

Modelagem da diferença de gols por distribuições discretas: uma aplicação ao Campeonato Brasileiro

Lucas Pereira Belo^{1*}, Jonas Firmiano da Silva¹, Kamilla Andrade de Oliveira², Paulo César Emiliano³

¹ Universidade Federal de Viçosa, Departamento de Estatística

² Universidade Federal do Maranhão, Departamento de Engenharia Agrícola

³ Universidade Federal de Viçosa, Departamento de Estatística

¹ Discente do Programa de Pós graduação em estatística aplicada e biometria, * lucas.p.belo@ufv.br

INTRODUÇÃO

A diferença entre os gols de mandantes e visitantes, agregada por rodada, permite descrever o saldo de gols associado ao mando de campo. Como essa variável assume valores inteiros positivos, nulos ou negativos, sua modelagem requer distribuições discretas capazes de representar sua localização, dispersão e assimetria.

OBJETIVO

Comparar quatro distribuições discretas para descrever a diferença de gols por rodada do Campeonato Brasileiro da Série A entre 2006 e 2025.

METODOLOGIA

Os dados foram obtidos de Duque (2025) e referem-se, ao número de gols marcados pelos mandantes e pelos visitantes em cada uma das 38 rodadas do campeonato brasileiro para os anos de 2006 a 2025. Ao efetuar a diferença entre estas variáveis obteve-se a variável sob estudo que é a diferença entre o número de gols marcados pelos mandantes e os visitantes. O período escolhido refere-se aos campeonatos que contaram com a presença de 20 equipes com 380 jogos por ano (exceção feita a 2016 que teve 379 jogos).

Normal discreta

Se X é uma variável aleatória com distribuição Normal de média μ e variância σ^2 , isto é $X \sim N(\mu, \sigma^2)$ com função de distribuição acumulada $\Phi_X(\cdot)$, a função de probabilidade da normal discreta é dada por:

$$P[X = x] = \Phi\left(\frac{x - \mu + 1}{\sigma}\right) - \Phi\left(\frac{x - \mu}{\sigma}\right), \quad x \in \mathbb{Z}$$

Skew-normal discreta

A distribuição skew-normal contínua, proposta por Azzalini (1985) tem função densidade de probabilidade (fdp) dada por

$$f(x) = \frac{2}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left(\lambda \frac{x - \mu}{\sigma}\right),$$

em que $\phi(\cdot)$ e $\Phi(\cdot)$ representam, respectivamente, a função densidade e a função distribuição acumulada da normal padrão, enquanto $\mu, \sigma > 0$ e λ correspondem aos parâmetros de localização, escala e assimetria.

Skellam

Proposta por Skellam (1964) é a distribuição exata da diferença entre duas variáveis aleatórias independentes com distribuição de Poisson. Sejam $X \sim \text{Poi}(\lambda_1)$ e $Y \sim \text{Poi}(\lambda_2)$, com X e Y independentes, então $D = X - Y \sim \text{Skellam}(\lambda_1, \lambda_2)$ cuja função densidade de probabilidade é dada por;

$$P[D = d] = e^{-(\lambda_1 + \lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{d/2} I_{|d|}(2\sqrt{\lambda_1 \lambda_2}), \quad d \in \mathbb{Z},$$

em que $I_d(\cdot)$ denota a função de Bessel modificada de primeira espécie de ordem d .

Diferença Weibull discreta

Sejam $X \sim \text{WD}(q_1, \beta^1)$ e $Y \sim \text{WD}(q_2, \beta^2)$ variáveis aleatórias independentes, em que $\text{DW}(\cdot)$ denota a distribuição Weibull discreta de Nakagawa e Osaki (1975) cuja função de probabilidade é dada por;

$$P[X = x] = q^{x^\beta} - q^{(x+1)^\beta}, \quad x \in \mathbb{N}$$

com $0 < q < 1$ e $\beta > 0$.

Os parâmetros foram estimados pelo método da máxima verossimilhança, no *software* R. Os modelos foram comparados pelos critérios de informação, *Akaike Information Criterion* (AIC), *Bayesian Information Criterion* (BIC) e inspeção gráfica.

RESULTADOS E DISCUSSÃO

Análise descritiva

Tabela 1: Estatísticas descritivas da diferença de gols por rodada do Campeonato Brasileiro (2006–2025).

n	$\bar{X}$	Md(X)	s(X)	X_1	Q_1	Q_3	$X_{(n)}$	A(X)	K(X)
760	4,89	5	4,91	-13	2	8	22	0,02	3,36

Notação: n = número de observações; $\bar{X}$ = média amostral; Md(X) = mediana; s(X) = desvio padrão amostral; X_1 = valor mínimo; Q_1 = primeiro quartil; Q_3 = terceiro quartil; $X_{(n)}$ = valor máximo; A(X) = coeficiente de assimetria; K(X) = coeficiente de curtose.

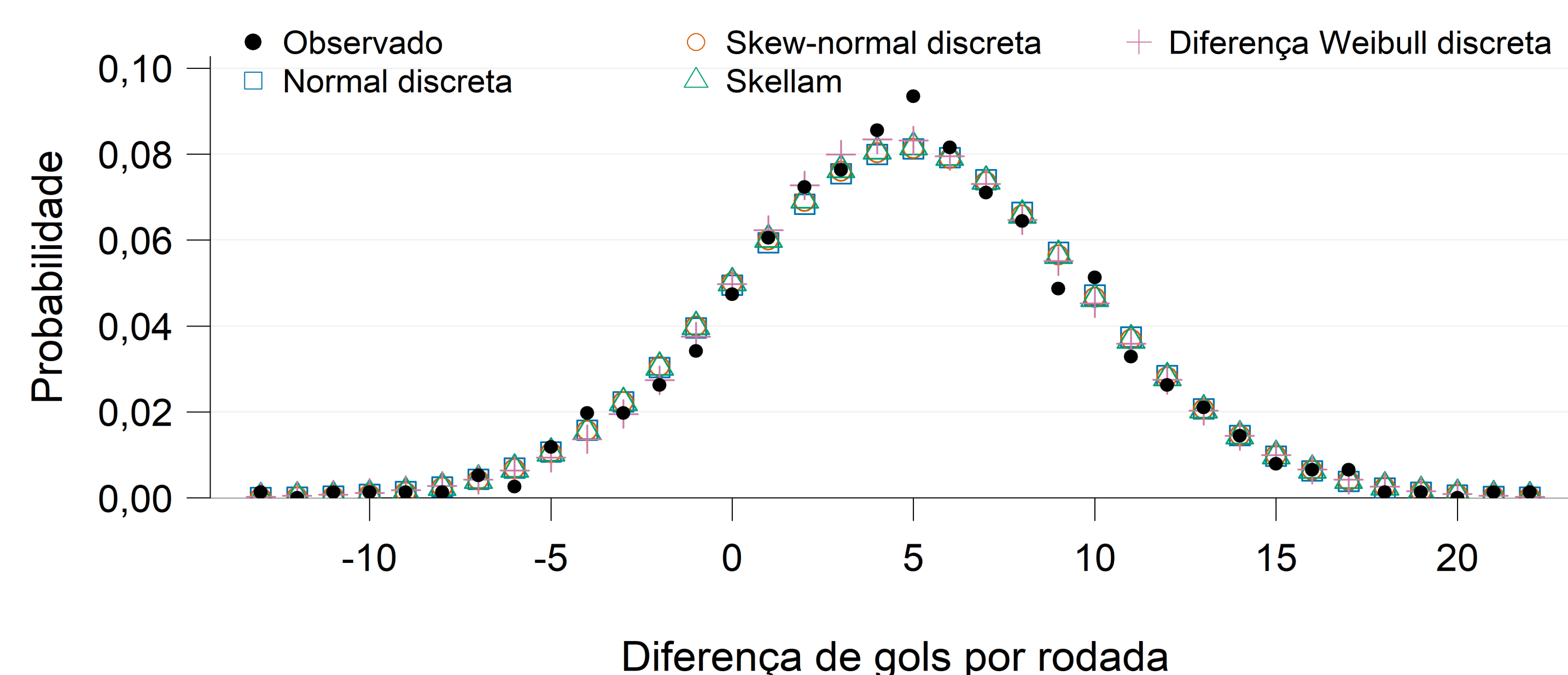
Tabela 2: Estimativa dos parâmetros ajustados.

Normal Discreta	$\hat{\mu} = 5,93$	$\hat{\sigma} = 1,59$	---	---
Skew-normal discreta	$\hat{\mu} = 3,24$	$\hat{\sigma} = 5,35$	$\hat{\lambda} = 0,58$	---
Skellam	$\hat{\lambda}_1 = 2,67$	$\hat{\lambda}_2 = 2,26$	---	---
Diferença Weibull discreta	$\hat{q}_1 = 5,01$	$\hat{\beta}_1 = 0,8$	$\hat{q}_2 = 1,57$	$\hat{\beta}_1 = 0,22$

Tabela 3: Critérios de seleção dos modelos ajustados.

Modelo	AIC	BIC
Skellam	4578,40	4587,66
Normal discreta	4579,16	4588,43
Diferença Weibull discreta	4580,07	4598,60
Skew-normal discreta	4581,01	4594,91

Figura 1: Frequências observadas e probabilidades ajustadas pelos modelos probabilísticos



CONCLUSÃO

Entre os modelos avaliados, a Skellam apresentou os menores valores dos critérios de informação, com pequena diferença em relação a normal discreta. Os resultados favorecem sua utilização para descrever a diferença de gols por rodada no período analisado.

REFERÊNCIAS BIBLIOGRÁFICAS

DUQUE, A. *Brasileirão Dataset: Campeonato Brasileiro de Futebol*. 2025. Arquivo campeonato-brasileiro-full.csv. Acesso em: 07 ago. 2026.

AZZALINI, A. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, [Board of the Foundation of the *Scandinavian Journal of Statistics*, Wiley], v. 12, n. 2, p. 171–178, 1985. ISSN 03036898, 14679469. Disponível em: .

SKELLAM, J. G. The frequency distribution of the difference between two poisson variates belonging to different populations. *Journal of the Royal Statistical Society*, v. 109, n. 3, p. 296–296, 1946. Disponível em: .

NAKAGAWA, T.; OSAKI, S. The discrete weibull distribution. *IEEE Transactions on Reliability*, R-24, n. 5, p. 300–301, 1975.

AGRADECIMENTOS

