderio de cada adversário das partidas, atribuindo os pesos de acordo com o ranking da FIFA. Ao findar, desta etapa, foi calculado, um fator ataque e um fator defesa de cada selecionado, para aplicar a distribuição de probabilidade de Poisson, no qual a sua finalidade era atribuir probabilidades de gols a serem marcados por uma seleção em um confronto do torneio, o placar mais provável de um duelo, é definido pela maior probabilidade de gols a serem marcados por cada seleção no embate. E assim pode-se prever as prováveis quartas de finais, e recalculando o fator defesa e ataque de acordo com dados previstos durante a competição, foram realizadas as previsões das semis e final, até encontrar o provável campeão da copa América de 2019.

Palavras-chave: Probabilidades, ponderada, Poisson

ABSTRACT

The purpose of this study was to predict the 2019 Copa America, the most important football tournament in the Americas. In order to create the model, there was an initial need to analyze and collect the score data scored and suffered by the 12 (twelve) selections participating in the event, the space observed were the FIFA 2018 FIFA World Cup

friendly matches where it was obtained the weighted average of the goals, analyzing the power of each opponent of the matches, assigning the weights according to the FIFA ranking. At the end of this step, an attack factor and a defense factor of each selected were calculated to apply the Poisson probability distribution, in which its purpose was to assign odds of goals to be scored by a selection in a tournament match, the most likely score of a duel, is defined by the greater likelihood of goals to be scored by each team in the match. And so it is possible to predict the likely quarters of the finals, and recalculating the defense and attack factor according to data predicted during the competition, semis and final predictions were made, until finding the probable champion of the 2019 America Cup.

Keywords: Probabilities, weighted, Poisson

1-Introdução

A Copa América é um evento futebolístico do continente sul-americano, sendo bastante tradicional disputada desde o ano de 1916 sendo organizada pela CONMEBOL desde seus primórdios, percebendo essa relevância surgiu a ideia de prever os resultados da edição de 2019 que terá como país-sede o Brasil,

Bibliograficamente não existem muitas pesquisas de previsões futebolísticas, em especial a competições entre seleções, especialmente no que é referente a a copa américa, as pesquisas com maior relevância são relacionadas a copa do mundo como afirma Filho, Suzuki, Saraiva, Louzada e Salasar (2017,p 77).

Na literatura, poucos são os artigos que dizem respeito às previsões de placares da Copa do Mundo, tais como Dyte & Clarke (2000); Suzuki et al. (2009); Bastos & da Rosa (2013). O trabalho de Dyte & Clarke (2000) propôs um modelo de regressão de Poisson log-linear que considera o ranking da FIFA como covariável. Uma abordagem bayesiana para prever os resultados das partidas, utilizando a opinião dos especialistas e o ranking da FIFA como informações a priori, foi proposta por Suzuki et al. (2009). Já Bastos & da Rosa (2013) apresentam uma

metodologia bayesiana do modelo Poisson-Gama para a previsão dos jogos da Copa do Mundo de 2010.

O modelo para a previsão deste certame , teve como base a pesquisa de Amaral & Nascimento(2018), que usaram a modelagem de Poisson para a previsão da copa do mundo FIFA 2018.

Para realizar a previsão coletou-se os dados dos jogos amistosos pós-copa 2018, das 12(doze) seleções participantes do certame , calculou-se a partir disso a média ponderada de gols feitos e sofridos de cada uma, os pesos das ponderações foram adicionados analisando o poder de cada seleção adversária nesses amistosos, analisando esta força a partir do ranking da FIFA(no momento da pesquisa) , e em seguida obtemos os fatores de ataque e de defesa de cada equipe. Para calcular as probabilidades dos jogos usou-se a distribuição de Poisson, onde calculou-se uma média λ , relacionando os fatores das seleções em cada confronto para prever a quantidade mais provável de gols marcados de cada seleção no embate, e assim determinar o placar mais provável da partida.

Foram postos como resultados todos os placares com maior probabilidade de acontecimento nos jogos e conseqüentemente simular a classificação de cada grupo , para posteriormente realizar a previsão das fases subsequentes(quartas, semi e final), tendo a necessidade de calcular o fator ataque e defesa considerando os dados acontecidos nas fases de grupos para assim prever todos os jogos dessas etapas do torneio. Observando o resultado mais provável de cada jogo , e corrigindo o empate nas fases finais, pois tende haver um vencedor, anotou-se o resultado mais provável de cada duelo, para que ao passar das fases, encontrarmos o campeão da Copa América 2019.

2. Método e Análise dos dados

Para a previsão dos jogos da copa da América, usou-se a Distribuição de Poisson, que é um modelo de distribuição discreta, que faz uma aproximação de probabilidade de fatos da realidade, através de uma média λ em relação a um certo intervalo de tempo k , representado

por um número inteiro não-negativo do qual quer encontrar a ocorrência de sua probabilidade, abaixo temos a fórmula da distribuição:

$$P(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$$

Aplicando-a para as partidas da Copa América, criou-se um modelo, associando a média λ , a relação dos fatores de ataque e defesa de cada seleção. Em relação a k , ele representa a quantidade de gols que uma seleção pode marcar em um jogo, para ajuste do modelo considerou-se k de 0 a 10, e foi calculado em cada caso, a probabilidade de ocorrência dessa quantidades de gols para cada seleção em todos os jogos da copa.

2.1 Descrição do modelo:

Teve embasamento nos dados de todas as seleções participantes da competição através dos jogos amistosos pós-copa do Mundo FIFA 2018. A média λ foi calculada, a partir das relações do fator ataque e fator defesa das seleções em um determinado embate do torneio, a obtenção desses fatores se dá através da média ponderada de gols feitos e sofridos das equipes nesses amistosos, os pesos da média ponderada foram atribuídos por influencia do Ranking da FIFA(atualizado em Abril de 2019), os critérios utilizados foram os seguintes:

Posição ranking seleção adversária x	Peso
$1 \geq x \geq 10$	3
$11 \geq x \geq 50$	2
$x \geq 51$	1

Tabela 1: Distribuição dos pesos

Prosseguindo, fez-se os cálculo das médias ponderadas de gols feitos e sofridos, para exemplificar com foi realizado esse procedimento tomamos os dados da seleção brasileira

Advérsário	GF	GS	Peso	GF*Peso	GS*Peso
EUA	2	0	2	4	0
El Salvador	5	0	1	5	0

Arábia	2	0	1	2	0
Argentina	1	0	2	2	0
Uruguai	1	0	2	2	0
Camarões	1	0	1	1	0
Panamá	1	1	1	1	1
Rep. Tcheca	3	1	2	6	2
Soma	16	2	12	23	

3 Tabela 2:Jogos da seleção brasileira pós-copa

A média ponderada de gols feitos é calculado pela razão da $\frac{\sum GF*Peso}{\sum Peso}$, que se dada por $23/12=1,916666667$.

Já a média ponderada de gols sofridos é obtido por $\frac{\sum GS*Peso}{\sum Peso}$, que é igual a $3/12=0,25$, prosseguindo as etapas, deve-se agora obter os fatores ataque e defesa, que será dado respectivamente pela razão das médias ponderadas de uma seleção, pela média ponderada geral, novamente usemos como exemplo a seleção brasileira:

FA(Fator ataque): Média GF ponderada Brasil/ Média GF ponderada geral:

$$FA = 1,916666667 / 1,435714286 = 1,334991708$$

FD(Fator defesa): Média GS ponderada Brasil/ Média GS ponderada geral:

$$FD = 0,25 / 1,071428571 = 0,233333333$$

Esse procedimento foi realizado com todas outras 11(onze) seleções e os dados dos fatores estão distribuídos na tabela a seguir:

Seleção	Fator ataque	Fator defesa
Brasil	1,334991708	0,233333333
Bolívia	0,569877883	1,018181818
Venezuela	1,114427861	0,933333333
Peru	0,895522388	1,333333333
Argentina	0,88225539	0,435555556
Colômbia	1,27694859	0,933333333
Paraguai	0,386954118	2,177777778
Qatar	1,160862355	0,725925926

Uruguai	1,144278607	1,333333333
Equador	0,905472637	0,653333333
Japão	1,519674355	0,848484848
Chile	0,750095675	1,507692308

Tabela 3: Fatores ataque e defesa das seleções

Terminada a etapa de cálculos dos fatores, deu-se início a previsão das partidas, para a realizar esta tarefa, obteve a média λ de cada seleção, tendo como intuito aplicar na distribuição de Poisson, como modelo tem-se o cálculo das médias λ do segundo jogo do Brasil contra a seleção Peruana:

λ Bra= FA Brasil * FD Peru* Média ponderada de gols feitos geral*fator correção=2,555555556

λ Per= FA Peru * FD Brasil*

Média ponderada de gols feitos geral*fator correção=0,3

E assim da mesma maneira foi feito para os outras partidas, onde em um confronto entre seleções x e y , a média λ da equipe x é calculado pelo produto, do fator de ataque de x (FAx) com o fator de defesa de y (FDy) a média ponderada de gols das seleções e o fator correção de 5%, e a média λ da equipe y é o produto da Força de ataque de y (FAy), com o fator de defesa de x (FDx) , a média ponderada de gols feitos das seleções e a correção de 5%. Todas essas médias λ , foram aplicadas na distribuição de Poisson, afim de saber a quantidade de gols mais prováveis que um seleção x pode marcar na seleção y e vice-versa, sendo calculadas as probabilidade de marcar de 0 á 10 gols, usamos como os dados obtidos acima, das seleções do Brasil e Peru:

Brasil	1,857	7,7	19,8	2,5,4	21,6	13,0	7,1	3,0	0,2	0,1	0,0	0,0	100,00%
Peru	1,708	74,1	22,2	3,3	0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,00%

Tabela 4: Previsão do jogo Brasil x Peru

Analisando esta tabela e as probabilidades de ocorrência de gols em cada coluna, percebe-se que o número mais provável de gols que o

Brasil faça contra a Peru é 2(dois) gols, com 25,4%, e em relação a equipe peruana é mais provável que não marque gols contra o Brasil, com 22,2 %, logo o placar com maior chance de acontecer, é o jogo terminar, com vitória brasileira por 2x0, com a probabilidade de aproximadamente 5,6%. Da mesma maneira fez-se com todos os 18 jogos da primeira fase, com a ajuda do Excel, sempre observando o placar com maior probabilidade para cada partida, através da distribuição das probabilidades em relação a previsão de gols marcados por uma seleção. Utilizando o placar mais provável de cada partida da 1ª fase da copa, simulou-se a classificação de cada grupo na copa, para posteriormente, prever as fases finais da copa do mundo (quartas, semi e final). Em cada grupo, passam de fase as duas primeiras seleções e os dois melhores terceiros colocados da competição.

Grupo A	JG	P	GF	GS	SG
BRASIL	3	9	4	0	4
VENEZUELA	3	6	3	1	2
BOLÍVIA	3	1	1	3	-2
PERU	3	1	2	5	-3
Grupo B	J	P	GF	GS	SG
ARGENTINA	3	7	4	0	0
QATAR	3	5	4	1	3
COLOMBIA	3	4	4	2	2
PARAGUAI	3	0	0	8	-8
Grupo C	JG	P	GF	GS	SG
JAPÃO	3	7	6	2	4
EQUADOR	3	5	3	2	1
URUGUAI	3	4	4	4	0
CHILE	3	0	1	6	-5

■ Classificados

P- Pontos	GS-Gols sofridos
GF-Gols feitos	JG- Jogos
SG-Saldo de gols	

Com a simulação da fase de grupos terminada, o próximo passo será a previsão das fases mata-mata(quartas, semi e final), havendo a necessidade de calcular novamente os fatores ataque e defesa incluindo os dados previstos na primeira fase, e com o decorrer da competição, ir adicionando os dados que foram previstos na fase anterior, em todas essas fases finais ainda tiveram influência os dados iniciais obtidos.

Para exemplificar, aqui está a obtenção dos fatores da seleção argentina, o procedimento será idêntico ao cálculo da fase inicial, com a inclusão dos valores previstos na fase de grupos :

Argentina	Pesos	GF ponderados	GS ponderados
Amistosos	15	19	7
Copa América	5	6	0
Total	20	25	7

Tabela 5: Dados Seleção argentina para as quartas

Seleções(12)	Pesos	GF ponderados	GS ponderados
Amistosos	140	201	150
Copa América	66	71	77
Total	206	272	224

Tabela 6: Dados seleções para as quartas

Analisando os dados das tabela, a média de gols feitos ponderada da seleção Argentina é igual a $25/20=1,25$ e a média de gols sofridos ponderada é igual a $7/20=0,35$.

Analogamente observando os dados das seleções temos que a média de gols feitos ponderada é $272/206=1,3204$ e a média de gols sofridos ponderada é igual a $224/206=1,0874$. Dada por encerrada a etapa, a obtenção das médias de gols ponderadas, deu-se início o cálculo dos fatores de ataque e defesa de cada seleção participante das quartas de final, o cálculo é similar ao realizado na fase de grupos.

FA= Média de GF ponderada Argentina/Média de GF ponderada seleções

$$FA=1,25/1,3204= 0,946691176$$

Da mesma maneira:

FD= Média de GS ponderada Argentina/Média de GS ponderada seleções

$$FD=0,35/1,0874= 0,321875$$

O mesmo processo aconteceu para as outras 7 (sete) seleções participantes das quartas de final , no mesmo passo do que ocorreu na primeira fase, realizou-se o cálculo da média λ , que no confronto entre as seleções x e y nesta etapa do certame é calculada por:

$\lambda_x = (FA_x) * (FD_y) * \text{Média de gols feitos seleções amistosos e copa américa}$

$\lambda_y = (FA_y) * (FD_x) * \text{Média de gols feitos seleções amistosos e copa américa}$

Para o decorrer das outras fases da competição, e necessário recalcular os fatores e a média λ de cada seleção em uma partida, os dados utilizados continuam sendo os dos amistosos e os previstos da copa américa nas fases anteriores, porém o espaço amostral vai diminuir, os dados coletados vão ser obtidos apenas das seleções que participaram da fase anterior, por exemplo para prever as semi, usam-se os dados das equipes que participaram das quartas e para a previsão da final, os dados das seleções que participaram das semi.

Para as semi- finais

- $FA_x = \text{Média de gols feitos ponderada de x nos amistosos e na copa américa / média de gols feitos ponderada pelas seleções das quartas nos amistosos e na copa américa.}$

- $FD_x = \text{Média de gols sofridos por x nos amistosos e na copa américa / média de gols sofridos na copa pelas seleções das quartas nos amistosos e na copa américa.}$

E a média λ para esta fase é calculada por:

$\lambda_x = FA_x * FD_y * \text{Média de gols feitos das seleções das quartas nos amistosos e na copa américa.}$

$\lambda_y = FA_y * FD_x * \text{Média de gols feitos das seleções das quartas nos amistosos e na copa américa.}$

Para a final:

• FA_x =Média de gols feitos ponderada de x nos amistosos e na copa américa /média de gols feitos ponderada pelas seleções das semis nos amistosos e na copa américa.

• FD_x = Média de gols sofridos por x nos amistosos e na copa américa /média de gols sofridos na copa pelas seleções das semis nos amistosos e na copa américa.

E a média λ para esta fase é calculada por:

$\lambda_x = FA_x * FD_y$ Média ponderada de gols feitos das seleções das semis nos amistosos e na copa américa.

$\lambda_y = FA_y * FD_x$ Média ponderada de gols feitos das seleções semis nos amistosos e na copa américa.

Empate a partir das quartas

Em torneios que tem fases eliminatórias como é o caso da copa América, tende haver um vencedor, mesmo que o placar mais para uma partida seja um empate, para contornar essa adversidade, tomamos como exemplo; o confronto entre Argentina e Equador previsto no modelo para as quartas de final:

	Gols	0	1	2	3	4	5	6	7	8	9	10	Prob Tot
Argentina	0,962	38,2%	36,8%	17,7%	5,7%	1,4%	0,3%	0,0%	0,0%	0,0%	0,0%	0,0%	100%
Equador	0,465	62,8%	29,2%	6,8%	1,1%	0,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100%

Tabela 7:Previsão nas quartas de Argentina x Equador antes da correção.

O placar mais provável é o empate de 0x0, mas como necessita-se existir um seleção vencedora para o empate, para suprir tal necessidade foi calculada a média λ associa a defesa de casa seleção;

$\lambda_{defArg} = (FD_{Arg}) * (FA_{Equa}) * Média ponderada de gols sofridos seleções amistosos e copa américa$

$$\lambda_{\text{defArg}} = 0,391847826^* \quad 0,926609848^*$$

$$0,893203883 = 0,324313447.$$

$\lambda_{\text{defEqua}} = (\text{FDEqua}) * (\text{FAArg}) * \text{Média ponderada de gols sofridos seleções amistosos e copa américa}$

$$\lambda_{\text{defEqua}} = 0,769701087^*$$

$$0,975378788^* \cdot 0,893203883 = 0,670572917 .$$

A partir dessa obtenção foi calculada uma nova média λ :

$$\lambda_{\text{Arg novo}} = \lambda_{\text{Arg}} + \lambda_{\text{defEqua}} = 1,632699275$$

$$\lambda_{\text{Equa novo}} = \lambda_{\text{Equa}} + \lambda_{\text{defArg}} = 0,78963274$$

E assim aplicando na distribuição de Poisson tem-se que:

		Gols	0	1	2	3	4	5	6	7	8	9	10	Prob. Tot
Equador	Argentina	1,63	19,5	31,9	26,0	14,2	5,8	1,9	0,5	0,1	0,0	0,0	0,0	100,00%
	Equador	0,79	45,4	35,9	14,2	3,7	0,7	0,1	0,0	0,0	0,0	0,0	0,0	100,00%

Tabela 8:Previsão nas quartas de Argentina x Equador pós-correção.

Com esse processo, percebe-se que é mais provável vitória da seleção Argentina, pelo placar de 1x0 classificando-se para a fase subsequente. No entanto há casos em que esse procedimento não alterará o empate para a partida, nesse caso analisará qual equipe tem a maior média λ entre os oponentes, e essa seleção será a favorita para o confronto. No modelo, o provável campeã é o equipe brasileira, tendo como vice campeão o selecionado argentino, e as seleções semifinalistas, as seleções do Qatar e do Japão.

3-Conclusão

Ao final das previsões, percebe-se que a utilização do ranking da FIFA, influenciou no resultado, visto que a provável final, foi disputada pelas seleções com melhor posicionamento, todavia percebe-se também

que os dados mais recentes tiveram relevância, pois duas seleções em bom momento e em ascensão chegaram às semifinais, o Qatar e o Japão. Vale ressaltar que o modelo de Poisson, não consegue detectar problemas extra-campo, como as distâncias percorridas entre um jogo e outro, lesões e poderio extra do mandante nesse caso o Brasil. Tem-se como sugestões para trabalhos futuros, previsões eventos esportivos coletivos.

4. Referências

ARRUDA, M. L. (2000), Poisson, Bayes, Futebol e DeFinetti, Tese de Mestrado, IME-USP.

AURÉLIO NASCIMENTO DA SILVA ALVES, DÂMOCLES ; SILVA DO AMARAL, LUCAS . Modelagem estatística para previsão de resultados de jogo de futebol: uma aplicação a copa do mundo da FIFA 2018. Diálogos: Revista de Estudos Culturais e da Contemporaneidade, v. 2, p. 75-91, 2018.

EMONET, B (2000), Revisiting Statistical Applications in Soccer, Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland.

FILHO, Ciro Alexandre Olivieri; SUZUKI, Adriano Kamimura; LOUZADA, Francisco; SARAIVA, Erlandson Ferreira; SALASAR, Luis Ernesto Bueno. Uma abordagem bayesiana para previsão de resultados de jogos de futebol: uma aplicação ao campeonato inglês. Revista Brasileira de Biometria / Biometric Brazilian Journal, Lavras, MG, Departamento de Estatística da Universidade Federal de Lavras - UFLA, v. 35, n. 1, p. 76-97, 2017. Disponível em: < <http://www.biometria.ufla.br/index.php/BBJ/article/view/296> >.