



UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Exatas  
Departamento de Estatística

Luiz Felipe Avelar Lima

## **Avaliação da performance de modelos bayesianos para dados de campeonatos de futebol**

Belo Horizonte  
Julho de 2023



Luiz Felipe Avelar Lima

## **Avaliação da performance de modelos bayesianos para dados de campeonatos de futebol**

Trabalho apresentado à banca examinadora  
da UFMG, como requisito para a obtenção  
do título de bacharel em Estatística.

Orientador: Cristiano de Carvalho Santos

Belo Horizonte  
Julho de 2023

# Avaliação da performance de modelos bayesianos para dados de campeonatos de futebol

## Resumo

Com o desenvolvimento da estatística computacional, a análise preditiva tem ganhado destaque por meio do uso de modelos capazes de determinar o resultado de partidas esportivas e a classificação de times em um campeonato a partir de fatores como resultados históricos, equipes mandantes e visitantes, números de gols ou pontos marcados e eficiência defensivas e ofensivas de cada equipe. Na literatura de modelos mistos, pode-se destacar Karlis e Ntzoufras (2003) que propuseram um modelo assumindo a distribuição Poisson bivariada para modelar os números de gols marcados pelas equipes mandante e visitante em uma partida de futebol onde é considerada a associação entre o número de gols marcados pelas equipes por meio da introdução de efeitos aleatórios. Neste artigo, os autores utilizaram a abordagem clássica através do algoritmo EM para obter as estimativas de máxima verossimilhança. Ainda dentro de modelos mistos, mas assumindo a abordagem bayesiana, existem modelagens como a proposta por Baio e Blangiardo (2010) para dados de futebol, em que o modelo hierárquico assume duas variáveis Poisson condicionalmente independentes para modelar o número de gols marcados por cada equipe em uma partida. Ao mesmo tempo em que efeitos aleatórios representando as capacidades defensivas e ofensivas dos times são responsáveis por introduzir a associação entre os resultados obtidos por um mesmo time.

Neste trabalho são ajustados os modelos bayesianos hierárquicos propostos por Baio e Blangiardo (2010), Poisson misto e Poisson misto com misturas para os efeitos aleatórios para os dados do Campeonato Brasileiro de Futebol (Brasileirão) dos anos de 2020 a 2022, extraídos da plataforma *Kaggle*. Estes modelos também foram estendidos assumindo a distribuição Binomial Negativa e um modelo zero inflado para o número de gols marcados por cada time em cada partida. Os quatro modelos foram implementados em linguagem de programação Stan e Jags através dos pacotes RStan (Stan Development Team, 2022) e RJags (Martyn Plummer, 2022) para o software R (R Core Team, 2022) e a avaliação da qualidade da implementação entre softwares é feita a partir de critérios como tamanho efetivo de amostra e graus de convergência das cadeias de Markov. Já a avaliação de performance entre modelos é feita a partir de critérios como a estimativa pontual e intervalar da pontuação de cada time ao longo da competição e dos gols marcados por partida, erro quadrático médio e viés dessas estimativas, assim como o critério de seleção de modelos *Leave-one-out cross-validation* (LOO-CV) proposto por Vehtari A, Gelman A, Gabry J (2017).

Os resultados obtidos indicam que os quatro modelos implementados são muito similares com base nos critérios de comparação escolhidos, dentre eles o modelo Poisson misto com mistura apresenta resultados relativamente melhores. Tratando-se da comparação entre softwares, o Stan apresenta resultados melhores em termos de tamanho efetivo de amostra, enquanto que em termos de graus de convergência das cadeias de Markov ambos os softwares apresentam resultados muito próximos.



# Sumário

<b>1. Introdução</b>	<b>5</b>
<b>2. Metodologia</b>	<b>6</b>
2.1. Modelo Poisson misto	6
2.2. Modelo Poisson misto com mistura	8
2.3. Modelo Binomial Negativa	9
2.4. Modelo Poisson misto zero inflado	10
<b>3. Aplicação a dados reais</b>	<b>10</b>
3.1. Comparação entre modelos	11
3.1.1. Ajuste e critérios	11
3.1.2. Brasileirão 2020	12
3.1.3. Brasileirão 2021	19
3.1.4. Brasileirão 2022	23
3.2. Comparação entre softwares	24
<b>4. Considerações finais</b>	<b>28</b>
<b>5. Referências Bibliográficas</b>	<b>29</b>
<b>6. Apêndice</b>	<b>31</b>

## 1. Introdução

Com o desenvolvimento da estatística computacional, a análise preditiva tem ganhado destaque por meio do uso de modelos capazes de determinar o resultado de partidas esportivas e a classificação de times em um campeonato a partir de fatores como resultados históricos, equipes mandantes e visitantes, números de gols ou pontos marcados e eficiência defensivas e ofensivas de cada equipe. Na literatura mais recente, encontram-se exemplos como o de Prasetio e Harlili (2016) que desenvolveram um modelo de regressão logística para prever os resultados do Campeonato Inglês de Futebol (Premier League) para a temporada 2015/2016 a partir de variáveis que mensuram o poderio defensivo e ofensivo dos times jogando como mandante ou como visitante. Também encontramos o trabalho de Manish, Bhagat e Pramila (2021) que faz uso de algoritmos de machine learning e deep learning para prever a performance de jogadores baseado em sua posição de atuação.

Na literatura de modelos mistos, pode-se destacar Karlis e Ntzoufras (2003) que propuseram um modelo assumindo a distribuição Poisson bivariada para modelar os números de gols marcados pelas equipes mandante e visitante em uma partida de futebol onde é considerada a associação entre o número de gols marcados pelas equipes por meio da introdução de efeitos aleatórios. Neste artigo, os autores utilizaram a abordagem clássica através do algoritmo EM para obter as estimativas de máxima verossimilhança. Também há variações para outros esportes, como o vôlei, em que os modelos podem ser expandidos levando em consideração as características intrínsecas do esporte, e.g. quantidade de sets e o número total de pontos marcados por cada time. Gabrio (2021) propõe um modelo composto por três módulos: o número total de pontos são modelados por variáveis Poisson, enquanto que variáveis binárias indicativas de que o jogo teve cinco sets ou o time mandante foi vencedor são modelados através de regressões logísticas. Ainda dentro de modelos mistos, mas assumindo a abordagem bayesiana, existem modelagens como a proposta por Baio e Blangiardo (2010) para dados de futebol, em que o modelo hierárquico assume duas variáveis Poisson condicionalmente independentes para modelar o número de gols marcados por cada equipe em uma partida. Ao mesmo tempo em que efeitos aleatórios representando as capacidades defensivas e ofensivas dos times são responsáveis por introduzir a associação entre os resultados obtidos por um mesmo time.

Neste trabalho são ajustados os modelos bayesianos hierárquicos propostos por Baio e Blangiardo (2010), Poisson misto e Poisson misto com misturas para os efeitos aleatórios para os dados do Campeonato Brasileiro de Futebol (Brasileirão) dos anos de 2020 a 2022, extraídos da plataforma *Kaggle*. Estes modelos também foram estendidos assumindo a distribuição Binomial Negativa e um modelo zero inflado para o número de gols marcados por cada time em cada partida. Tais extensões permitem avaliar como a distribuição Binomial Negativa acomoda uma variabilidade maior dos dados em comparação a distribuição Poisson e como ajustar um conjunto de dados com partidas em que há um excesso de partidas em que algum dos times não marcou gols. Os quatro modelos foram implementados em linguagem de programação Stan e Jags através dos pacotes RStan (Stan Development Team, 2022) e RJags (Martyn Plummer, 2022) para o software R (para o software R (R Core Team, 2022) e a avaliação da qualidade da implementação entre softwares é feita a partir de critérios como tamanho efetivo de amostra e graus de convergência das cadeias de Markov. Já a avaliação de performance entre

modelos é feita a partir de critérios como a estimativa pontual e intervalar da pontuação de cada time ao longo da competição e dos gols marcados por partida, erro quadrático médio e viés dessas estimativas, assim como o critério de seleção de modelos *Leave-one-out cross-validation* (LOO-CV) proposto por Vehtari A, Gelman A, Gabry J (2017).

## 2. Metodologia

Nesta seção serão apresentados os modelos mistos propostos por Baio e Blangiardo (2010), modelos Poisson misto e Poisson misto com mistura, assim como as suas possíveis extensões, sendo elas os modelos Binomial Negativa misto e Poisson misto zero inflado.

### 2.1. Modelo Poisson misto

O modelo Poisson misto proposto por Baio e Blangiardo (2010) assume que para o número de gols  $y_{g1}$  e  $y_{g2}$  marcados pelas equipes mandante e visitante em um jogo  $g$  do campeonato de futebol é modelado pela distribuição Poisson tal que  $y_{gj}|\theta_{gj} \sim \text{Poisson}(\theta_{gj})$ ,  $j = 1$  ou  $2$  e  $g = 1, \dots, G$ , em que  $\theta_g = (\theta_{g1}, \theta_{g2})$  representa a taxa de ocorrência de gols no jogo para os dois times ( $j = 1$  para o mandante e  $j = 2$  para o visitante) e  $G$  representa o total de jogos do campeonato.

Neste modelo, as taxas de ocorrência de gols das duas equipes no jogo  $g$  são modeladas através de função de ligação logarítmica tal que:

$$\log \theta_{g1} = \text{home} + \text{att}_{h(g)} + \text{def}_{a(g)},$$

$$\log \theta_{g2} = \text{att}_{a(g)} + \text{def}_{h(g)},$$

em que *home* é um efeito fixo que quantifica a vantagem de se disputar uma partida em seu mando de campo, enquanto que os parâmetros *att* e *def* são efeitos aleatórios que quantificam as capacidades ofensivas e defensivas, respectivamente, dos times mandante (índice  $h$ ) e visitante (índice  $a$ ) na  $g$ -ésima partida.

Para os efeitos aleatórios de ataque e defesa, para cada time  $t$ , são especificadas distribuições normais:

$$\text{att}_t \sim \text{Normal}(\mu_{\text{att}}, \sigma_{\text{att}}),$$

$$\text{def}_t \sim \text{Normal}(\mu_{\text{def}}, \sigma_{\text{def}}),$$

em que  $\sigma_{\text{att}}$  e  $\sigma_{\text{def}}$  representam o desvio padrão da distribuição. Para garantir que o modelo seja identificável, também é necessário impor alguma restrição. Uma

possibilidade considerada por Baio e Blangiardo (2010) é assumir  $\sum_{t=1}^T att_t = 0$  e  $\sum_{t=1}^T def_t = 0$ . Outra possibilidade, assumida neste trabalho por resultar em melhor convergência do modelo, é assumir que  $att_{20} = 0$  e  $def_{20} = 0$ .

Tratando-se da abordagem Bayesiana, para completar a especificação do modelo, é necessário definir as distribuições a priori para os parâmetros do modelo. Para o parâmetro de efeito fixo *home* assume-se  $home \sim Normal(0, 10)$  em que o desvio-padrão é pré-especificado.

Por fim, são especificadas distribuições a prioris para os demais hiperparâmetros, isto é, assumimos  $\mu_{att} \sim Normal(0, 10)$ ,  $\mu_{def} \sim Normal(0, 10)$ ,  $\sigma_{att} \sim Cauchy(0; 2, 5)$  e  $\sigma_{def} \sim Cauchy(0; 2, 5)$ . Vale ressaltar que Baio e Blangiardo (2010) sugerem usar a distribuição a priori Gama inversa para  $\sigma^2$  ao assumir uma distribuição gama para a precisão, entretanto, no Stan foram utilizadas as distribuições a priori Cauchy para  $\sigma$  como recomendado em Gelman et al. (2008) para distribuições a priori de parâmetros de desvio padrão.

## 2.2. Modelo Poisson misto com mistura

Baio e Blangiardo (2010) também propõem um modelo mais complexo de Poisson misto com mistura de distribuições para os efeitos aleatórios a fim de evitar o problema de encolhimento (overshrinkage), que ocorre quando o modelo hierárquico bayesiano tende a subestimar a variabilidade entre as equipes e produzir estimativas muito próximas da média. Por exemplo, se em um campeonato há três equipes disputando o título e pontuando bem acima da média e três equipes lutando contra o rebaixamento e pontuando bem abaixo da média, o modelo Poisson misto subestima a grande variabilidade que essas equipes apresentam em comparação com o restante dos times, subestimando a pontuação dos melhores times e superestimando o pontuação dos piores times.

Desta forma, o modelo mais sofisticado proposto pelos autores divide os times em três subgrupos, aqueles que possuem propensão a marcar gols acima da média e propensão a conceder gols abaixo da média, aqueles que possuem propensão a marcar gols abaixo da média e propensão a conceder gols acima da



média, aqueles que possuem propensão a marcar e conceder gols próximas a média, atribuindo uma distribuição de probabilidade diferente para os efeitos aleatórios a cada um destes subgrupos. Para isso define-se as seguintes misturas para os efeitos aleatórios:

$$att_t = \sum_{k=1}^3 \pi_{kt}^{att} \times Normal(\mu_k^{att}, \tau_k^{att}), \quad def_t = \sum_{k=1}^3 \pi_{kt}^{def} \times Normal(\mu_k^{def}, \tau_k^{def}),$$

sendo  $k$  o índice que indica a qual dos três subgrupos cada time pertence e  $\pi^{att}$ ,  $\pi^{def}$ , respectivamente, os vetores de parâmetros de probabilidade cada time pertencer a cada um dos subgrupos de ataque e de defesa, de forma que assumimos distribuições a priori  $\pi^{att} \sim Dirichlet(1, 1, 1)$  e  $\pi^{def} \sim Dirichlet(1, 1, 1)$ .

Agora, os hiperparâmetros de média de cada subgrupo é representado por meio de distribuições a priori truncadas. Para os times que têm alta propensão a concederem gols e baixa propensão a marcarem gols assumimos:

$$\mu_1^{att} \sim truncNormal(0, 10, -3, 0) \text{ e } \mu_1^{def} \sim truncNormal(0, 10, 0, 3).$$

Para os times que têm alta propensão a marcarem gols e baixa propensão a concederem gols:

$$\mu_3^{att} \sim truncNormal(0, 10, 0, 3) \text{ e } \mu_3^{def} \sim truncNormal(0, 10, -3, 0).$$

E as distribuições a priori atribuídas aos hiperparâmetros  $\sigma_k^{att}$  e  $\sigma_k^{def}$ , para  $k = 1, 3$ , continuam sendo uma  $Cauchy(0; 2, 5)$ .

Para times medianos, tanto para ataque quanto para defesa, assumimos que  $\mu_2^{att} = \mu_2^{def} = 0$  e  $\sigma_2^{att} = \sigma_2^{def} = 0,01$ . Tais especificações visam definir uma massa pontual no valor zero para os efeitos aleatórios de times com essas categorizações.

O modelo descrito acima diferencia-se daquele proposto por Baio e Blangiardo (2010) no que tange o uso da distribuição Normal para estimar os efeitos aleatórios. Os autores propõem, em seu lugar, o uso da distribuição t-student

não-central com quatro graus de liberdade, porém o seu uso foi preterido a fim de garantir a convergência do modelo.

### 2.3. Modelo Binomial Negativa

Lawless (1987) investiga o uso da distribuição Binomial Negativa em análises de regressão visto que esta distribuição acomoda uma maior variabilidade dos dados em comparação a distribuição Poisson para cenários de contagem.

Tendo em vista este comportamento da distribuição, uma extensão do modelo descrito na seção anterior é o modelo Binomial Negativa misto que possui a mesma estrutura do modelo Poisson misto exceto na distribuição assumida para o número de gols marcados por cada equipe em um jogo. Neste caso, serão estimados como variáveis Binomiais Negativas independentemente distribuídas:

$$y_{gj} | \mu_{gj} \sim NB(\mu_{gj}, \phi).$$

Trata-se de uma parametrização alternativa para a distribuição Binomial Negativa, por meio de um parâmetro de localização (média) e um parâmetro que controla a superdispersão dos dados ( $\phi$ ) em relação ao quadrado da média. Sendo ao hiperparâmetro  $\phi$  atribuído a distribuição a priori  $\phi \sim Cauchy(0; 2, 5)$ .

### 2.4. Modelo Poisson misto zero inflado

Lambert (1992) propõe um modelo Poisson para a contagem de dados com excesso de zeros, em que com probabilidade  $p$  o valor observado é zero e com probabilidade  $1 - p$  o valor observado é proveniente de uma variável Poisson aleatória.

Uma segunda extensão ao modelo descrito na Seção 2.1 é o modelo Poisson misto zero inflado em que possui a mesma estrutura do modelo Poisson misto adicionando a contagem de dados com excesso de zeros, de forma que:

$$\begin{cases} y_{gj} \sim 0 \text{ com probabilidade } p_j \\ y_{gj} | \theta_{gj} \sim Poisson(\theta_{gj}) \text{ com probabilidade } 1 - p_j \end{cases}$$

Assim como os outros parâmetros dos modelos bayesianos, também é necessário adicionar distribuições a priori aos parâmetros  $p_j \sim \text{Beta}(0, 1; 0, 4)$ . Desta forma o valor esperado para  $p_j$  é a esperança da distribuição Beta,  $E[p_j] = \frac{0, 1}{0, 1 + 0, 4} = 0, 2$ .

### 3. Aplicação a dados reais

Para o ajuste dos modelos a dados reais foi utilizado o dataset “Campeonato Brasileiro de Futebol” disponibilizado na plataforma *Kaggle* pelo usuário Adão Duque. A partir do dataset original, os dados foram filtrados para cada uma das edições de interesse (2020, 2021 e 2022), selecionadas as colunas pertinentes aos modelos (rodada, placar e times envolvidos) e adicionadas colunas de identificadores únicos (com sufixo id) a cada time a fim de seguir a modelagem estabelecida na Seção 2. Além disso, foi necessário fornecer alguns parâmetros referentes à estrutura do campeonato, sendo eles  $T = 20$  e  $G = 380$ , ou seja, o campeonato é disputado por 20 times em 380 jogos, todos os times se enfrentam duas vezes, uma como mandante e outra como visitante.

Tabela 1 - Amostra do conjunto de dados

round	home_team_name	away_team_name	home_team_score	away_team_score	home_team_id	away_team_id
1	Botafogo-RJ	Bahia	1	2	5	4
1	Corinthians	Atletico-GO	0	0	8	2
1	Coritiba	Internacional	0	1	9	15
1	Flamengo	Atletico-MG	0	1	10	3
1	Fortaleza	Athletico-PR	0	2	12	1

#### 3.1. Comparação entre modelos

##### 3.1.1. Ajuste e critérios

A comparação de performance entre os modelos descritos na Seção 2 foi feita através da implementação no software Stan. Os modelos foram ajustados com uma única cadeia de Markov com 5000 iterações para essa cadeia e período de aquecimento (*burn-in*) igual a 2500. Sendo assim, o Stan amostrou 2500 observações simuladas a partir da distribuição a posteriori de cada um dos

parâmetros do modelo e da distribuição preditiva a posteriori do placar de cada um dos jogos.

Os modelos descritos são definidos considerando o número de gols marcados por cada equipe em cada jogo, mas, por se tratar de um campeonato de pontos corridos, fez-se necessário criar novas variáveis seguindo as regras de pontuação estabelecidas na competição: Vitória - 3 pontos; Derrota - 0 ponto e Empate - 1 ponto a fim de computar a pontuação estimada das equipes ao longo da disputa.

Dado os 2500 resultados amostrados da distribuição a posteriori, foram computadas as pontuações acumuladas ao longo do campeonato para cada time em cada um dos cenários simulados de modo que a mediana a posteriori foi definida como a estimativa pontual e a estimativa intervalar foi definida como o intervalo de credibilidade 95%.

Os critérios de comparação Erro Quadrático Médio (EQM) e Viés foram calculados considerando tanto a pontuação quanto os gols marcados por partida. Para o primeiro cenário, o cálculo foi feito de acordo com as equações

$$EQM_{pontos} = \frac{1}{T} \sum_{t=1}^T (P_{38t} - \hat{P}_{38t})^2, \quad Viés_{pontos} = \frac{1}{T} \sum_{t=1}^T P_{38t} - \hat{P}_{38t},$$

em que,  $T$  denota os 20 times participantes do campeonato,  $P_{38t}$  denota a pontuação observada na última rodada da competição (trigésima-oitava) pelo  $t$ -ésimo time e  $\hat{P}_{38t}$  denota a pontuação estimada nesta mesma rodada para este mesmo time. Para o segundo cenário, o cálculo foi feito de acordo com as equações

$$EQM_{gols} = \frac{1}{2G} \sum_{g=1}^G (Y_{g1} - \hat{Y}_{g1})^2 + (Y_{g2} - \hat{Y}_{g2})^2,$$

$$Viés_{gols} = \frac{1}{2G} \sum_{g=1}^G (Y_{g1} - \hat{Y}_{g1}) + (Y_{g2} - \hat{Y}_{g2}),$$

em que,  $G$  denota as 380 partidas do campeonato,  $Y_{g1}$  denota o número observado de gols marcados pelo time mandante na  $g$ -ésima partida e  $Y_{g2}$  o número observado de gols marcados pelo visitante na respectiva partida,  $\hat{Y}_{g1}$  denota o número de gols estimado para o time mandante na  $g$ -ésima partida e  $\hat{Y}_{g2}$  o número de gols estimado para o time visitante na respectiva partida.

Para o cálculo do critério de seleção de modelos *leave-one-out cross-validation* (LOO-CV), foi utilizado o pacote *loo* em linguagem de programação R que implementa o método proposto por Vehtari A, Gelman A, Gabry J (2017):

$$elpd_{loo} = \sum_{i=1}^n \log p(y_i | y_{-i})$$
, onde  $p(y_i | y_{-i}) = \int p(y_i | \theta) p(\theta | y_{-i}) d\theta$  é a densidade preditiva *leave-one-out* dos dados sem a  $i$ -ésima observação, ou seja, sem a observação da qual está sendo feita a predição. Este método leva em consideração o valor esperado do logaritmo da densidade preditiva do modelo para o cálculo do critério de seleção. Calculado o critério  $elpd_{loo}$ , é possível obter o critério de seleção  $looic = -2 \times elpd_{loo}$ . Deste modo, maiores valores de  $elpd_{loo}$  e menores valores de  $looic$  indicam os melhores modelos.

### 3.1.2. Brasileiro 2020

Pode ser observado na Tabela 2, abaixo, que para os dados do Campeonato Brasileiro de 2020 os modelos Poisson misto e Poisson misto com mistura foram aqueles que apresentaram melhor ajuste. O primeiro apresentou menor viés para pontuação e menor  $looic$ , enquanto o segundo apresentou menores EQM tanto para pontuação quanto para viés e menor EQM para gols marcados. É importante frisar que apesar do modelo Poisson misto ter apresentado melhor ajuste segundo o critério LOO-CV, todos os quatro modelos apresentaram valores de  $looic$  muito próximos e com diferenças menores que o erro padrão da medida. Segundo a documentação do pacote *loo*, é recomendável considerar diferenças maiores que duas vezes o erro padrão como sendo significativas, o que não ocorre neste caso.

Tabela 2 - Comparação entre modelos para a edição de 2020 do Brasileiro

Modelo	Pontuação		Gols		LOO - Estimação		LOO - Erro padrão	
	EQM	Viés	EQM	Viés	elpd_loo	looic	elpd_loo	looic
Poisson Misto	42,75	-0,25	2,07	0,93	-1060,00	2120,00	17,80	35,50
Binomial Negativa Misto	42,00	-0,40	0,94	-0,06	-1061,00	2122,10	17,60	35,30
Poisson Misto Zero Inflado	44,55	-0,35	2,69	0,44	-1060,60	2121,20	17,60	35,30
Poisson Misto com Mistura	35,00	-0,30	0,92	-0,07	-1062,20	2124,40	17,80	35,70

Aprofundando um pouco mais na análise dos modelos que obtiveram as melhores estimações, podemos observar, na Tabela 3, que os efeitos aleatórios de ataque estimados pelos modelos distaram bastante em valores absolutos. O Bahia, por exemplo, possui um efeito aleatório de ataque mediano de 0,25 segundo o modelo Poisson misto enquanto que, segundo o modelo Poisson misto com mistura, este efeito seria igual a zero. Comparando este valor com o time de referência (Vasco), isso significa que, segundo o modelo Poisson misto, o Bahia tem uma propensão a marcar gols de  $\exp(0,25) = 1,28$  vezes maior do que a propensão do Vasco em marcar caso ambos os times enfrentassem o mesmo adversário em seus respectivos mandos de campo. Enquanto que, segundo o modelo Poisson misto com mistura, essa propensão seria  $\exp(0) = 1$ , ou seja, igual a do Vasco.

Tabela 3 - Estimativas a posteriori dos parâmetros de efeitos aleatórios de ataque para a edição de 2020 do Brasileirão

Time	Modelo Poisson misto			Modelo Poisson misto com mistura		
	Mediana	2,5%	97.5%	Mediana	2,5%	97.5%
Athletico-PR	0,09	-0,23	0,40	-0,01	-0,40	0,14
Athletico-GO	0,13	-0,18	0,44	-0,01	-0,44	0,19
Athletico-MG	0,44	0,13	0,74	0,32	-0,00	0,55
Bahia	0,25	-0,07	0,54	0,00	-0,24	0,35
Botafogo-RJ	0,01	-0,35	0,33	-0,17	-0,54	0,02
Bragantino	0,26	-0,05	0,55	0,01	-0,09	0,39
Ceara	0,32	0,02	0,61	0,18	-0,03	0,43
Corinthians	0,20	-0,12	0,50	0,00	-0,19	0,30
Coritiba	-0,01	-0,35	0,31	-0,17	-0,58	0,02
Flamengo	0,49	0,18	0,78	0,34	0,07	0,59
Fluminense	0,33	0,03	0,61	0,20	-0,03	0,45
Fortaleza	0,04	-0,31	0,35	-0,12	-0,48	0,07
Goiás	0,15	-0,17	0,44	-0,00	-0,32	0,26
Gremio	0,30	-0,02	0,59	0,02	-0,04	0,41
Internacional	0,40	0,11	0,70	0,27	-0,01	0,51
Palmeiras	0,28	-0,02	0,57	0,01	-0,08	0,37
Santos	0,29	-0,01	0,58	0,09	-0,09	0,41
Sao Paulo	0,38	0,08	0,67	0,25	-0,01	0,49
Sport	-0,00	-0,36	0,31	-0,15	-0,56	0,02
Vasco	0,00	0,00	0,00	0,00	0,00	0,00

Agora, analisando os efeitos aleatórios de defesa na Tabela 4, podemos observar um comportamento similar aos dos efeitos aleatórios de ataque. O Ceará, por exemplo, que é o time com maior diferença absoluta para os efeitos entre os dois modelos, possui um efeito aleatório de defesa mediano de -0,15 segundo o modelo Poisson misto enquanto que, segundo o modelo Poisson misto com mistura, este efeito seria igual a 0,29. Comparando este valor com o time de referência (Vasco), isso significa que, segundo o modelo Poisson misto, o Ceará tem uma propensão a conceder gols de  $\exp(-0,15) = 0,86$  vezes a propensão do Vasco em conceder gols caso ambos os times enfrentassem o mesmo adversário em seus respectivos mandos de campo. Enquanto que, segundo o modelo Poisson misto com mistura, essa propensão seria  $\exp(0,29) = 1,33$  vezes a do Vasco, ou seja, para o primeiro modelo o Ceará tem uma pior defesa do que o Vasco e no segundo modelo, o contrário.

Cabe destacar que, em virtude da modelagem construída, é possível notar que os efeitos aleatórios do modelo Poisson misto com misturas apresentam claramente três clusters de valores em que os efeitos podem estar concentrados (acima de zero, próximos a zero ou abaixo de zero), enquanto que a distribuição dos efeitos para o modelo Poisson misto se dá de forma mais homogênea.

Tabela 4 - Estimativas a posteriori dos parâmetros de efeitos aleatórios de defesa para a edição de 2020 do Brasileirão

Time	Modelo Poisson misto			Modelo Poisson misto com mistura		
	Mediana	2,5%	97.5%	Mediana	2,5%	97.5%
Athletico-PR	-0,27	-0,55	0,01	-0,00	-0,31	0,20
Athletico-GO	-0,19	-0,46	0,08	0,01	-0,15	0,38
Athletico-MG	-0,19	-0,46	0,08	0,22	-0,02	0,49
Bahia	-0,09	-0,34	0,19	-0,01	-0,36	0,12
Botafogo-RJ	-0,07	-0,33	0,22	0,29	-0,01	0,53
Bragantino	-0,23	-0,49	0,04	-0,00	-0,26	0,24
Ceara	-0,15	-0,40	0,13	0,29	-0,01	0,52
Corinthians	-0,19	-0,46	0,08	-0,00	-0,29	0,19
Coritiba	-0,13	-0,38	0,14	-0,01	-0,41	0,13
Flamengo	-0,16	-0,43	0,11	0,29	-0,01	0,53
Fluminense	-0,21	-0,48	0,06	-0,00	-0,27	0,21
Fortaleza	-0,21	-0,47	0,05	-0,00	-0,33	0,15
Goiás	-0,06	-0,32	0,23	-0,00	-0,27	0,23
Gremio	-0,23	-0,51	0,05	-0,00	-0,33	0,17
Internacional	-0,26	-0,56	0,00	0,10	-0,07	0,44
Palmeiras	-0,25	-0,54	0,01	-0,15	-0,56	0,05
Santos	-0,15	-0,41	0,12	0,38	0,10	0,64
Sao Paulo	-0,22	-0,48	0,05	-0,23	-0,68	0,01
Sport	-0,16	-0,41	0,11	-0,00	-0,28	0,20
Vasco	0,00	0,00	0,00	0,00	0,00	0,00

Também podem ser analisadas as estimativas a posteriori para os parâmetros fixos dos modelos, como mostrado nas tabelas abaixo.



Tabela 5 - Estimativas a posteriori dos parâmetros para o Modelo Poisson misto para a edição de 2020 do Brasileirão

Parâmetro	Mediana	2,5%	97,5%
home	0,27	0,14	0,40
mu_att	0,23	-0,00	0,44
mu_def	-0,18	-0,38	0,04
sigma_att	0,20	0,11	0,34
sigma_def	0,11	0,04	0,22

Tabela 6 - Estimativas a posteriori dos parâmetros para o modelo Poisson misto com mistura para a edição de 2020 do Brasileirão

Parâmetro	Mediana	2,5%	97,5%
home	0,28	0,12	0,40
mu_att1	-0,16	-1,14	-0,01
mu_att3	0,27	0,03	0,48
mu_def1	0,14	0,01	2,26
mu_def3	-0,16	-1,56	-0,01
sigma_att1	0,21	0,01	2,14
sigma_att3	0,11	0,01	0,64
sigma_def1	0,18	0,01	6,35
sigma_def3	0,14	0,01	4,43

É possível notar que para ambos os modelos o efeito fixo de jogar uma partida em seu mando de campo (*home*) é positivo, indicando que times que jogam em casa possuem uma taxa de ocorrência de gols maior. Ademais, para o modelo Poisson misto com mistura há uma distinção clara entre os valores estimados a posteriori para os parâmetros  $\mu$ , atribuindo taxas de ocorrência de marcar ou conceder gols diferentes a cada subgrupo de equipes. As estimativas a posteriori dos parâmetros dos demais modelos encontram-se no Apêndice.

Por fim, pode-se analisar graficamente como ficaria a pontuação acumulada de cada time ao longo do campeonato estimada por cada modelo.

Figura 1 - Pontuação estimada pelo modelo Poisson misto para a edição de 2020 do Brasileiro

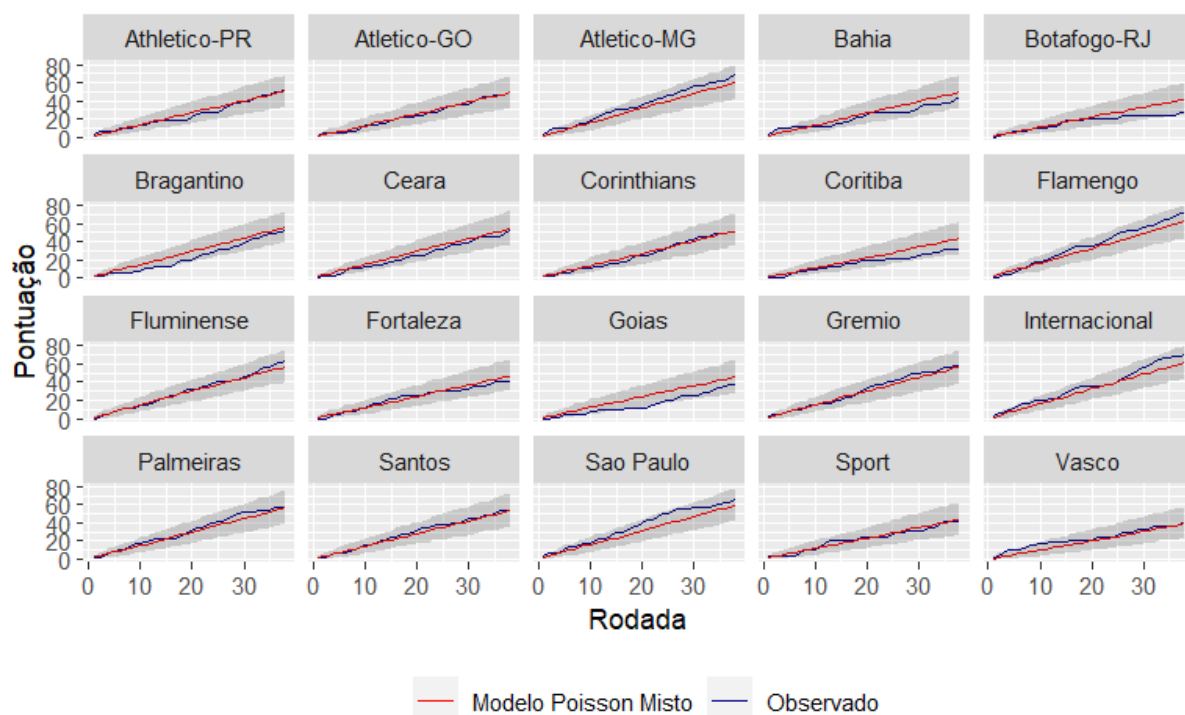
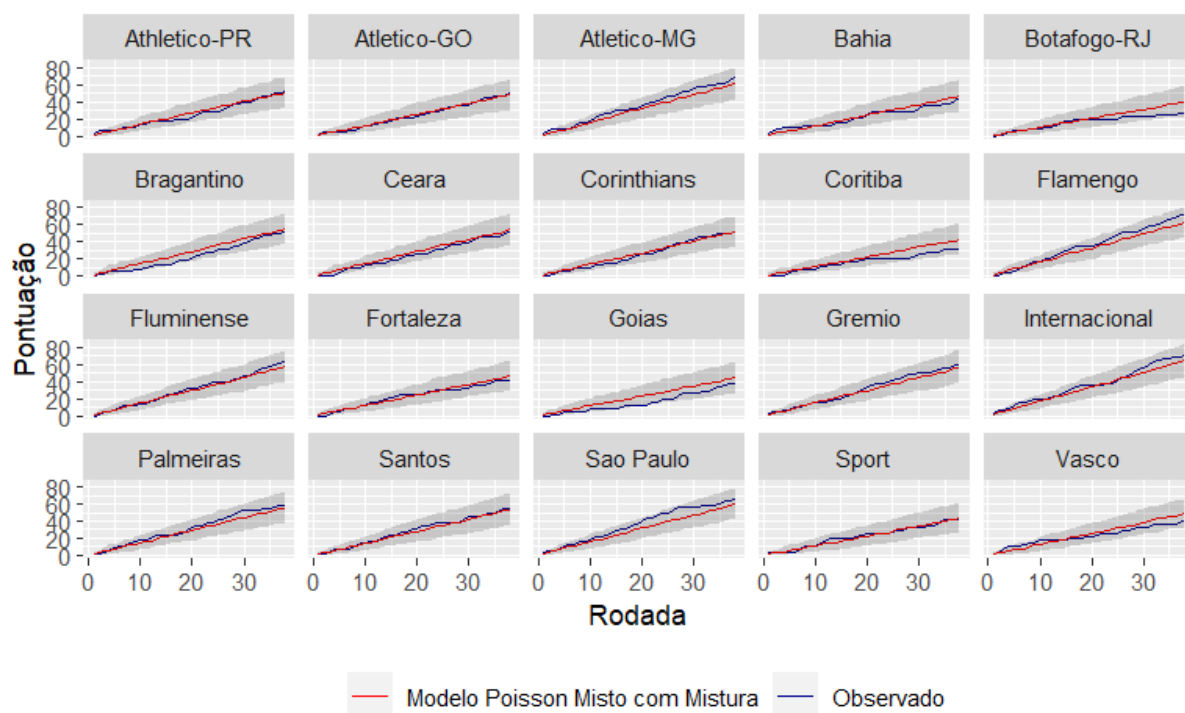


Figura 2 - Pontuação estimada pelo modelo Poisson misto com mistura para a edição de 2020 do Brasileiro



Podemos observar que para ambos os modelos, as pontuações observadas encontram-se dentro do intervalo de credibilidade estimado (área sombreada) e que apesar dos valores absolutos dos efeitos aleatórios de ataque e defesa distarem entre os modelos, a sua combinação equaciona em estimativas semelhantes. Segundo o modelo Poisson misto, os quatro primeiros colocados e, portanto, aqueles times classificados à Copa Libertadores, foram Flamengo, Internacional, Atlético-MG e São Paulo, o que coincide com os times que realmente terminaram a competição nessa classificação. Olhando para os quatro últimos classificados, aqueles times rebaixados à Série B, o modelo acertou três deles: Vasco, Botafogo e Coritiba, entretanto o modelo estima que Sport seria rebaixado no lugar do Goiás. Já o modelo Poisson misto com mistura também acerta os quatro primeiros colocados e três das quatro equipes rebaixadas, estimando que o Vasco se livrou do rebaixamento e o Sport seria rebaixado.

### **3.1.3. Brasileirão 2021**

Tendo a Tabela 7 como referência para avaliar a performance dos modelos em relação ao ajuste dos dados do Campeonato Brasileiro de 2021, podemos observar que, desta vez, cada um dos modelos apresentou melhor performance segundo algum dos critérios, não sendo possível destacar com clareza qual o melhor modelo para este conjunto de dados. O modelo Poisson misto apresenta menor EQM para gols, o modelo Binomial Negativa apresenta menor viés tanto para pontuação quanto para gols, o modelo Poisson misto zero inflado apresenta menor viés para pontuação (empatado com o modelo anterior) e o modelo Poisson misto com mistura apresenta menor EQM para pontuação e menor *looic*. Novamente, é importante frisar que todos os modelos possuem qualidade de ajuste muito próximas segundo o critério *looic*, com diferenças menores que o erro padrão da estimativa.

Tabela 7 - Comparação entre modelos para a edição de 2021 do Brasileirão

Modelo	Pontuação		Gols		LOO - Estimção		LOO - Erro padrão	
	EQM	Viés	EQM	Viés	elpd_loo	looic	elpd_loo	looic
Poisson Misto	61,60	-0,30	0,65	-0,55	-1017,40	2034,80	17,10	34,20
Binomial Negativa Misto	69,50	-0,20	1,14	-0,05	-1017,50	2035,00	16,90	33,90
Poisson Misto Zero Inflado	62,20	-0,20	2,05	0,96	-1018,00	2035,90	17,00	34,00
Poisson Misto com Mistura	49,90	-0,30	2,06	0,95	-1016,10	2032,20	16,90	33,90

Visto que os critérios de avaliação da qualidade do ajuste foram muito próximos e que a análise da edição anterior do campeonato foi feita para os modelos Poisson misto e Poisson misto com mistura, vamos agora analisar os modelos Binomial Negativa misto e Poisson misto zero inflado.

Avaliando a pontuação acumulada ao longo do campeonato para ambos os modelos, nas figuras abaixo,

Figura 3 - Pontuação estimada pelo modelo Binomial Negativa misto para a edição de 2021 do Brasileiro

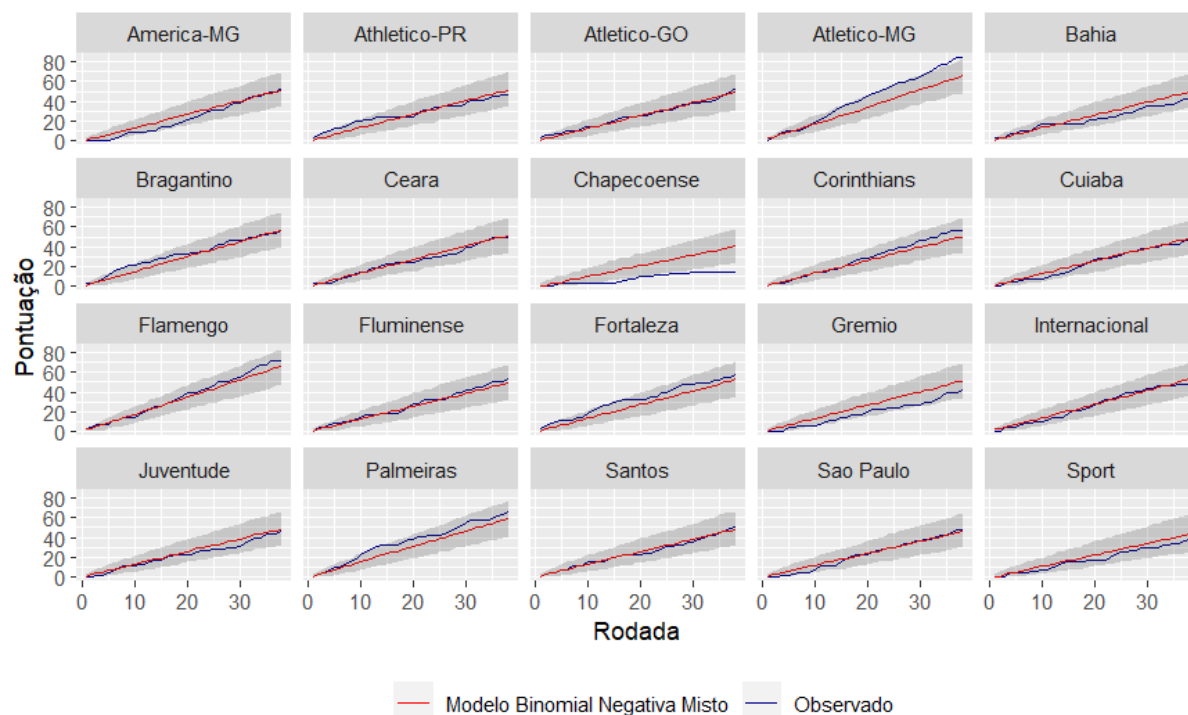
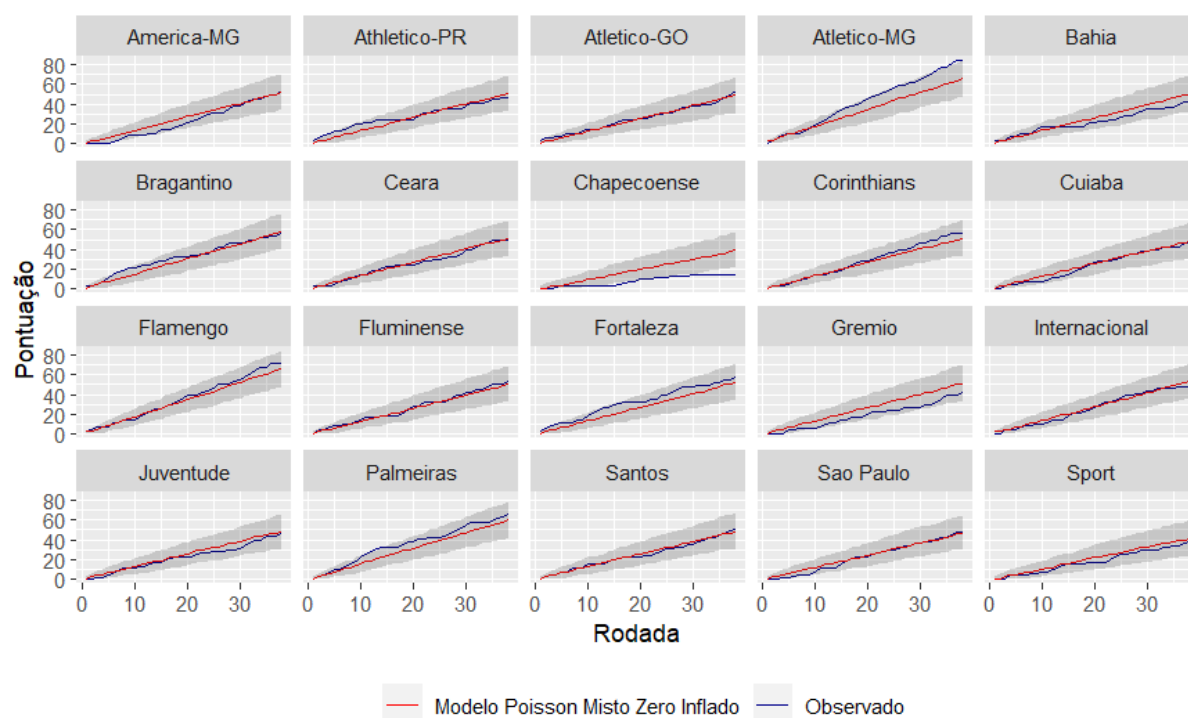


Figura 4 - Pontuação estimada pelo modelo Poisson misto zero inflado para a edição de 2021 do Brasileiro



pode-se observar que para os dois os modelos existem dois únicos times em que a pontuação observada ficou fora dos intervalos de credibilidade 95%, sendo eles o Atlético-MG e a Chapecoense.

Observando na Tabela 8, Para o modelo Binomial Negativa, a partir da trigésima-terceira rodada, a pontuação observada para o Atlético-MG é sempre acima do limite superior do intervalo de credibilidade 95%, enquanto que para o modelo Poisson misto zero inflado, a partir da trigésima-terceiro rodada, a pontuação observada é maior que o limite superior do intervalo de credibilidade 95% para todas as rodadas, exceto a trigésima-quinta e a trigésima-oitava, em que os valores são exatamente iguais. Outro ponto que chama a atenção nesta tabela é que os valores de mediana a posteriori para a pontuação são iguais entre os dois modelos, para todas as rodadas selecionadas.

Tabela 8 - Pontuação estimada para o Atlético-MG para a edição de 2021 do Brasileiro

Time	Rodada	Pontuação observada	Modelo Binomial Negativa			Modelo Poisson Misto Zero Inflado		
			Mediana	2,5%	97,5%	Mediana	2,5%	97,5%
Atlético-MG	33	74	57,00	41,00	72,00	57,00	40,00	73,53
	34	77	59,00	42,00	75,00	59,00	42,00	76,00
	35	78	60,00	44,00	76,00	60,00	43,00	78,00
	36	81	62,00	46,00	78,00	62,00	44,00	80,00
	37	84	64,00	47,00	80,53	64,00	46,00	82,00
	38	84	66,00	48,00	83,00	66,00	47,00	84,00

Para entender mais a fundo por que os valores de mediana para a pontuação estão iguais, podemos olhar para os parâmetros fixos dos modelos:

Tabela 9 - Estimativas a posteriori para os parâmetros dos modelos Binomial Negativa e Poisson misto zero inflado para a edição de 2021 do Brasileirão

Parâmetro	Modelo Binomial Negativa			Modelo Poisson Misto Zero Inflado		
	Mediana	2,5%	97,5%	Mediana	2,5%	97,5%
home	0,25	0,13	0,38	0,19	0,10	0,38
mu_att	0,14	-0,11	0,35	0,06	-0,07	0,36
mu_def	-0,15	-0,37	0,05	-0,25	-0,38	0,03
sigma_att	0,21	0,12	0,34	0,18	0,12	0,36
sigma_def	0,07	0,01	0,19	0,06	0,02	0,21
bn_phi	96,84	14,11	1708,00			
p_zero_home				0,00	0,00	0,02
p_zero_away				0,00	0,00	0,10

Podemos observar que para ambos os modelos, o efeito fixo de jogar uma partida em seu mando de campo (*home*) é positivo, indicando que times que jogam em casa tendem a marcar mais gols. Todavia, podemos observar que para o modelo Poisson misto zero inflado, os parâmetros que indicam excesso de zeros para os times mandante e visitante, respectivamente,  $p\_zero\_home$ ,  $p\_zero\_away$ , possuem valores medianos iguais a zero e intervalos de credibilidade 95% iguais a  $[0; 0,02]$  e  $[0; 0,10]$ . Isto indica que o modelo está atribuindo uma probabilidade muito pequena para os times terem excesso de partidas em que não marcam gols, sendo, neste caso, um modelo praticamente equivalente ao modelo Poisson misto. As estimativas a posteriori dos parâmetros dos demais modelos encontram-se no Apêndice.

### 3.1.4. Brasileirão 2022

Tendo a Tabela 10 como referência para avaliar a performance dos modelos em relação ao ajuste dos dados do Campeonato Brasileiro de 2022, podemos observar que, novamente, cada um dos modelos apresentou melhor performance segundo algum dos critérios, não sendo possível destacar com clareza qual o melhor modelo para este conjunto de dados. O modelo Poisson misto apresentou melhor EQM para o número de gols marcados e melhor *looic*, o modelo Binomial Negativa misto apresentou menor viés para pontuação, o modelo Poisson misto zero inflado apresentou menor viés para o número de gols marcados e o modelo Poisson misto com mistura apresentou menor EQM para pontuação.

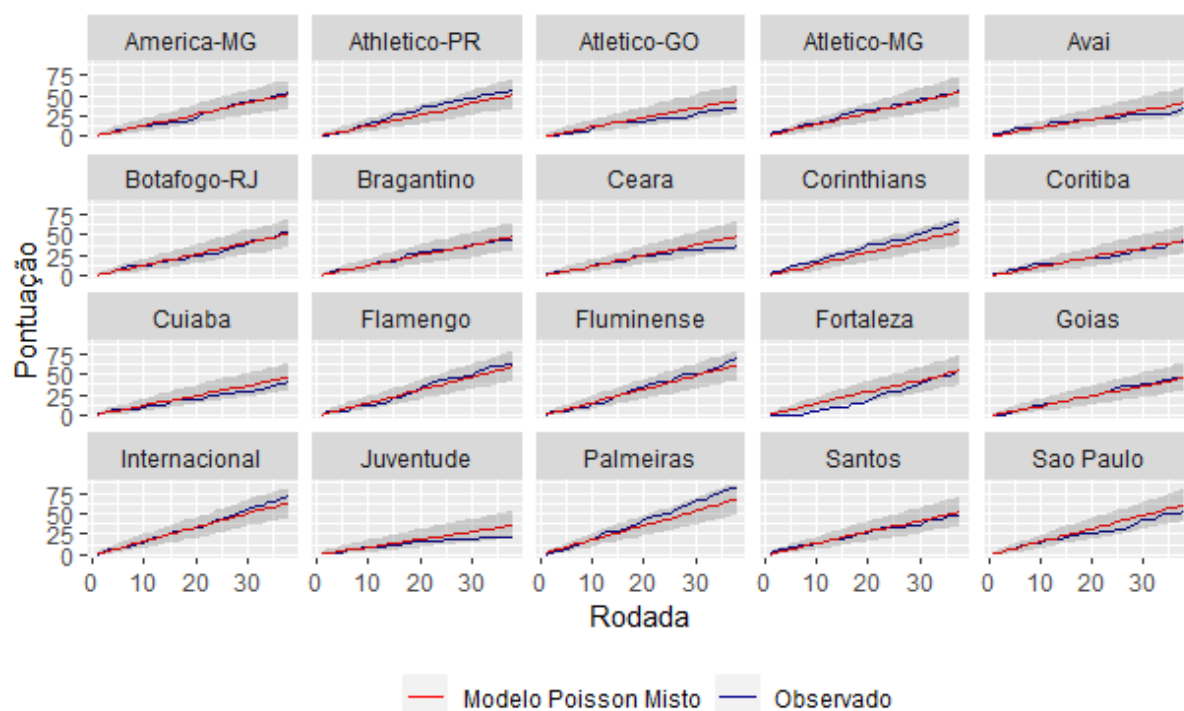
Tabela 10 - Comparação entre modelos para a edição de 2022 do Brasileirão

Modelo	Pontuação		Gols		LOO - Estimação		LOO - Erro padrão	
	EQM	Viés	EQM	Viés	elpd_loo	looic	elpd_loo	looic
Poisson Misto	54,25	-0,35	0,75	-0,55	-1026,70	2053,40	15,70	31,30
Binomial Negativa Misto	52,90	-0,20	1,19	-1,04	-1027,40	2054,80	15,60	31,10
Poisson Misto Zero Inflado	52,55	-0,25	0,90	-0,05	-1026,90	2053,80	15,60	31,30
Poisson Misto com Mistura	42,55	-0,35	0,90	-0,07	-1027,20	2054,40	15,70	31,30

Baseado no critério de menor *looic* para escolha do modelo, vamos analisar a estimativa a posteriori para a pontuação acumulada ao longo do campeonato para o modelo Poisson misto. Na Figura 5, pode-se observar que o modelo acerta a classificação dos três primeiros colocados, respectivamente, Palmeiras, Internacional e Fluminense, porém erra a classificação do quarto time, estimando ser o São Paulo, enquanto o time observado foi o Corinthians. Em relação aos quatro times rebaixados na competição, o modelo também acerta três equipes: Juventude, Avaí e Atlético-GO. A quarta equipe rebaixada estimada pelo modelo foi o Coritiba, enquanto que no campeonato real o Ceará foi o clube rebaixado. Além disso, os intervalos de credibilidade 95% contêm o valor observado para a pontuação final de todos os times do campeonato.



Figura 5 - Pontuação estimada pelo modelo Poisson misto para a edição de 2022 do Brasileirão



Avaliando o parâmetros estimados a posteriori pelo modelo, presentes na Tabela 11, é interessante ressaltar que para o conjunto de dados do Brasileirão 2022 as medianas a posteriori dos parâmetros  $\mu_{att}$  e  $\mu_{def}$  têm sinais opostos em relação aos outros anos do campeonato, desta vez a mediana de  $\mu_{att}$  é um valor negativo e a mediana de  $\mu_{def}$  é um valor positivo. As estimativas a posteriori dos parâmetros dos demais modelos encontram-se no Apêndice.

Tabela 11 - Estimativas a posteriori para os parâmetros estimados pelo modelo Poisson misto para a edição de 2022 do Brasileirão

Parâmetro	Mediana	2,5%	97,5%
home	0,37	0,24	0,50
mu_att	-0,18	-0,41	0,05
mu_def	0,12	-0,12	0,34
sigma_att	0,18	0,09	0,31
sigma_def	0,20	0,10	0,34

A razão deste comportamento dos parâmetros  $\mu$  fica mais evidente quando são analisados os efeitos aleatórios de ataque e defesa, como mostra a tabela 12. Para esta edição do campeonato, o time de referência foi o São Paulo, ao qual o modelo estimou que faria um bom campeonato, terminando na quarta colocação. Por isso, como todos os efeitos são em referência a este time, a interpretação dos parâmetros é explicada como, em mediana, os times que possuem efeito aleatório de ataque negativo têm menor taxa de ocorrência de gols marcados do que a equipe do São Paulo teria caso enfrentassem o mesmo adversário em seus respectivos mandos de campo. A mesma interpretação é válida para os efeitos aleatórios de defesa, em mediana, os times que possuem efeito aleatório de defesa positivo têm maior taxa de ocorrência de concederem gols do que a equipe do São Paulo teria caso enfrentassem o mesmo adversário em seus respectivos mandos de campo.

Tabela 12 - Estimativas a posteriori para os efeitos aleatórios de ataque e defesa pelo modelo Poisson misto para a edição de 2022 do Brasileirão

Time	Ataque			Defesa		
	Mediana	2,5%	97,5%	Mediana	2,5%	97,5%
America-MG	-0,23	-0,55	0,07	0,05	-0,27	0,34
Athletico-PR	-0,14	-0,44	0,16	0,16	-0,15	0,47
Athletico-GO	-0,24	-0,56	0,05	0,28	-0,02	0,58
Athletico-MG	-0,18	-0,50	0,13	0,00	-0,32	0,32
Avai	-0,31	-0,66	-0,00	0,30	0,02	0,61
Botafogo-RJ	-0,22	-0,55	0,08	0,09	-0,22	0,40
Bragantino	-0,12	-0,42	0,18	0,30	0,01	0,59
Ceara	-0,32	-0,65	-0,01	0,06	-0,26	0,35
Corinthians	-0,19	-0,49	0,11	-0,01	-0,37	0,31
Coritiba	-0,25	-0,56	0,06	0,31	-0,01	0,61
Cuiaba	-0,36	-0,72	-0,06	0,07	-0,26	0,37
Flamengo	0,01	-0,29	0,30	0,04	-0,29	0,35
Fluminense	0,04	-0,25	0,35	0,07	-0,27	0,39
Fortaleza	-0,16	-0,47	0,13	0,03	-0,28	0,34
Goiias	-0,24	-0,55	0,06	0,22	-0,09	0,52
Internacional	-0,02	-0,31	0,29	-0,09	-0,45	0,25
Juventude	-0,38	-0,73	-0,08	0,41	0,11	0,71
Palmeiras	0,07	-0,24	0,38	-0,15	-0,53	0,19
Santos	-0,19	-0,51	0,12	0,06	-0,28	0,36
Sao Paulo	0,00	0,00	0,00	0,00	0,00	0,00

### 3.2. Comparação entre softwares

A comparação de performance entre os softwares Stan e Jags foi feita considerando os dados para a edição de 2022 do Brasileirão e apenas a parte da estimação dos parâmetros, ou seja, a predição do placar das partidas não foi considerada. As configurações para ajuste dos modelos foram quatro cadeias de Markov com 5000 iterações para cada cadeia e período de aquecimento (*burn-in*) igual a 2500.

Os critérios definidos para avaliar a qualidade do ajuste foram os graus de convergência das cadeias de Markov (Brooks e Gelman, 1998), calculado pela estatística  $\hat{R}$ , e o tamanho efetivo de amostra, calculado pela estatística  $n_{eff}$ . No

primeiro método são comparados os parâmetros estimados pelo modelo entre cadeias e dentro das cadeias de Markov, caso haja divergências de estimativas para os parâmetros,  $\hat{R}$  será maior que 1. Procuramos por valores próximos a 1 e menores que 1,05 para indicar a convergência do modelo. Já a estatística  $n_{eff}$  visa estimar o número de amostras obtidas da distribuição a posteriori que são efetivamente independentes, visto que nos métodos MCMC (*Monte Carlo Markov Chain*) as amostras são comumente autocorrelacionadas dentro de uma mesma cadeia. Neste caso, procuramos por maiores valores da estatística  $n_{eff}$ .

É necessário destacar que o software Jags não possui uma implementação nativa para a distribuição de Cauchy, como alternativa para que ambos os softwares fizessem uso das mesmas distribuições a priori foi avaliado o uso da distribuição Gamma pelo Stan e o uso da distribuição t-student com um grau de liberdade pelo Jags para as distribuições a priori de desvios-padrão. Em ambos os casos, os modelos não convergiram de forma satisfatória. Em virtude disso, os modelos foram especificados utilizando a distribuição Gamma para os hiperparâmetros no Jags, assim como feito por Baio e Blangiardo (2010), e a distribuição Cauchy no Stan.

Para o modelo Poisson misto, como observado na Tabela 13, os valores de  $\hat{R}$  são iguais a 1 para todos os parâmetros fixos em ambos os softwares, exceto para o  $\sigma_{att}$  que apresenta um valor ligeiramente maior de 1,01 no Stan. Em relação à estatística  $n_{eff}$ , nota-se que o Stan fornece maiores valores de número efetivo de amostras para aqueles parâmetros em que a distribuição especificada foi igual entre os dois softwares. Quando é comparado os parâmetros  $\sigma_{att}$  e  $\sigma_{def}$ , o Jags fornece valores maiores para  $n_{eff}$ .

Tabela 13 - Comparação entre softwares para o modelo Poisson misto

Parametros	Stan		Jags	
	Rhat	n.eff	Rhat	n.eff
home	1,00	5290,00	1,00	3488,00
mu_att	1,00	816,00	1,00	589,00
mu_def	1,00	881,00	1,00	629,00
sigma_att	1,01	719,00	1,00	7749,00
sigma_def	1,00	3364,00	1,00	7740,00

Para o modelo Binomial Negativa misto, como observado na Tabela 14, os valores de  $\hat{R}$  são todos menores que 1,05 para todos os parâmetros fixos em ambos os softwares, exceto para o parâmetro  $\phi$  no Jags que apresenta um valor de 1,06, além disso este mesmo parâmetro apresenta um  $n_{eff}$  destoante entre os dois softwares, indicando a não convergência no Jags do parâmetro. Em relação à estatística  $n_{eff}$ , nota-se o mesmo comportamento do modelo Poisson misto, em que o Stan fornece maiores valores de número efetivo de amostras para aqueles parâmetros em que a distribuição especificada foi igual entre os dois softwares. E, quando comparado os parâmetros  $\sigma_{att}$  e  $\sigma_{def}$ , o Jags fornece valores maiores para  $n_{eff}$ .

Tabela 14 - Comparação entre softwares para o modelo Binomial Negativa

Parametros	Stan		Jags	
	Rhat	n.eff	Rhat	n.eff
home	1,00	6363,00	1,00	3427,00
mu_att	1,01	663,00	1,01	562,00
mu_def	1,01	693,00	1,01	576,00
sigma_att	1,00	3761,00	1,00	6825,00
sigma_def	1,00	3739,00	1,00	7498,00
bn_phi	1,00	4883,00	1,06	309,00

Para o modelo Poisson misto zero inflado, como observado na Tabela 15, os valores de  $\hat{R}$  são todos menores que 1,05 para os parâmetros fixos em ambos os softwares, exceto para os parâmetros que denotam a probabilidade de excesso de zeros,  $p_{zero\_home}$  e  $p_{zero\_away}$ , para os times mandante e visitante, respectivamente, no Jags. Estes parâmetros apresentam valores extremamente altos para  $\hat{R}$ , além de um número efetivo de amostras muito pequeno em relação ao número de amostras configuradas na execução do modelo, tudo isso indica que estes parâmetros não convergiram. Em relação à estatística  $n_{eff}$ , nota-se um comportamento um pouco diferente, para este modelo  $\mu_{att}$  e  $\mu_{def}$  apresentam  $n_{eff}$  maiores no Jags do que no Stan.

Tabela 15 - Comparação entre softwares para o modelo Poisson misto zero inflado

	Stan		Jags	
Parâmetros	Rhat	n.eff	Rhat	n.eff
home	1,00	4333,00	1,00	3439,00
mu_att	1,01	542,00	1,00	669,00
mu_def	1,01	589,00	1,00	671,00
sigma_att	1,00	2563,00	1,00	7365,00
sigma_def	1,00	3101,00	1,00	7968,00
p_zeros_home	1,00	7483,00	1,19	87,00
p_zeros_away	1,00	10630,00	1,21	200,00

Analizando o *traceplot* das cadeias dos parâmetros  $\sigma_{att}$ ,  $\sigma_{def}$ ,  $p_1$  e  $p_2$  tanto para o Jags quanto para o Stan, como visto nas figuras abaixo, pode-se observar que os dois últimos parâmetros citados não convergiram no software Jags.

Figura 6 - *Traceplot* para as estimativas a posteriori de parâmetros do modelo Poisson misto zero inflado em Jags

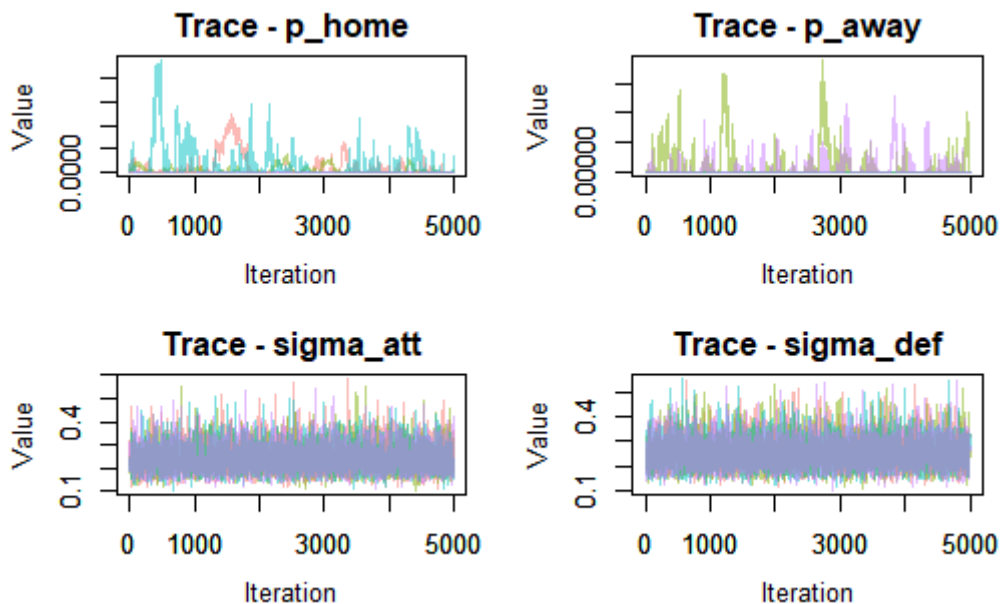
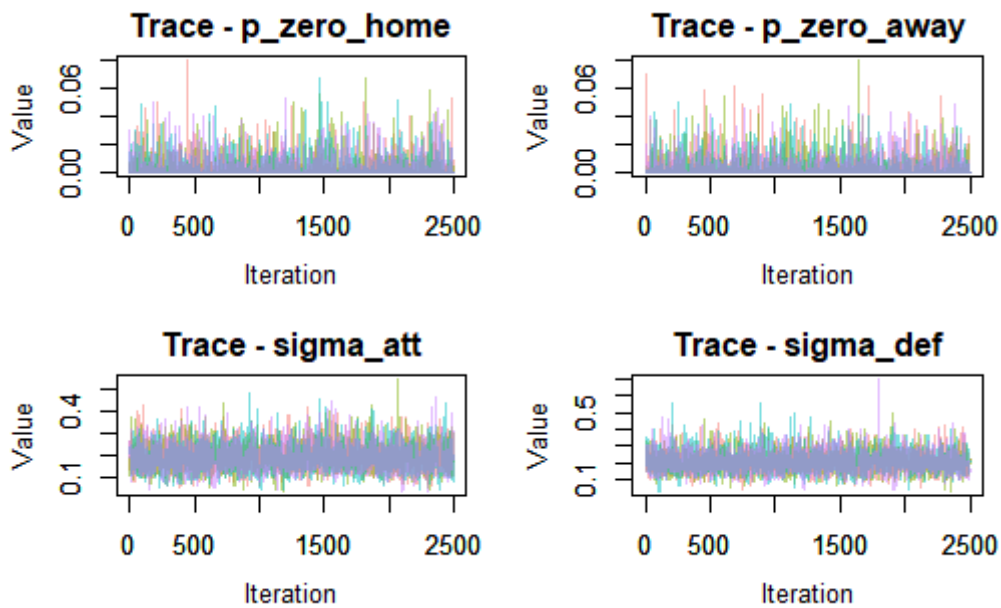


Figura 7 - *Traceplot* para as estimativas a posteriori de parâmetros do modelo Poisson misto zero inflado em Stan



#### 4. Considerações finais

A partir dos ajustes dos modelos propostos aos dados do Campeonato Brasileiro de Futebol, foi possível constatar que o modelo Poisson misto com mistura apresentou resultados relativamente melhores em termos de EQM e Viés em relação aos outros modelos. Em termos de *LOO-CV* não foi possível determinar que algum modelo seja melhor que os outros, visto que a diferença de valores para a estimativa deste critério foi menor que o erro padrão desta estimativa. De modo geral, é prudente afirmar que todos os modelos são similares.

Possíveis melhorias aos modelos podem ser implementadas a partir da definição de distribuições a priori mais informativas, como foi visto para o caso do modelo Poisson misto zero inflado, em que os parâmetros  $p_j$  não foram capazes de atribuir uma porcentagem considerável de placares iguais a zero.

Outro desdobramento a estes modelos seria uma modelagem que acomodasse diferentes valores ao efeito fixo *home*. Por ser modelado como um efeito fixo para todos os times disputando o campeonato, este efeito não contempla as possíveis diferenças que podem existir entre mandos de campo. Por exemplo, times que têm torcidas mais numerosas e, portanto, uma propensão maior a saírem vitoriosos em seu mando de campo.

Em relação a comparação entre modelos, foi possível constatar que, de maneira geral, o Stan fornece amostras menos autocorrelacionadas que o Jags, o

que pode ser melhorado adicionando um *lag* durante a amostragem neste último software. Também foi nítido observar que devido a implementação distinta de distribuições a priori para os desvios-padrões, os modelos não são efetivamente comparáveis, o que torna a comparação enviesada e incompleta. Pode-se avaliar, como alternativa, alguma terceira distribuição a ser usada como priori de desvios-padrões de modo a tornar os modelos comparáveis entre os dois softwares.

## 5. Referências bibliográficas

D. Prasetio and D. Harlili, "Predicting football match results with logistic regression," 2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA), Penang, Malaysia, 2016, pp. 1-5.

Karlis, D., & Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3), 381–393.

Gabrio, A. (2021). Bayesian hierarchical models for the prediction of volleyball results. *Journal of Applied Statistics*, 48(2), 301–321.

Baio, G., & Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, 37(2), 253–264.

Duque, A. (2022). Campeonato Brasileiro de futebol. Disponível em: <https://www.kaggle.com/datasets/adaoduke/campeonato-brasileiro-de-futebol>.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Stan Development Team (2022). RStan: the R interface to Stan. R package version 2.21.7. <https://mc-stan.org/>.



Martyn Plummer (2022). rjags: Bayesian Graphical Models using MCMC. R package version 4-13. <https://CRAN.R-project.org/package=rjags>

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing*, 27(5), 1413–1432.

Lawless, Jerald F. “Negative Binomial and Mixed Poisson Regression.” *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, vol. 15, no. 3, 1987, pp. 209–25.

Diane Lambert (1992) Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing, *Technometrics*, 34:1, 1-14.

Stephen P. Brooks & Andrew Gelman (1998) General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics*, 7:4, 434-455.

Vehtari A, Gabry J, Magnusson M, Yao Y, Bürkner P, Paananen T, Gelman A (2023). “loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models.” R package version 2.6.0, <URL: <https://mc-stan.org/loo/>>.

## **6. Apêndice**

Neste apêndice serão apresentados figuras e tabelas complementares às análises discutidas na Seção 3, assim como os códigos dos modelos implementados em Jags e Stan e a classificação final das edições 2020, 2021 e 2022 do Campeonato Brasileiro de Futebol.

Figura 8 - Pontuação estimada pelo modelo Binomial Negativa misto para a edição de 2020 do Brasileirão

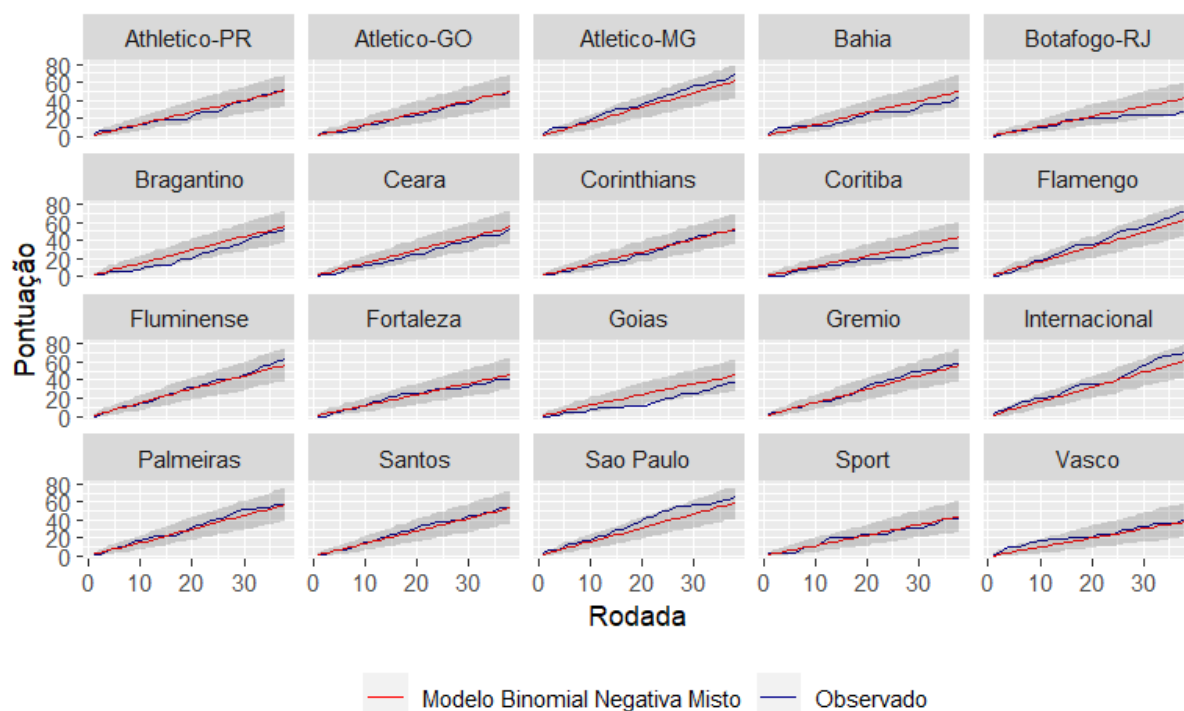


Figura 9 - Pontuação estimada pelo modelo Poisson misto zero inflado para a edição de 2020 do Brasileirão

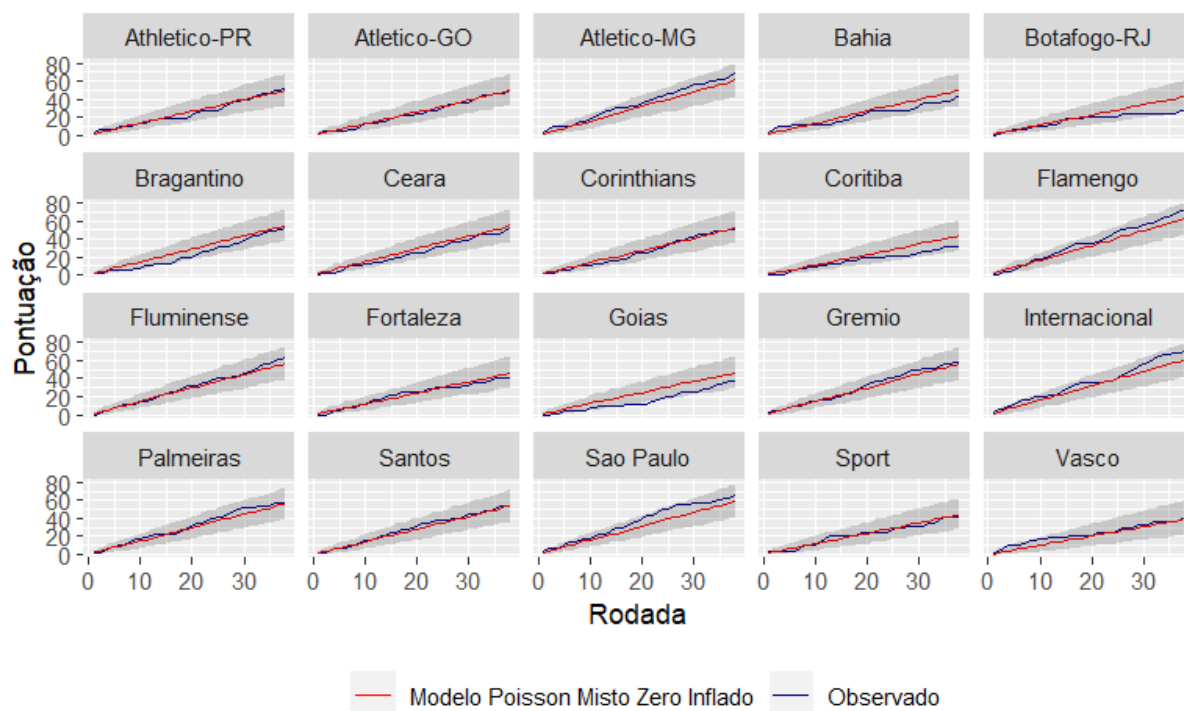


Tabela 16 - Estimativas a posteriori dos parâmetros para o modelo Modelo Binomial Negativa para a edição de 2020 do Brasileirão

<b>Parâmetro</b>	<b>Mediana</b>	<b>2,5%</b>	<b>97,5%</b>
home	0,28	0,15	0,40
mu_att	0,24	-0,01	0,47
mu_def	-0,20	-0,41	0,06
sigma_att	0,21	0,10	0,35
sigma_def	0,10	0,04	0,21
bn_phi	84,92	18,59	3223,00

Tabela 17 - Estimativas a posteriori dos parâmetros para o modelo Poisson misto zero inflado para a edição de 2020 do Brasileirão

<b>Parâmetro</b>	<b>Mediana</b>	<b>2,5%</b>	<b>97,5%</b>
home	0,27	0,13	0,40
mu_att	0,23	-0,02	0,46
mu_def	-0,17	-0,39	0,07
sigma_att	0,20	0,10	0,34
sigma_def	0,09	0,02	0,21
p_zero_home	0,00	0,00	0,02
p_zero_away	0,00	0,00	0,04

Figura 10 - Pontuação estimada pelo modelo Poisson misto para a edição de 2021 do Brasileiro

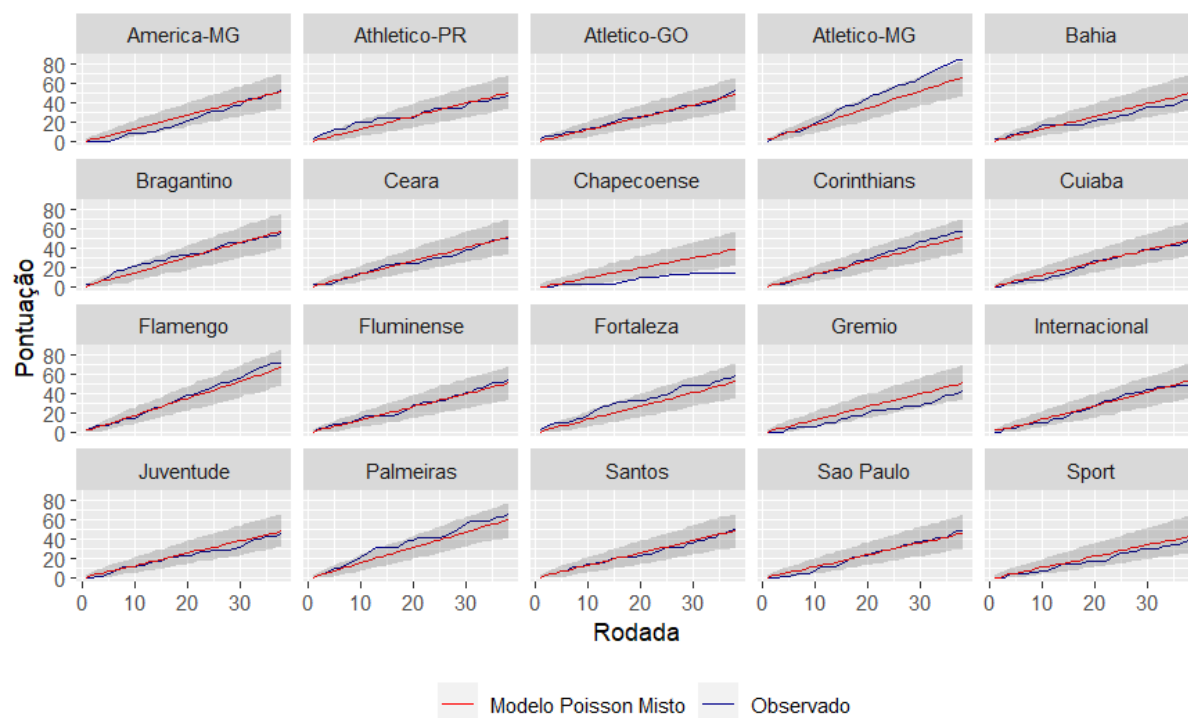


Figura 11 - Pontuação estimada pelo modelo Poisson misto com mistura para a edição de 2021 do Brasileiro

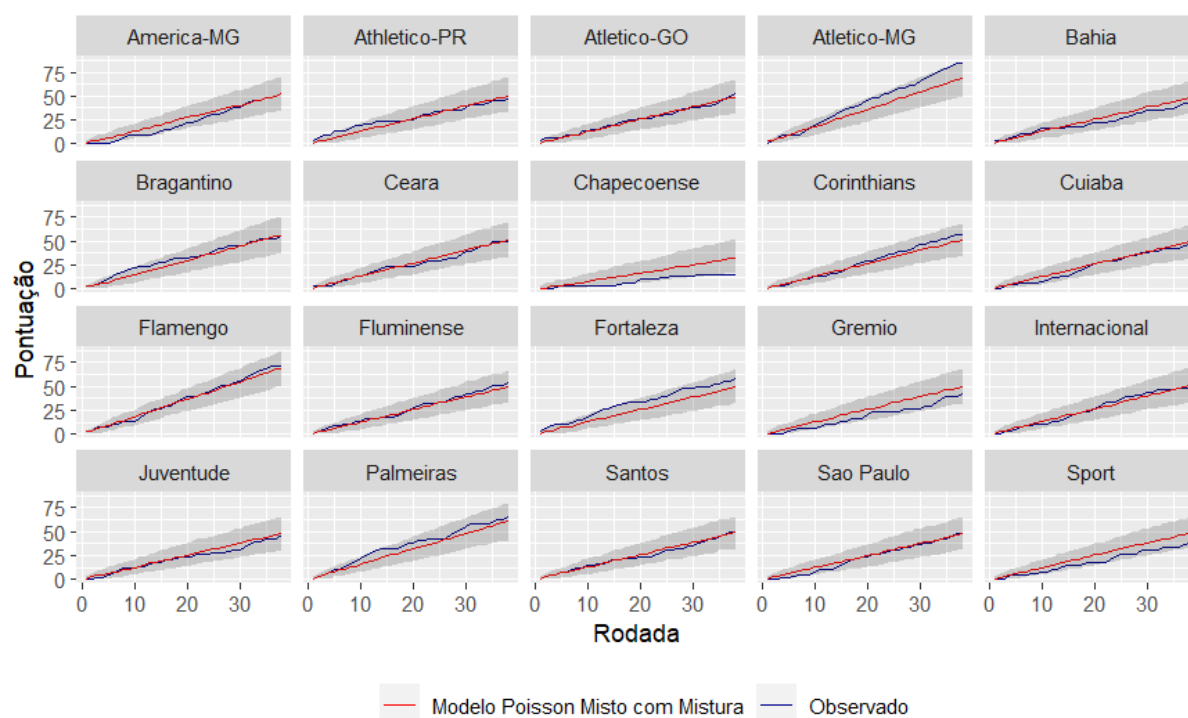


Tabela 18 - Estimativas a posteriori dos parâmetros para o modelo Poisson misto para a edição de 2021 do Brasileirão

<b>Parâmetro</b>	<b>Mediana</b>	<b>2,5%</b>	<b>97,5%</b>
home	0,26	0,13	0,39
mu_att	0,11	-0,11	0,32
mu_def	-0,16	-0,35	0,05
sigma_att	0,22	0,12	0,37
sigma_def	0,09	0,03	0,20

Tabela 19 - Estimativas a posteriori dos parâmetros para o modelo Poisson misto com mistura para a edição de 2021 do Brasileirão

<b>Parâmetro</b>	<b>Mediana</b>	<b>2,5%</b>	<b>97,5%</b>
home	0,21	0,10	0,35
mu_att1	-0,15	-1,33	-0,01
mu_att3	0,25	0,02	0,82
mu_def1	0,13	0,00	2,19
mu_def3	-0,15	-0,45	-0,01
sigma_att1	0,15	0,01	2,46
sigma_att3	0,25	0,03	1,24
sigma_def1	0,30	0,03	4,90
sigma_def3	0,08	0,01	0,75

Figura 12 - Pontuação estimada pelo modelo Binomial Negativa misto para a edição de 2022 do Brasileiro

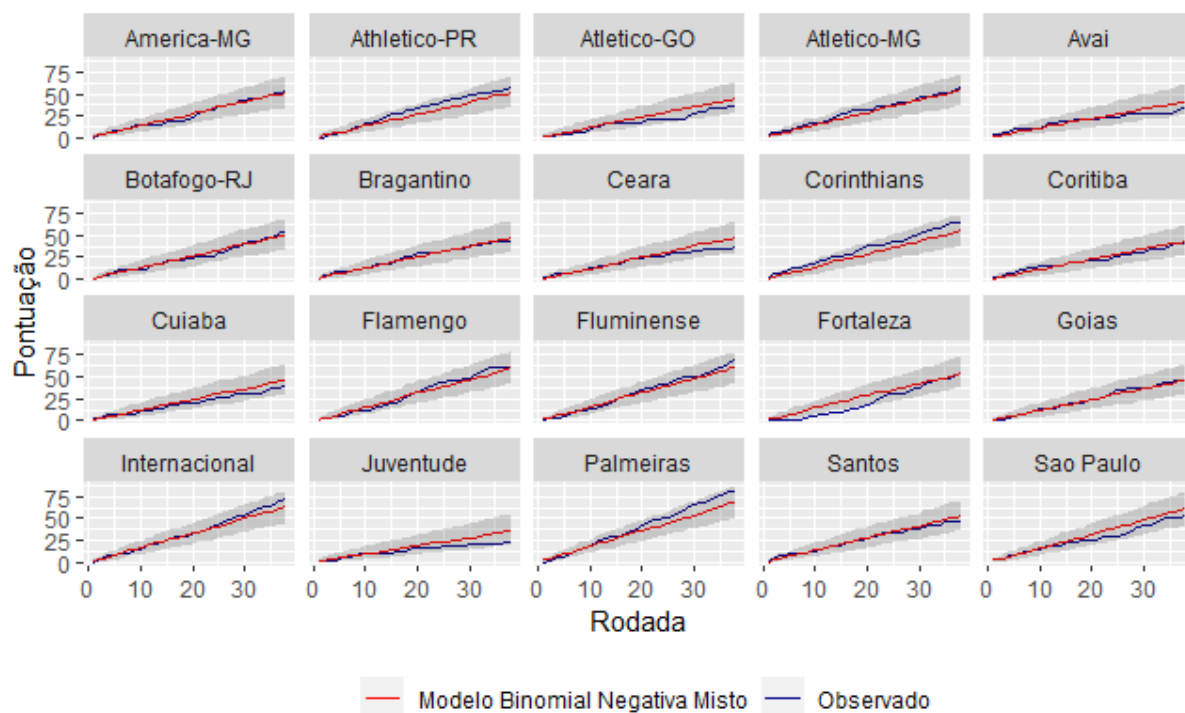


Figura 13 - Pontuação estimada pelo modelo Poisson misto zero inflado para a edição de 2022 do Brasileiro

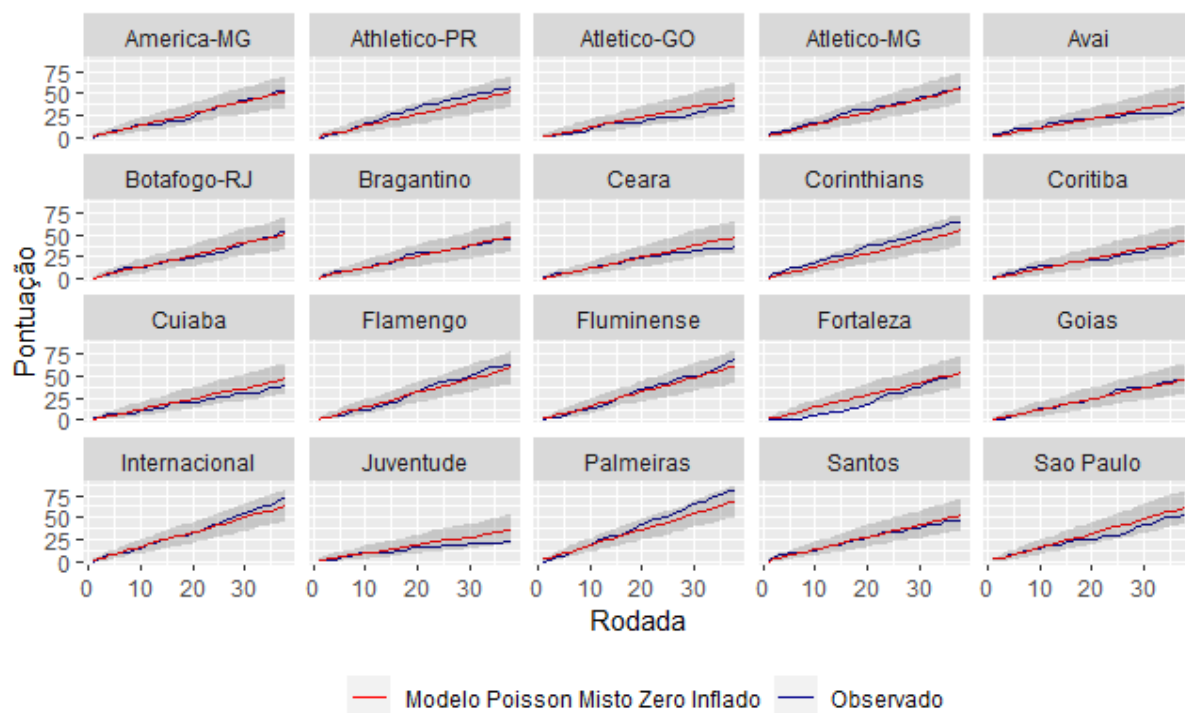


Figura 14 - Pontuação estimada pelo modelo Poisson misto com mistura para a edição de 2022 do Brasileirão

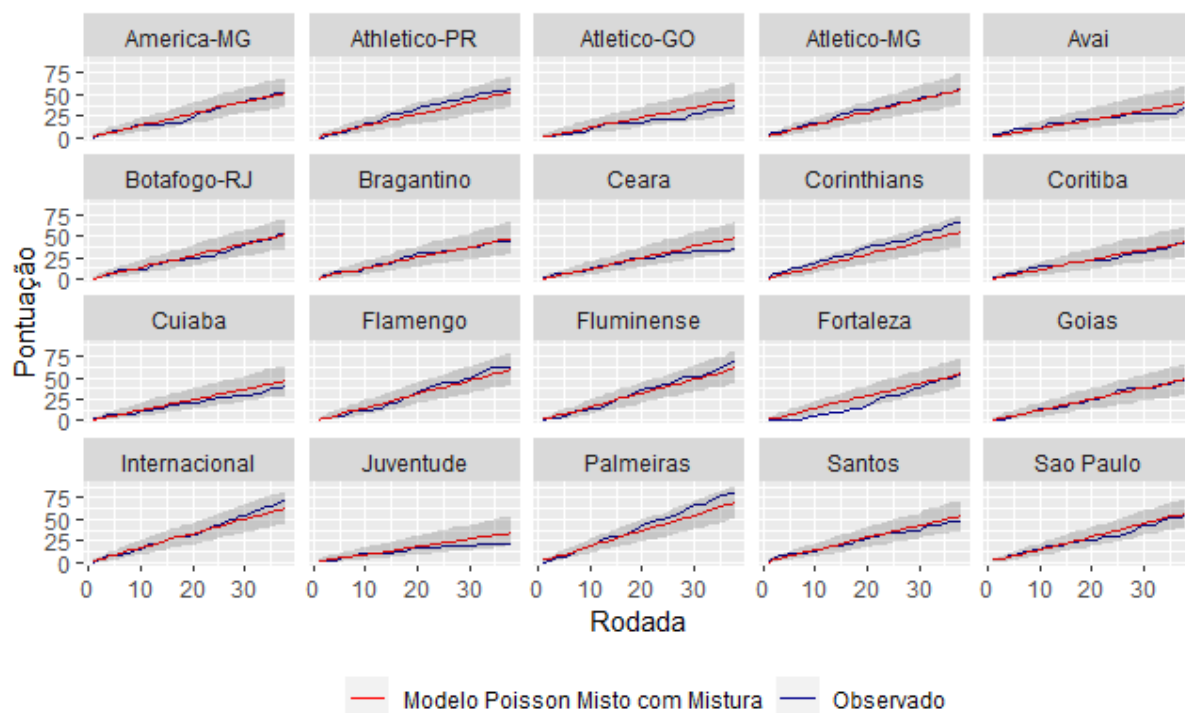


Tabela 20 - Estimativas a posteriori dos parâmetros para o modelo Binomial Negativa para a edição de 2022 do Brasileirão

Parâmetro	Mediana	2,5%	97,5%
home	0,37	0,24	0,50
mu_att	-0,18	-0,41	0,04
mu_def	0,11	-0,11	0,33
sigma_att	0,18	0,09	0,31
sigma_def	0,20	0,10	0,34
bn_phi	117,10	21,73	3260,71

Tabela 21 - Estimativas a posteriori dos parâmetros para o modelo Poisson misto zero inflado para a edição de 2022 do Brasileirão

<b>Parâmetro</b>	<b>Mediana</b>	<b>2,5%</b>	<b>97,5%</b>
home	0,37	0,24	0,51
mu_att	-0,19	-0,44	0,04
mu_def	0,12	-0,09	0,35
sigma_att	0,18	0,09	0,30
sigma_def	0,20	0,10	0,34
p_zero_home	0,00	0,00	0,02
p_zero_away	0,00	0,00	0,02

Tabela 22 - Estimativas a posteriori dos parâmetros para o modelo Poisson misto com mistura para a edição de 2022 do Brasileirão

<b>Parâmetro</b>	<b>Mediana</b>	<b>2,5%</b>	<b>97,5%</b>
home	0,36	0,23	0,48
mu_att1	-0,25	-0,67	-0,03
mu_att3	0,14	0,01	2,07
mu_def1	0,26	0,02	0,62
mu_def3	-0,17	-1,90	-0,01
sigma_att1	0,11	0,01	0,94
sigma_att3	0,30	0,02	4,89
sigma_def1	0,18	0,01	0,97
sigma_def3	0,21	0,00	3,52



Tabela 23 - Classificação final do Brasileirão edição 2020

<b>Time</b>	<b>Pontuação</b>	<b>Classificação</b>
Flamengo	71	1
Internacional	70	2
Atletico-MG	68	3
Sao Paulo	66	4
Fluminense	64	5
Gremio	59	6
Palmeiras	58	7
Santos	54	8
Athletico-PR	53	9
Bragantino	53	10
Ceara	52	11
Corinthians	51	12
Atletico-GO	50	13
Bahia	44	14
Sport	42	15
Fortaleza	41	16
Vasco	41	17
Goias	37	18
Coritiba	31	19
Botafogo-RJ	27	20

Tabela 24 - Classificação final do Brasileirão edição 2021

<b>Time</b>	<b>Pontuação</b>	<b>Classificação</b>
Atletico-MG	84	1
Flamengo	71	2
Palmeiras	66	3
Fortaleza	58	4
Corinthians	57	5
Bragantino	56	6
Fluminense	54	7
America-MG	53	8
Atletico-GO	53	9
Santos	50	10
Ceara	50	11
Internacional	48	12
Sao Paulo	48	13
Athletico-PR	47	14
Cuiaba	47	15
Juventude	46	16
Gremio	43	17
Bahia	43	18
Sport	38	19
Chapecoense	15	20

Tabela 25 - Classificação final do Brasileirão edição 2022

<b>Time</b>	<b>Pontuação</b>	<b>Classificação</b>
Palmeiras	81	1
Internacional	73	2
Fluminense	70	3
Corinthians	65	4
Flamengo	62	5
Athletico-PR	58	6
Athletico-MG	58	7
Fortaleza	55	8
Sao Paulo	54	9
America-MG	53	10
Botafogo-RJ	53	11
Santos	47	12
Goiás	46	13
Bragantino	44	14
Coritiba	42	15
Cuiaba	41	16
Ceara	37	17
Athletico-GO	36	18
Avai	35	19
Juventude	22	20

## Código para o modelo Poisson misto implementado em Stan

```
data {
  int<lower=1> G;
  int<lower=1> T;
  int<lower=0, upper=T> h[G];
  int<lower=0, upper=T> a[G];
  int<lower=0> y1[G];
  int<lower=0> y2[G];
}
parameters {
  real home;
  real mu_att;
  real mu_def;
  real<lower=0> sigma_att;
  real<lower=0> sigma_def;
  vector[T - 1] att_raw;
  vector[T - 1] def_raw;
}
transformed parameters {
  vector[T] att;
  vector[T] def;
  for (t in 1:(T-1)) {
    att[t] = att_raw[t];
    def[t] = def_raw[t];
  }
  att[T] = 0;
  def[T] = 0;
}
model {
  for (g in 1:G) {
    y1[g] ~ poisson_log(home + att[h[g]] + def[a[g]]);
    y2[g] ~ poisson_log(att[a[g]] + def[h[g]]);
  }
  for (t in 1:(T - 1)) {
    att_raw[t] ~ normal(mu_att, sigma_att);
    def_raw[t] ~ normal(mu_def, sigma_def);
  }
  home ~ normal(0, 10);
  mu_att ~ normal(0, 10);
  mu_def ~ normal(0, 10);
  sigma_att ~ cauchy(0, 2.5);
  sigma_def ~ cauchy(0, 2.5);
}
```

```

}
generated quantities {
  vector[G] y1_tilde;
  vector[G] y2_tilde;
  vector[G] log_lik;
  for (g in 1:G) {
    y1_tilde[g] = poisson_log_rng(home + att[h[g]] + def[a[g]]);
    y2_tilde[g] = poisson_log_rng(att[a[g]] + def[h[g]]);
    log_lik[g] = poisson_log_lpmf(y1[g] | home + att[h[g]] +
def[a[g]]) +
    poisson_log_lpmf(y2[g] | att[a[g]] + def[h[g]]);
  }
}

```

Código para o modelo Binomial Negativa misto implementado em Stan

```

data {
  int<lower=1> G;
  int<lower=1> T;
  int<lower=0, upper=T> h[G];
  int<lower=0, upper=T> a[G];
  int<lower=0> y1[G];
  int<lower=0> y2[G];
}

parameters {
  real home;
  real mu_att;
  real mu_def;
  real<lower=0> bn_size;
  real<lower=0> sigma_att;
  real<lower=0> sigma_def;
  vector[T - 1] att_raw;
  vector[T - 1] def_raw;
}

transformed parameters {
  vector[T] att;
  vector[T] def;
  for (t in 1:(T-1)) {
    att[t] = att_raw[t];
    def[t] = def_raw[t];
  }
}

```

```

}

att[T] = 0;
def[T] = 0;
}

model {
  for (g in 1:G) {
    y1[g] ~ neg_binomial_2_log(home + att[h[g]] + def[a[g]],
bn_size);
    y2[g] ~ neg_binomial_2_log(att[a[g]] + def[h[g]], bn_size);
  }
  for (t in 1:(T - 1)) {
    att_raw[t] ~ normal(mu_att, sigma_att);
    def_raw[t] ~ normal(mu_def, sigma_def);
  }

  home ~ normal(0, 10);
  bn_size ~ cauchy(0, 2.5);
  mu_att ~ normal(0, 10);
  mu_def ~ normal(0, 10);
  sigma_att ~ cauchy(0, 2.5);
  sigma_def ~ cauchy(0, 2.5);
}

generated quantities {
  vector[G] y1_tilde;
  vector[G] y2_tilde;
  vector[G] log_lik;
  for (g in 1:G) {
    y1_tilde[g] = neg_binomial_2_log_rng(home + att[h[g]] +
def[a[g]], bn_size);
    y2_tilde[g] = neg_binomial_2_log_rng(att[a[g]] + def[h[g]],
bn_size);
    log_lik[g] = neg_binomial_2_log_lpmf(y1[g] | home +
att[h[g]] + def[a[g]], bn_size) +
neg_binomial_2_log_lpmf(y2[g] | att[a[g]] + def[h[g]],
bn_size);
  }
}

```

```

data {
  int<lower=1> G;
  int<lower=1> T;
  int<lower=0, upper=T> h[G];
  int<lower=0, upper=T> a[G];
  int<lower=0> y1[G];
  int<lower=0> y2[G];
}

parameters {
  real <lower=0, upper=1> p_zero_home;
  real <lower=0, upper=1> p_zero_away;
  real home;
  real mu_att;
  real mu_def;
  real<lower=0> sigma_att;
  real<lower=0> sigma_def;
  vector[T - 1] att_raw;
  vector[T - 1] def_raw;
}

transformed parameters {
  vector[T] att;
  vector[T] def;

  for (t in 1:(T-1)) {
    att[t] = att_raw[t];
    def[t] = def_raw[t];
  }

  att[T] = 0;
  def[T] = 0;
}

model {
  for (g in 1:G) {
    if (y1[g] == 0) {
      target += log_sum_exp(
        bernoulli_lpmf(1 | p_zero_home),
        bernoulli_lpmf(0 | p_zero_home) +
        poisson_log_lpmf(y1[g] | home +

```

```

att[h[g]] + def[a[g]])
    );

    } else {
        target += bernoulli_lpmf(0 | p_zero_home) +
        poisson_log_lpmf(y1[g] | home + att[h[g]] + def[a[g]]);

    }

    if (y2[g] == 0) {
        target += log_sum_exp(
            bernoulli_lpmf(1 | p_zero_away),

            bernoulli_lpmf(0 | p_zero_away) +
            poisson_log_lpmf(y2[g] | att[a[g]] +
def[h[g]])
        );
    } else {
        target += bernoulli_lpmf(0 | p_zero_away) +
        poisson_log_lpmf(y2[g] | att[a[g]] + def[h[g]]);
    }

}
for (t in 1:(T - 1)) {
    att_raw[t] ~ normal(mu_att, sigma_att);
    def_raw[t] ~ normal(mu_def, sigma_def);
}

home ~ normal(0, 10);
mu_att ~ normal(0, 10);
mu_def ~ normal(0, 10);
sigma_att ~ cauchy(0, 2.5);
sigma_def ~ cauchy(0, 2.5);
p_zero_home ~ beta(0.1, 0.4);
p_zero_away ~ beta(0.1, 0.4);
}

generated quantities {
    vector[G] y1_tilde;
    vector[G] y2_tilde;
    vector[G] log_lik_home;
    vector[G] log_lik_away;
    vector[G] log_lik;
}

```



```

for (g in 1:G) {
  y1_tilde[g] = poisson_log_rng(home + att[h[g]] + def[a[g]]);
  y2_tilde[g] = poisson_log_rng(att[a[g]] + def[h[g]]);

  if (y1[g] == 0) {
    log_lik_home[g] = log_sum_exp(log(p_zero_home), log(1 -
p_zero_home) + poisson_log_lpmf(y1[g] | home + att[h[g]] +
def[a[g]]));

    } else {

    log_lik_home[g] = log(1 - p_zero_home) +
poisson_log_lpmf(y1[g] | home + att[h[g]] + def[a[g]]);

    }

    if (y2[g] == 0) {
      log_lik_away[g] = log_sum_exp(log(p_zero_away), log(1 -
p_zero_away) + poisson_log_lpmf(y2[g] | att[a[g]] + def[h[g]]));

      } else {

      log_lik_away[g] = log(1 - p_zero_away) +
poisson_log_lpmf(y2[g] | att[a[g]] + def[h[g]]);

      }

      log_lik[g] = log_lik_home[g] + log_lik_away[g];
    }
  }
}

```

Código para o modelo Poisson misto com mistura implementado em Stan

```

data {
  int<lower=1> G;
  int<lower=1> T;
  int<lower=1> C;
  int<lower=0, upper=T> h[G];
  int<lower=0, upper=T> a[G];
  int<lower=0> y1[G];
  int<lower=0> y2[G];
}

```

```

parameters {
  real home;
  real<lower=-3, upper=0> mu_att1;
  real<lower=0, upper=3> mu_att3;
  real<lower=0, upper=3> mu_def1;
  real<lower=-3, upper=0> mu_def3;
  real<lower=0> sigma_att[C];
  real<lower=0> sigma_def[C];
  vector[T] att_raw;
  vector[T] def_raw;
  simplex[3] pi_att[T];
  simplex[3] pi_def[T];
}

transformed parameters {
  vector[T] att;
  vector[T] def;

  for (t in 1:(T-1)) {
    att[t] = att_raw[t];
    def[t] = def_raw[t];
  }
  att[T] = 0;
  def[T] = 0;
}

model {
  real m_att[C];
  real m_def[C];

  home ~ normal(0, 10);
  mu_att1 ~ normal(0, 10) T[-3,0];
  mu_att3 ~ normal(0, 10) T[0,3];
  mu_def1 ~ normal(0, 10) T[0,3];
  mu_def3 ~ normal(0, 10) T[-3,0];

  sigma_att ~ cauchy(0, 2.5);
  sigma_def ~ cauchy(0, 2.5);

  for (t in 1:T) {

```

```

    pi_att[t] ~ dirichlet(rep_vector(1, C));
    pi_def[t] ~ dirichlet(rep_vector(1, C));

    m_att[1] = log(pi_att[t, 1]) + normal_lpdf(att[t] | mu_att1,
sigma_att[1]);
    m_att[2] = log(pi_att[t, 2]) + normal_lpdf(att[t] | 0,
0.01);
    m_att[3] = log(pi_att[t, 3]) + normal_lpdf(att[t] | mu_att3,
sigma_att[3]);
    m_def[1] = log(pi_def[t, 1]) + normal_lpdf(def[t] | mu_def1,
sigma_def[1]);
    m_def[2] = log(pi_def[t, 2]) + normal_lpdf(def[t] | 0,
0.01);
    m_def[3] = log(pi_def[t, 3]) + normal_lpdf(def[t] | mu_def3,
sigma_def[3]);

    target += log_sum_exp(m_att) + log_sum_exp(m_def);
}

for (g in 1:G) {
    y1[g] ~ poisson_log(home + att[h[g]] + def[a[g]]);
    y2[g] ~ poisson_log(att[a[g]] + def[h[g]]);
}

}

generated quantities {
    vector[G] y1_tilde;
    vector[G] y2_tilde;
    vector[G] log_lik;
    for (g in 1:G) {
        y1_tilde[g] = poisson_log_rng(home + att[h[g]] + def[a[g]]);
        y2_tilde[g] = poisson_log_rng(att[a[g]] + def[h[g]]);
        log_lik[g] = poisson_log_lpmf(y1[g] | home + att[h[g]] +
def[a[g]]) +
            poisson_log_lpmf(y2[g] | att[a[g]] + def[h[g]]);
    }
}

```

Código para o modelo Poisson misto implementado em Jags

```

model
{
  for(g in 1:G) {
    y1[g] ~ dpois(lambda1[g])
    lambda1[g] = exp(home + att[h[g]] + def[a[g]])

    y2[g] ~ dpois(lambda2[g])
    lambda2[g] = exp(att[a[g]] + def[h[g]])
  }

  for(t in 1:(T-1)) {
    att_raw[t] ~ dnorm(mu_att, tau_att)
    def_raw[t] ~ dnorm(mu_def, tau_def)
    att[t] = att_raw[t]
    def[t] = def_raw[t]
  }

  att[T] = 0
  def[T] = 0

  tau_eff = 1/sigma_sq_eff

  mu_att ~ dnorm(0, tau_eff)
  mu_def ~ dnorm(0, tau_eff)

  home ~ dnorm(0, tau_eff)

  tau_att ~ dgamma(a_att, b_att)
  tau_def ~ dgamma(a_def, b_def)

  sigma_att = 1/sqrt(tau_att)
  sigma_def = 1/sqrt(tau_def)
}

```

Código para o modelo Binomial Negativa misto implementado em Jags

```

model
{
  for(g in 1:G) {
    y1[g] ~ dnegbin(bn_size / (lambda1[g] + bn_size), bn_size)
    lambda1[g] = exp(home + att[h[g]] + def[a[g]])
  }
}

```

```

    y1_tilde[g] ~ dnegbin(bn_size / (lambda1[g] + bn_size),
bn_size)

    y2[g] ~ dnegbin(bn_size / (lambda2[g] + bn_size), bn_size)
    lambda2[g] = exp(att[a[g]] + def[h[g]])

    y2_tilde[g] ~ dnegbin(bn_size / (lambda1[g] + bn_size),
bn_size)
  }

  for(t in 1:(T-1)) {
    att_raw[t] ~ dnorm(mu_att, tau_att)
    def_raw[t] ~ dnorm(mu_def, tau_def)
    att[t] = att_raw[t]
    def[t] = def_raw[t]
  }

  att[T] = 0
  def[T] = 0

  tau_eff = 1/sigma_sq_eff
  tau_size = 1/sigma_sq_eff_size

  mu_att ~ dnorm(0, tau_eff)
  mu_def ~ dnorm(0, tau_eff)

  home ~ dnorm(0, tau_eff)

  bn_size ~ dt(0, sigma_sq_eff_size, 1) T(0, )

  tau_att ~ dgamma(a_att, b_att)
  tau_def ~ dgamma(a_def, b_def)

  sigma_att = 1/sqrt(tau_att)
  sigma_def = 1/sqrt(tau_def)
}

```

Código para o modelo Poisson misto zero inflado implementado em Jags

```

{
  C <- 10000

```

```

for(g in 1:G) {

  zeros1[g] ~ dpois(fg1[g])
  fg1[g] <- -log(L1[g]) + C
  lambda1[g] = exp(home + att[h[g]] + def[a[g]])
  L1[g] <- p_home * equals(y1[g], 0) + (1 - p_home) *
(exp(-lambda1[g]) * lambda1[g] ^ y1[g]) / y1_factorial[g]

  zeros2[g] ~ dpois(fg2[g])
  fg2[g] <- -log(L2[g]) + C
  lambda2[g] = exp(att[a[g]] + def[h[g]])
  L2[g] <- p_away * equals(y2[g], 0) + (1 - p_away) *
(exp(-lambda2[g]) * lambda2[g] ^ y2[g]) / y2_factorial[g]

  y1_indicator[g] ~ dbin(p_home, 1)
  y1_aux[g] ~ dpois(lambda1[g])
  y1_tilde[g] = y1_aux[g] * (1 - y1_indicator[g])

  y2_indicator[g] ~ dbin(p_away, 1)
  y2_aux[g] ~ dpois(lambda2[g])
  y2_tilde[g] = y2_aux[g] * (1 - y2_indicator[g])
}

for(t in 1:(T-1)) {
  att_raw[t] ~ dnorm(mu_att, tau_att)
  def_raw[t] ~ dnorm(mu_def, tau_def)
  att[t] = att_raw[t]
  def[t] = def_raw[t]
}

att[T] = 0
def[T] = 0

tau_eff = 1/sigma_sq_eff
tau_size = 1/sigma_sq_eff_size

mu_att ~ dnorm(0, tau_eff)
mu_def ~ dnorm(0, tau_eff)

home ~ dnorm(0, tau_eff)

```

```

tau_att ~ dgamma(a_att, b_att)
tau_def ~ dgamma(a_def, b_def)

sigma_att = 1/sqrt(tau_att)
sigma_def = 1/sqrt(tau_def)

p_home ~ dbeta(alpha, beta)
p_away ~ dbeta(alpha, beta)

}

```

Código para o modelo Poisson misto com mistura implementado em Jags

```

model
{
  for(g in 1:G) {
    y1[g] ~ dpois(lambda1[g])
    lambda1[g] = exp(home + att[h[g]] + def[a[g]]) # h e a são
passados nos dados

    y2[g] ~ dpois(lambda2[g])
    lambda2[g] = exp(att[a[g]] + def[h[g]])
  }

  for(t in 1:(T-1)) {

    att_raw[t] ~ dnorm(mu_att[zeta_att[t]], tau_att[zeta_att[t]])
    def_raw[t] ~ dnorm(mu_def[zeta_def[t]], tau_def[zeta_def[t]])

    zeta_att[t] ~ dcat(p_att[])
    zeta_def[t] ~ dcat(p_def[])

    att[t] = att_raw[t]
    def[t] = def_raw[t]
  }

  att[T] = 0
  def[T] = 0

  tau_eff = 1/sigma_sq_eff ## sigma_sq_eff passado nos dados

  mu_att[1] ~ dnorm(0, tau_eff)T(-3,0) #I() truncado

```

```

mu_def[1] ~ dnorm(0, tau_eff)T( 0,3) #I() truncado

mu_att[2] = 0
mu_def[2] = 0

mu_att[3] ~ dnorm(0, tau_eff)T( 0,3) #I()
mu_def[3] ~ dnorm(0, tau_eff)T(-3,0) #I()


home ~ dnorm(0, tau_eff)

tau_att[1] ~ dgamma(a_att, b_att) ## a_att e b_att vem nos dados
tau_att[2] = 10000
tau_att[3] ~ dgamma(a_att, b_att)
tau_def[1] ~ dgamma(a_def, b_def)
tau_def[2] = 10000
tau_def[3] ~ dgamma(a_def, b_def)

sigma_att[1] = 1/sqrt(tau_att[1])
sigma_att[2] = 1/sqrt(tau_att[2])
sigma_att[3] = 1/sqrt(tau_att[3])
sigma_def[1] = 1/sqrt(tau_def[1])
sigma_def[2] = 1/sqrt(tau_def[2])
sigma_def[3] = 1/sqrt(tau_def[3])

p_att ~ ddirich(c(1, 1, 1))
p_def ~ ddirich(c(1, 1, 1))

}

```