

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA

Modelo Hierárquico Bayesiano para a previsão  
de resultados de futebol

**Matheus Bertoldi Diani**

**Orientador: Luis Ernesto Bueno Salasar**

Trabalho de Conclusão de Curso apresentado  
como parte dos requisitos para obtenção do  
título de Bacharel em Estatística.

**São Carlos**  
**Julho de 2025**



# Agradecimentos

Agradeço à minha família pela oportunidade de estudar na UFSCar, ao meu orientador pela ajuda importante nesta reta final e aos amigos que fiz nesta jornada. Toda essa vivência não apenas me trouxe conhecimento, mas me moldou de uma forma que levarei para toda a vida.

# Resumo

Este trabalho desenvolve e compara dois modelos estatísticos para prever resultados do Campeonato Brasileiro de futebol, o modelo hierárquico bayesiano e o modelo Poisson Independente. Ambas as abordagens modelam o número de gols marcados por cada equipe utilizando a distribuição de Poisson, com base em parâmetros de força ofensiva, defensiva e um efeito de mando de campo. A avaliação dos modelos foi realizada em duas frentes, incluindo uma análise de qualidade de ajuste aos dados da temporada completa e uma avaliação preditiva fora da amostra, usando um esquema de atualização sequencial para prever a segunda metade do campeonato.

Os resultados revelam um *trade-off* entre as duas abordagens. O modelo hierárquico, por seu efeito de regularização, produziu previsões mais estáveis e consistentes, mas apresentou uma suavização excessiva no ajuste, subestimando o desempenho das equipes nos extremos da tabela. Em contrapartida, o modelo Poisson Independente demonstrou maior sensibilidade às equipes, resultando em melhor ajuste geral, mas com maior variabilidade preditiva. Ambos os modelos, contudo, mostraram limitações, como a dificuldade em prever empates. Adicionalmente, uma análise de cenários demonstrou a capacidade dos modelos em capturar a evolução das probabilidades de título, G4 e rebaixamento ao longo da reta final do campeonato.

Por fim, discutem-se extensões para aprimorar as previsões. Sugere-se a exploração de distribuições mais flexíveis, como a Binomial Negativa para tratar a superdispersão, e modelos com inflação de empates. Uma frente de pesquisa promissora envolve a incorporação de informações do mercado de apostas, seja na construção de prioris informativas, como benchmark de desempenho ou na simulação do retorno financeiro de uma estratégia de apostas. Este estudo avança na estatística aplicada ao futebol, ao dissecar os desafios práticos e mapear as possibilidades que novas abordagens de modelagem oferecem para o campo.

**Palavras-chave:** *futebol, previsão, inferência bayesiana, modelo hierárquico, distribuição de Poisson, Brasileirão.*



# Sumário

<b>1</b>	<b>Introdução</b>	<b>7</b>
<b>2</b>	<b>Modelos Bayesianos para dados do Campeonato Brasileiro de Futebol</b>	<b>9</b>
2.1	Modelos adotados . . . . .	10
2.1.1	Modelo Hierárquico . . . . .	10
2.1.2	Modelo Poisson Independente . . . . .	13
2.2	Inferência Bayesiana . . . . .	14
2.2.1	Previsão Bayesiana vs. Frequentista . . . . .	14
2.2.2	Obtenção da Distribuição Preditiva via Simulação . . . . .	15
2.2.3	Método MCMC e Estimação dos Parâmetros . . . . .	16
2.2.4	Método de estimação via . . . . .	17
2.2.5	Crítérios de Convergência . . . . .	21
2.3	Métodos de avaliação da qualidade de ajuste . . . . .	22
2.4	Métodos de avaliação da qualidade preditiva . . . . .	23
2.4.1	Regras de escore ( <i>Scoring Rules</i> ) . . . . .	23
2.4.2	Medidas de desempenho do modelo . . . . .	24
2.4.3	Teste Kolmogorov-Smirnov . . . . .	26
2.4.4	Medidas de Concordância . . . . .	27
2.4.5	Diagrama de Calibração . . . . .	29
2.4.6	Análise de Cenários para Desfechos Finais . . . . .	31
<b>3</b>	<b>Análise de qualidade de ajuste dos modelos</b>	<b>33</b>
3.1	Análise Descritiva . . . . .	34
3.2	Análise da Cadeia . . . . .	37
3.3	Simulações Modelo Hierárquico . . . . .	37
3.3.1	Estimação dos Parâmetros . . . . .	37

3.3.2	Predição de Jogos e Probabilidades de Vitória . . . . .	41
3.3.3	Comparação entre Pontos Preditos e Observados . . . . .	42
3.3.4	Avaliação do Modelo Hierárquico . . . . .	43
3.4	Simulações Modelo Poisson Independente . . . . .	49
3.4.1	Estimação dos Parâmetros . . . . .	49
3.4.2	Predição de Jogos e Probabilidades de Vitória . . . . .	53
3.4.3	Comparação entre Pontos Preditos e Observados . . . . .	54
3.4.4	Avaliação do Modelo Poisson Independente . . . . .	54
3.5	Comparação Modelo Poisson Independente e Modelo Hierárquico . . . . .	59
3.5.1	Avaliação do Ajuste . . . . .	59
3.5.2	Teste Qui-Quadrado para Avaliação do Ajuste . . . . .	60
3.5.3	Comparação dos Pontos Acumulados . . . . .	61
3.5.4	Análise dos Gols Previstos e Observados . . . . .	63
<b>4</b>	<b>Análise da qualidade preditiva dos modelos</b>	<b>67</b>
4.1	O Conjunto de Dados para Avaliação . . . . .	68
4.2	Análise das Cadeias . . . . .	68
4.3	Regras de escore ( <i>Scoring Rules</i> ) . . . . .	69
4.4	Análise de desempenho do modelo . . . . .	72
4.5	Teste KS . . . . .	77
4.6	Medidas de Concordância . . . . .	78
4.7	Diagrama de Calibração . . . . .	80
4.8	Probabilidades para a Classificação Final . . . . .	80
4.8.1	Análise das Probabilidades Simuladas . . . . .	81
<b>5</b>	<b>Conclusão</b>	<b>87</b>
	<b>Referências Bibliográficas</b>	<b>89</b>
<b>A</b>	<b>Cadeias de Markov e MCMC</b>	<b>91</b>
A.1	Traceplots das cadeias de Markov . . . . .	91
<b>B</b>	<b>Modelo</b>	<b>95</b>
<b>C</b>	<b>Detalhamento da Distribuição a Posteriori</b>	<b>99</b>

# Capítulo 1

## Introdução

O Campeonato Brasileiro de Futebol, considerado a principal competição do esporte no Brasil, é amplamente reconhecido como uma das ligas mais competitivas do mundo. Composto por 20 equipes, o torneio adota o formato de pontos corridos, no qual cada time enfrenta todos os demais em partidas de ida e volta ao longo de 38 rodadas. O sistema de pontuação premia a vitória com três pontos e o empate com um, não conferindo pontos em caso de derrota. Ao término da competição, a equipe com maior pontuação é consagrada campeã. Além do título, as posições na tabela final definem as vagas para competições continentais de prestígio, como a Copa Libertadores para os quatro primeiros colocados, e determinam o rebaixamento das quatro equipes com menor desempenho para a Série B do ano seguinte.

O interesse pela previsão dos resultados das partidas é uma característica amplamente compartilhada entre os entusiastas do esporte, independentemente de seu nível de envolvimento. As discussões acerca dessas previsões são elementos centrais da cultura esportiva, variando desde análises informais e subjetivas realizadas por torcedores até debates mais aprofundados e técnicos conduzidos por especialistas, frequentemente veiculados em programas esportivos especializados ou em outras plataformas de mídia.

Ao longo das últimas décadas, diversos métodos estatísticos foram propostos para a previsão de resultados de futebol, abrangendo tanto abordagens frequentistas quanto bayesianas. [Karlis e Ntzoufras \(2003\)](#), por exemplo, propuseram o uso de um modelo de Poisson bivariado para capturar a relação entre os gols marcados por ambas as equipes em uma partida, destacando como essa abordagem pode ser usada para analisar e prever resultados. Além disso, [Rue e Salvesen \(2000\)](#) utilizou métodos probabilísticos para prever e analisar retrospectivamente o desempenho de equipes em ligas, enquanto

[Santos-Fernandez \(2019\)](#) revisou o uso de métodos bayesianos em esportes, enfatizando as vantagens dessa abordagem para lidar com a incerteza inerente e a complexidade dos dados esportivos.

Neste contexto, a estatística bayesiana vem ganhando espaço nos últimos anos como ferramenta eficaz na previsão de resultados esportivos, devido à sua flexibilidade em incorporar informações prévias e à capacidade de ajustar modelos para capturar características específicas dos eventos. Nos trabalhos, como o de [Baio e Blangiardo \(2010\)](#) os autores aplicaram um modelo Hierárquico bayesiano para prever resultados em competições de futebol, considerando os gols marcados pelas equipes como variáveis com distribuição de Poisson. O modelo utiliza uma estrutura log-linear para modelar as taxas de gols, incorporando fatores como forças ofensivas e defensivas das equipes e o impacto do mando de campo, tratados como efeitos aleatórios com distribuições normais. Ademais, por meio do nível Hierárquico do modelo, é possível capturar correlações entre esses fatores, introduzindo dependências latentes que melhoram as previsões.

Neste trabalho, buscamos aplicar e avaliar um modelo Hierárquico bayesiano inspirado em [Baio e Blangiardo \(2010\)](#) para prever os resultados do Campeonato Brasileiro de Futebol. A análise é estruturada para modelar o número de gols marcados pelas equipes, incorporando fatores como forças ofensivas e defensivas, bem como o efeito do mando de campo. Além disso, realiza-se uma comparação com o modelo Poisson Independente. Nesse modelo, assume-se que o número de gols marcados por cada equipe segue uma distribuição de Poisson, mas sem a estrutura hierárquica que permite regularizar os parâmetros das equipes. Dessa forma, no modelo de Poisson Independente, cada equipe é tratada isoladamente, sem considerar relações entre os times que poderiam capturar melhor a variabilidade e dependências presentes no campeonato.

A comparação entre o modelo Poisson Independente e o modelo Hierárquico possibilita a avaliação os impactos da inclusão de níveis adicionais de estrutura na modelagem dos dados. Em particular, por meio dessa análise, verifica-se se a abordagem hierárquica melhora a precisão preditiva e a interpretação dos parâmetros, oferecendo um ajuste mais refinado ao considerar correlação latente entre as equipes e a variação nos desempenhos ao longo da temporada.

## Capítulo 2

# Modelos Bayesianos para dados do Campeonato Brasileiro de Futebol

Neste capítulo, estabelecemos a fundação metodológica do trabalho, com o detalhamento da arquitetura dos modelos estatísticos e dos critérios para sua avaliação e comparação. Na Seção 2.1 apresentamos os dois modelos centrais da análise: o Hierárquico Bayesiano, com sua estrutura multinível, e o modelo Poisson Independente, que serve como referência. Na sequência, na Seção 2.2 aprofundamos a abordagem de inferência, cobrindo os princípios da predição bayesiana, os métodos de estimação via MCMC e os critérios para diagnóstico de convergência.

As seções seguintes são dedicadas à avaliação dos modelos. O foco da Seção 2.3 é a qualidade do ajuste aos dados já observados. Na Seção 2.4, detalhamos as métricas utilizadas para medir a capacidade preditiva, o que inclui regras de escore, métricas de classificação, testes de discriminação e diagramas de calibração. Adicionalmente, nesta seção, introduzimos a metodologia de análise de cenários, empregada para simular a reta final do campeonato e projetar as probabilidades de desfechos como título e rebaixamento. Por fim, na Seção 2.5, apresentamos a fonte do conjunto de dados do Campeonato Brasileiro de 2023.

## 2.1 Modelos adotados

### 2.1.1 Modelo Hierárquico

Utilizamos um modelo Hierárquico para representar o número de gols marcados pelas equipes, modelando-os como variáveis com distribuição de Poisson, condicionadas aos parâmetros que descrevem a média de gols esperada em cada jogo.

Seja  $y_{g1}$  o número de gols marcados pela equipe mandante e  $y_{g2}$  o número de gols marcados pela equipe visitante no  $g$ -ésimo jogo. Condicionalmente aos parâmetros  $\theta_{g1}$  e  $\theta_{g2}$ , assumimos que

$$y_{gj} \mid \theta_{gj} \sim \text{Poisson}(\theta_{gj}), \quad (2.1)$$

em que  $\theta_{gj}$  denota o número esperado de gols para a equipe o  $g$ -ésimo jogo para a equipe que joga como ( $j = 1$ ) mandante e ( $j = 2$ ) para a equipe visitante.

No modelo Hierárquico proposto,  $h(g)$  e  $a(g)$  indexam, respectivamente, o time mandante e o visitante no jogo  $g$ . Para facilitar a indexação, atribuímos os índices dos times foram atribuídos de acordo com o posicionamento de cada equipe na tabela final de classificação.

Os parâmetros  $\theta_{gj}$  são modelados por meio de uma estrutura log-linear que incorpora fatores relacionados ao desempenho ofensivo e defensivo das equipes, bem como ao efeito do mando de campo. A formulação logarítmica é descrita como

$$\log \theta_{g1} = \text{home} + \text{att}_{h(g)} + \text{def}_{a(g)}, \quad (2.2)$$

$$\log \theta_{g2} = \text{att}_{a(g)} + \text{def}_{h(g)}, \quad (2.3)$$

nessas equações,  $\text{home}$  representa o efeito fixo do mando de campo, assumido como constante para todas as equipes ao longo da temporada  $\text{att}_{h(g)}$  e  $\text{att}_{a(g)}$  indicam as forças ofensivas das equipes mandante e visitante, respectivamente. Da mesma forma,  $\text{def}_{h(g)}$  e  $\text{def}_{a(g)}$  representam as forças defensivas dessas equipes. Na Tabela 2.1, ilustramos como os dados de algumas partidas são estruturados para o modelo.

### Distribuições Priori e Identificação de Parâmetros

No contexto da inferência bayesiana, especificamos distribuições a priori para todos os parâmetros do modelo. No que segue, denotamos por  $\text{Normal}(\mu, \tau)$  a densidade Normal

Tabela 2.1: Rótulos dos times utilizados no modelo.

<b>g</b>	<b>Time Mandante</b>	<b>Time Visitante</b>	$h(g)$	$a(g)$	$y_{g1}$	$y_{g2}$
1	Palmeiras	Cuiaba	1	12	2	1
2	América-MG	Fluminense	20	7	0	3
3	Bragantino	Bahia	6	16	2	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
378	Santos	Fortaleza	17	10	1	2
379	São Paulo	Flamengo	11	4	1	0
380	Vasco	Bragantino	15	6	2	1

com média  $\mu$  e precisão  $\tau$  (inverso da variância). O efeito *home* foi modelado como um efeito fixo, utilizando uma distribuição Normal não informativa

$$home \sim \text{Normal}(0; 0,0001). \quad (2.4)$$

As forças ofensivas ( $att_t$ ) e defensivas ( $def_t$ ) de cada time  $t$  foram modeladas como efeitos aleatórios. Adotamos uma suposição de intercambialidade, o que significa que, a priori, não há distinção entre as equipes. Assim, assumimos que os parâmetros de cada time são realizações de uma mesma distribuição Normal geral

$$att_t \sim \text{Normal}(\mu_{att}, \tau_{att}), \quad def_t \sim \text{Normal}(\mu_{def}, \tau_{def}). \quad (2.5)$$

Para garantir a identificabilidade do modelo, impusemos restrições de soma-zero

$$\sum_{t=1}^{20} att_t = 0, \quad \sum_{t=1}^{20} def_t = 0. \quad (2.6)$$

Essa imposição das restrições de soma-zero é fundamental para fixar o referencial do modelo e garantir sua identificabilidade. Sem essas restrições, seria possível que uma mesma constante fosse somada simultaneamente a todos os parâmetros de ataque e defesa sem alterar as previsões, resultando em infinitas soluções equivalentes. Essa indeterminação faria com que o algoritmo de estimação, especialmente métodos baseados em *Markov Chain Monte Carlo* (MCMC), não conseguisse estabilizar os parâmetros em torno de um ponto fixo, prejudicando a convergência e gerando instabilidade na estimação. Além disso, a soma-zero confere interpretação direta aos parâmetros, permitindo que cada efeito ofensivo ou defensivo represente um desvio em relação à média geral do campeonato, facilitando a comparação entre as equipes.

Os hiperparâmetros dos efeitos ofensivos e defensivos ( $\mu_{att}, \mu_{def}, \tau_{att}, \tau_{def}$ ) foram mode-

lados com distribuições Normais para os parâmetros de localização e distribuições Gama para os parâmetros de precisão, utilizando-se a parametrização por forma ( $\alpha$ ) e taxa ( $\beta$ )

$$\mu_{att} \sim \text{Normal}(0; 0,0001), \quad \mu_{def} \sim \text{Normal}(0; 0,0001), \quad (2.7)$$

$$\tau_{att} \sim \text{Gama}(0,1; 0,1), \quad \tau_{def} \sim \text{Gama}(0,1; 0,1). \quad (2.8)$$

### Estrutura Hierárquica e Correlação Latente

A natureza hierárquica do modelo resulta em correlações latentes entre os jogos, uma vez que os parâmetros ofensivos e defensivos de cada equipe são compartilhados em todos os jogos que envolvem essa equipe. Assim, cada jogo contribui para a estimação dos hiperparâmetros  $\boldsymbol{\eta} = (\mu_{att}, \mu_{def}, \tau_{att}, \tau_{def})$ , que, por sua vez, determinam as taxas médias de gols esperados em toda a temporada.

A estrutura completa de dependências do modelo Hierárquico pode ser visualizada por meio do Gráfico Acíclico Direcionado (DAG, do inglês *Directed Acyclic Graph*) na Figura 2.1.

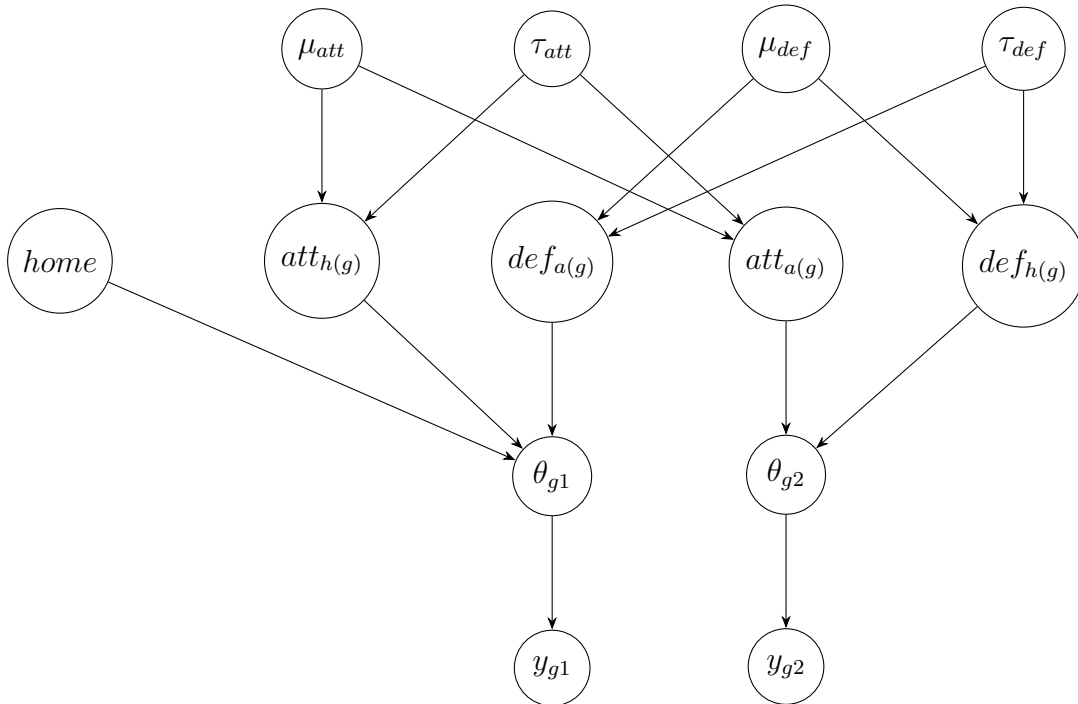


Figura 2.1: Estrutura de dependências do Modelo Hierárquico Bayesiano.

No diagrama da figura 2.1, os nós (círculos) representam as variáveis, e as setas indicam dependência condicional. No topo, os hiperparâmetros (como  $\mu_{att}$  e  $\tau_{att}$ ) influenciam as distribuições dos parâmetros de cada time ( $att_t$  e  $def_t$ ). Estes, por sua vez, juntamente com o efeito de mando de campo ( $home$ ), determinam a taxa de gols esperada ( $\theta_{g1}$  e  $\theta_{g2}$ ),

da qual os dados observados ( $y_{g1}$  e  $y_{g2}$ ) são uma realização. Por meio dessa estrutura, as equipes conseguem “emprestar força” estatística umas das outras, uma vez que todas são modeladas como provenientes de uma mesma distribuição populacional.

### 2.1.2 Modelo Poisson Independente

Para fins comparativos, também implementamos um modelo Poisson Independente para a previsão do número de gols marcados em cada partida. Nesse modelo, sejam  $y_{1g}$  e  $y_{2g}$  o número de gols marcados pelo time mandante e visitante, respectivamente, no jogo  $g$ . Assume-se que essas variáveis seguem distribuições de Poisson com taxa média de gols constantes, isto é

$$y_{gj} \mid \lambda_{gj} \sim \text{Poisson}(\lambda_{gj}). \quad (2.9)$$

As médias de gols esperadas são especificadas de forma simples, considerando um efeito global de mando de campo e parâmetros fixos de ataque e defesa para cada time. Assim, define-se

$$\log \lambda_{1g} = \textit{home} + \textit{att}_{h(g)} + \textit{def}_{a(g)}, \quad (2.10)$$

$$\log \lambda_{2g} = \textit{att}_{a(g)} + \textit{def}_{h(g)}. \quad (2.11)$$

Especificamos distribuições a priori não informativas para os parâmetros. Tanto o efeito de mando de campo (*home*) quanto os parâmetros de ataque ( $\textit{att}_t$ ) e defesa ( $\textit{def}_t$ ) de cada time, atribuímos uma priori Normal com média 0 e precisão 0,0001.

Diferentemente do modelo Hierárquico, essa abordagem não permite que os parâmetros de ataque e defesa variem de forma específica entre os times, nem incorpora a heterogeneidade inerente ao desempenho individual, sendo, portanto, menos flexível. No entanto, sua simplicidade a torna uma referência importante para avaliar os ganhos em precisão preditiva e interpretabilidade obtidos ou não pela modelagem hierárquica. Além disso, para garantir a comparabilidade, também impusemos as mesmas restrições de soma-zero.

A estrutura deste modelo é visualmente representada pelo diagrama na Figura 2.2. De forma semelhante ao modelo anterior, os parâmetros de ataque, defesa e mando de

campo determinam as taxas de gols  $(\lambda_{g1}, \lambda_{g2})$ , que por sua vez geram os dados observados  $(y_{g1}, y_{g2})$ .

A diferença crucial em relação ao modelo Hierárquico (Figura 2.1) é a ausência de um nível superior de hiperparâmetros. Nesta estrutura, os parâmetros de ataque e defesa de cada time são tratados como entidades fixas e independentes, sem uma distribuição comum que os relacione. É essa característica que define o modelo como “independente” e impede o compartilhamento de informação entre as equipes.

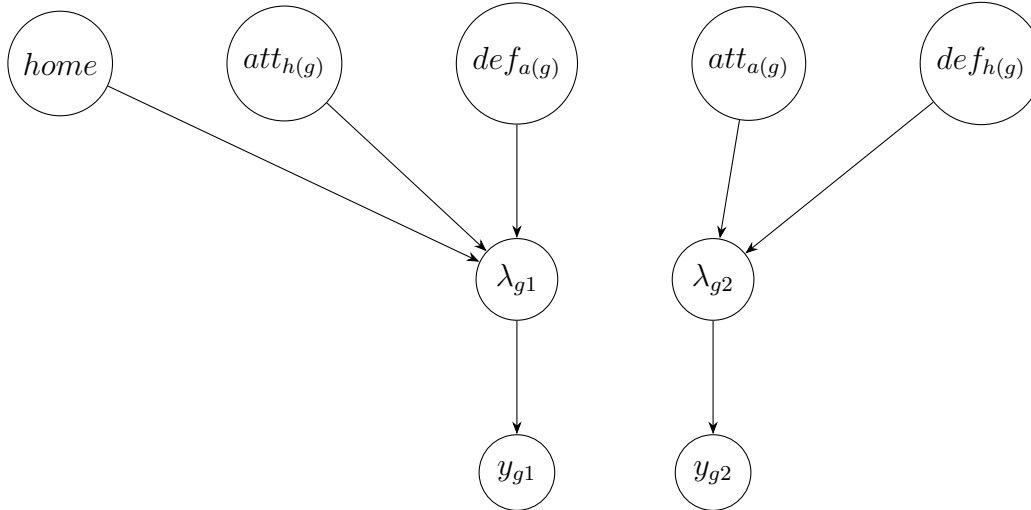


Figura 2.2: Estrutura de dependências do Modelo Independente Bayesiano.

## 2.2 Inferência Bayesiana

### 2.2.1 Previsão Bayesiana vs. Frequentista

A abordagem frequentista de previsão consiste em ajustar um modelo aos dados de treinamento para obter um estimador pontual  $\hat{\theta}$ . Para prever novos desfechos, assume-se que o verdadeiro parâmetro  $\theta$  é exatamente igual a  $\hat{\theta}$ , desprezando-se toda a incerteza inerente à sua estimação. Esse procedimento pode gerar uma falsa sensação de precisão, pois encapsula toda a variabilidade apenas na variância das observações, sem refletir a incerteza associada à estimação dos parâmetros.

Em contraste, a abordagem bayesiana trata o vetor de parâmetros  $\theta$  como uma variável aleatória, cuja incerteza após observar os dados  $D$  é descrita pela distribuição a posteriori,  $p(\theta | D)$ . A propagação dessa incerteza para as previsões de um novo dado  $x_{\text{novo}}$  é a essência da previsão bayesiana e ocorre formalmente por meio da distribuição preditiva a posteriori. Esta é definida pela seguinte integral

$$p(x_{\text{nov}} | D) = \int p(x_{\text{nov}} | \theta) p(\theta | D) d\theta. \quad (2.12)$$

Conforme a equação, a distribuição preditiva é formada pela média de todas as distribuições preditivas condicionais  $p(x_{\text{nov}} | \theta)$ , ponderada pela plausibilidade de cada  $\theta$  segundo a distribuição a posteriori  $p(\theta | D)$  (Santos-Fernandez, 2019). Ao invés de condicionar a predição a um único valor  $\hat{\theta}$ , a abordagem bayesiana integra sobre todos os valores possíveis de  $\theta$ , o que torna suas previsões costumam apresentar faixas de credibilidade mais realistas e informativas.

## 2.2.2 Obtenção da Distribuição Preditiva via Simulação

Na prática, a integral que define a distribuição preditiva raramente possui uma solução analítica fechada. No entanto, podemos aproximá-la por meio de um algoritmo de simulação em duas etapas. Para cada iteração  $s = 1, \dots, S$ :

1. **Amostragem da posteriori:** Gera-se um vetor de parâmetros  $\theta^{(s)}$  da distribuição a posteriori,  $p(\theta | D)$ .
2. **Amostragem da preditiva condicional:** Gera-se uma nova observação  $x_{\text{nov}}^{(s)}$  da distribuição do modelo condicionada ao parâmetro recém-gerado,  $p(x_{\text{nov}} | \theta^{(s)})$ .

Ao repetir esse processo  $S$  vezes, obtém-se uma amostra empírica da distribuição preditiva,  $\{x_{\text{nov}}^{(1)}, \dots, x_{\text{nov}}^{(S)}\}$ . Este conjunto de valores simulados é a aproximação da distribuição  $p(x_{\text{nov}} | D)$  e pode ser usado diretamente para fins de inferência.

Em termos práticos:

- Cada  $x_{\text{nov}}^{(s)}$  é um cenário de previsão de gols (por exemplo, gols do mandante e do visitante em um jogo futuro), gerado condicionalmente ao vetor de parâmetros  $\theta^{(s)}$ .
- A distribuição preditiva completa, representada pela amostra  $\{x_{\text{nov}}^{(1)}, \dots, x_{\text{nov}}^{(S)}\}$ , permite extrair quantis, intervalos de credibilidade e probabilidades de eventos específicos (por exemplo, a probabilidade de vitória do mandante).

Dessa forma, a incerteza nos parâmetros  $\theta$  é naturalmente propagada para as previsões, fornecendo intervalos de credibilidade que refletem tanto a variabilidade dos dados quanto a incerteza de estimação (Gelman *et al.*, 2013).

### 2.2.3 Método MCMC e Estimação dos Parâmetros

A inferência bayesiana baseia-se na construção da distribuição a posteriori dos parâmetros de interesse, a qual é obtida por meio da aplicação do Teorema de Bayes (Gelman *et al.*, 2013).

$$p(\theta | D) = \frac{p(D | \theta) \cdot p(\theta)}{p(D)}. \quad (2.13)$$

Em que:

- $p(\theta | D)$  é a distribuição a posteriori dos parâmetros  $\theta$  dado os dados observados  $D$ ;
- $p(D | \theta)$  é a função de verossimilhança;
- $p(\theta)$  é a distribuição a priori;
- $p(D)$  é a distribuição marginal dos dados.

Na maioria dos modelos estatísticos com múltiplos parâmetros e estrutura hierárquica, como o caso deste trabalho, a integral envolvida no cálculo de  $p(D)$  não possui forma analítica fechada. Assim, recorreremos ao Método MCMC para aproximar a a posteriori (Gelman *et al.*, 2013).

#### Cadeias de Markov

Uma *Cadeia de Markov* é um processo estocástico que gera uma sequência de valores  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n)}$  em que o próximo valor da cadeia depende apenas do estado atual, e não de estados anteriores. Isto é

$$P(\theta^{(t+1)} | \theta^{(t)}, \theta^{(t-1)}, \dots, \theta^{(1)}) = P(\theta^{(t+1)} | \theta^{(t)}). \quad (2.14)$$

#### MCMC: Amostragem via Cadeias de Markov

O objetivo ao se utilizar MCMC é construir uma cadeia de Markov cuja distribuição estacionária convirja para a distribuição a posteriori  $p(\theta | D)$ . Como a cadeia geralmente parte de um ponto arbitrário, as iterações iniciais não são consideradas amostras válidas da distribuição de interesse. Este “período” inicial de convergência, conhecido como *burn-in* (ou aquecimento), é portanto descartado.

Após essa fase, o resultado principal é a obtenção de uma amostra de valores,  $\{\theta^{(1)}, \dots, \theta^{(S)}\}$ , que pode ser tratada como sendo representativa da distribuição a posteriori. Uma

vez de posse dessa amostra, a inferência sobre qualquer característica da distribuição se torna uma aplicação direta do método de Monte Carlo. Assim, a amostra é utilizada para estimar as características de interesse do modelo, como médias, variâncias, quantis e intervalos de credibilidade, aproximando, por exemplo, o valor esperado de um parâmetro por meio de sua média amostral.

No trabalho, o *software OpenBUGS* (Lunn *et al.*, 2009) foi utilizado para executar amostragem MCMC nos modelos especificados. No *OpenBUGS*, o algoritmo mais apropriado entre essas abordagens é utilizado automaticamente, dependendo da estrutura de cada parâmetro do modelo.

### **Importância no Contexto do Trabalho**

Nos modelos bayesianos adotados neste trabalho, a presença de múltiplos níveis de parâmetros latentes e de dependências cruzadas entre ataque e defesa produz distribuições a posteriori cuja forma analítica é intratável. Por essa razão, o uso de algoritmos de MCMC torna-se indispensável, pois apenas por meio da geração de amostras sucessivas é possível explorar o espaço de parâmetros de forma eficiente e consistente, sem recorrer a aproximações que possam comprometer a precisão das estimativas.

Ao fornecer uma amostra representativa da distribuição a posteriori, o MCMC é o motor que permite a propagação completa da incerteza sobre os parâmetros. Essa amostra da posteriori é o ingrediente fundamental para, em um passo de simulação adicional, construir a distribuição preditiva. A capacidade de gerar amostras da posteriori é, portanto, essencial para que as previsões de gols e resultados futuros reflitam não apenas o comportamento médio do modelo, mas também a variabilidade inerente aos dados e aos parâmetros. Em suma, a adoção do MCMC neste trabalho justifica-se tanto pela superação das limitações analíticas quanto por ser a ferramenta que viabiliza a geração de distribuições preditivas robustas e intervalos de credibilidade realistas para embasar as conclusões do estudo.

## **2.2.4 Método de estimação via**

### **Poisson Hierarquica**

Em qualquer problema de inferência bayesiana, o objetivo central é obter a distribuição a posteriori dos parâmetros de interesse. Essa distribuição é definida pelo Teorema de Bayes

$$p(\theta | D) = \frac{p(D | \theta) p(\theta)}{\int p(D | \theta) p(\theta) d\theta}. \quad (2.15)$$

Para nosso modelo, o vetor completo de parâmetros é  $\theta = (\text{home}, \text{att.star}[1, \dots, 20], \text{def.star}[1, \dots, 20])$ . A expressão para o produto  $p(D | \theta) p(\theta)$ , que define a forma da distribuição a posteriori (sem a constante de normalização), é dada por

$$\begin{aligned} p(\theta | D) \propto & \underbrace{\prod_{g=1}^{380} [\text{Poisson}(y_{g1} | e^{\lambda_{g1}}) \times \text{Poisson}(y_{g2} | e^{\lambda_{g2}})]}_{\text{Verossimilhança dos dados}} \\ & \times \underbrace{\left[ \prod_{t=1}^{20} \text{Normal}(\text{att}_t | \mu_{\text{att}}, \tau_{\text{att}}) \right] \times \left[ \prod_{t=1}^{20} \text{Normal}(\text{def}_t | \mu_{\text{def}}, \tau_{\text{def}}) \right]}_{\text{Priors para os efeitos das equipes}} \\ & \times \underbrace{\text{Normal}(\text{home} | 0; 0, 0001)}_{\text{Prior para o efeito de mando}} \\ & \times \underbrace{\text{Normal}(\mu_{\text{att}} | 0; 0, 0001) \times \text{Gama}(\tau_{\text{att}} | 0, 1; 0, 1)}_{\text{Hiperpriors de ataque}} \\ & \times \underbrace{\text{Normal}(\mu_{\text{def}} | 0; 0, 0001) \times \text{Gama}(\tau_{\text{def}} | 0, 1; 0, 1)}_{\text{Hiperpriors de defesa}}. \end{aligned} \quad (2.16)$$

O desafio computacional surge quando os componentes dessa equação, particularmente a integral de 2.16, não podem ser resolvidos analiticamente. Este é precisamente o caso de modelos como o hierárquico adotado neste trabalho.

O denominador de expressão 2.15 é uma integral de normalização em um espaço de 45 dimensões (uma para cada parâmetro estocástico, ver Apêndice C). Além disso, a função de verossimilhança Poisson combinada com prioris Normais gera um integrando para o qual não existe uma primitiva fechada nas tabelas usuais de integrais. Logo, não há conjugação que reduza esta integral a uma forma conhecida (como Gama ou Normal) e o cálculo direto torna-se inviável.

Para contornar essa limitação, utilizamos o *software OpenBUGS*, que implementa algoritmos de MCMC para gerar amostras da distribuição a a posteriori e simular da distribuição preditiva.

O núcleo de amostragem do *OpenBUGS* percorre iterativamente cada nó estocástico do modelo, escolhendo o algoritmo de atualização mais adequado para cada parâmetro:

- **Conjugate Normal updater (Gibbs)** (Geman e Geman, 1984) para as médias dos efeitos de time  $\mu_{\text{att}} \sim N(0; 0,0001)$  e  $\mu_{\text{def}} \sim N(0; 0,0001)$ , cuja a posteriori condicional também é Normal e é amostrada diretamente.
- **Conjugate Gamma updater (Gibbs)** para as precisões  $\tau_{\text{att}} \sim \text{Gama}(0,01; 0,01)$  e  $\tau_{\text{def}} \sim \text{Gama}(0,01; 0,01)$ , cuja a posteriori condicional é uma Gama também amostrada diretamente.
- **Adaptive Metropolis updater** (Haario *et al.*, 2001) para o intercepto *home* e para os efeitos de ataque e defesa. Cada proposta é

$$\phi' \sim \mathcal{N}(\phi^{(s-1)}, \sigma_{\text{prop}}^2), \quad (2.17)$$

com  $\sigma_{\text{prop}}$  ajustado automaticamente na fase de adaptação,, com o objetivo de manter a taxa de aceitação em um nível considerado ótimo, geralmente em torno de 20% a 50% (Gilks *et al.*, 1996). A proposta é aceita com a razão de Metropolis–Hastings clássica (Metropolis *et al.*, 1953; Hastings, 1970).

$$r = \min\left(1, \frac{p(D | \phi') p(\phi')}{p(D | \phi) p(\phi)}\right). \quad (2.18)$$

O procedimento de cada iteração  $s \geq 1$  é, resumidamente:

1. Para cada nó estocástico:  $\mu_{\text{att}}$ ,  $\mu_{\text{def}}$ ,  $\tau_{\text{att}}$ ,  $\tau_{\text{def}}$ , *home*, *att.star*[*t*], *def.star*[*t*], aplica-se o algoritmo de atualização apropriado, aceitando ou rejeitando propostas conforme necessário.
2. Recalculam-se em bloco os nós determinísticos *att*[*t*], *def*[*t*],  $\theta[g, \cdot]$ .
3. Por fim, para cada jogo  $g = 1, \dots, 380$ , um novo placar  $y_{\text{pred},g}^{(s)} = (y_{\text{pred},g1}^{(s)}, y_{\text{pred},g2}^{(s)})$  é gerado a partir da distribuição preditiva condicional:

$$y_{\text{pred},g1}^{(s)} \sim \text{Poisson}(\theta_{g1}^{(s)})$$

$$y_{\text{pred},g2}^{(s)} \sim \text{Poisson}(\theta_{g2}^{(s)})$$

O conjunto desses 380 placares simulados resulta em um campeonato completo para a iteração  $s$ .

Dessa forma, o *OpenBUGS* combina atualizações de Gibbs nas partes conjugadas do modelo, Metropolis–Hastings adaptativo nos parâmetros não-conjugados e amostragem direta nos preditivos, produzindo após o *burn-in* uma amostra representativa da posteriori  $p(\theta \mid D)$  e da preditiva  $p(x_{\text{novo}} \mid D)$ , o que permite obter estimativas pontuais, intervalos de credibilidade e previsões de resultados futuros.

### Poisson Independente

No modelo Independente, cada parâmetro de ataque e defesa possui priori fixa e não há estrutura hierárquica compartilhada. A distribuição a posteriori conjunta dos parâmetros  $\theta = (\text{home}, \text{att.star}[1, \dots, 20], \text{def.star}[1, \dots, 20])$  é dada pela 2.15, e não possui forma analítica fechada devido à não-conjugação entre a função de verossimilhança Poisson e as prioris Normais.

Para cada parâmetro, o *OpenBUGS* utiliza:

- ***Adaptive Metropolis updater*** para os coeficientes *home*, *att.star*[*t*] e *def.star*[*t*].

Cada proposta é

$$\theta' \sim \mathcal{N}(\theta^{(s-1)}, \sigma_{\text{prop}}^2), \quad (2.19)$$

com  $\sigma_{\text{prop}}$  ajustado durante a adaptação para manter a taxa de aceitação ideal. A aceitação segue a razão de Metropolis–Hastings

$$r = \min\left(1, \frac{p(D \mid \theta') p(\theta')}{p(D \mid \theta) p(\theta)}\right). \quad (2.20)$$

O ciclo de cada iteração  $s \geq 1$  é:

1. Para cada nó estocástico: *home*, *att.star*[*t*], *def.star*[*t*], aplica-se o *Adaptive Metropolis 1d updater*, aceitando ou rejeitando propostas conforme necessário.
2. Recalculam-se em bloco os nós determinísticos *att*[*t*], *def*[*t*],  $\theta[g, \cdot]$ .
3. Gera-se uma única vez, para cada jogo a amostra preditiva, de forma similar ao modelo Hierárquico, resultando num campeonato simulado por iteração.

Dessa forma, o *OpenBUGS* produz, após o *burn-in*, uma amostra representativa de  $p(\theta \mid D)$  e de  $p(y_{\text{novo}} \mid D)$ , permitindo obter estimativas pontuais, intervalos de credibilidade e previsões de resultados futuros.

## 2.2.5 Critérios de Convergência

Para avaliar a convergência das simulações MCMC para a distribuição estacionária de interesse, adotou-se como principal ferramenta quantitativa o critério de Gelman-Rubin (Gelman *et al.*, 2013). A escolha deste método se justifica por sua abordagem robusta, que se diferencia de diagnósticos baseados em uma única cadeia. Enquanto critérios univariados (como o de Geweke ou Raftery-Lewis) analisam as propriedades de uma sequência individual, o critério de Gelman-Rubin opera de forma comparativa, avaliando se múltiplas cadeias, executadas em paralelo e iniciadas a partir de pontos diferentes e dispersos, convergem para a mesma distribuição.

A lógica fundamental é que, se as cadeias exploraram adequadamente o espaço de parâmetros e atingiram a distribuição alvo, a variabilidade entre elas deve ser pequena em comparação com a variabilidade dentro de cada uma. Uma grande diferença entre essas variâncias indicaria que as cadeias não se misturaram bem ou ficaram presas em modos locais distintos, sinalizando a falta de convergência. A estatística  $\hat{R}$ , quantifica essa relação, fornecendo uma medida para essa avaliação.

O critério de Gelman-Rubin é calculado para cada parâmetro do modelo como *home*, *att* de um time, de forma individual. Para isso, o algoritmo MCMC é executado para gerar  $m$  cadeias de Markov independentes, cada uma partindo de um valor inicial (ou “chute”) diferente para cada parâmetro. Após o *burn-in*, cada cadeia  $i$  (para  $i = 1, \dots, m$ ) fornece  $n$  valores para os parâmetros (*home*, *att*, etc). Denotando o  $j$ -ésimo valor da  $i$ -ésima cadeia por  $X_{i,j}$ , calculam-se a média ( $\bar{X}_i$ ) e a variância ( $s_i^2$ ) para cada cadeia

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{i,j} \quad \text{e} \quad s_i^2 = \frac{1}{n-1} \sum_{j=1}^n (X_{i,j} - \bar{X}_i)^2. \quad (2.21)$$

Adicionalmente, calcula-se a média global das médias de cada cadeia,  $\bar{X} = \frac{1}{m} \sum_{i=1}^m \bar{X}_i$ . Com esses componentes, definem-se as variâncias dentro e entre as cadeias,  $W$  e  $B$ , respectivamente

$$W = \frac{1}{m} \sum_{i=1}^m s_i^2, \quad B = \frac{n}{m-1} \sum_{i=1}^m (\bar{X}_i - \bar{X})^2. \quad (2.22)$$

A variância total do parâmetro ( $\hat{V}$ ) é então estimada como uma combinação de  $W$  e  $B$ , e a estatística final,  $\hat{R}$ , é calculada da seguinte forma

$$\hat{V} = \frac{n-1}{n}W + \frac{1}{n}B, \quad \hat{R} = \sqrt{\frac{\hat{V}}{W}}. \quad (2.23)$$

Os componentes dessa expressão podem ser interpretados da seguinte forma:

- ***W* (*Within-chain variance*):** É a média das variâncias de cada cadeia ( $s_i^2$ ), medindo a variabilidade média *dentro* de cada sequência de simulação.
- ***B* (*Between-chain variance*):** É a variância entre as médias das diferentes cadeias ( $\bar{X}_i$ ), capturando a variabilidade *entre* as sequências. Se as cadeias estiverem explorando regiões diferentes,  $B$  será alto.
- **$\hat{V}$  (**Variância total estimada**):** É uma estimativa da variância real do parâmetro, combinando as fontes de variação de dentro ( $W$ ) e de entre ( $B$ ) as cadeias.
- **$\hat{R}$  (**Fator de Redução da Escala Potencial**):** Compara a variância total estimada ( $\hat{V}$ ) com a variância média interna ( $W$ ). Se as cadeias convergiram, a variância entre elas ( $B$ ) será pequena, fazendo com que  $\hat{V} \approx W$  e, conseqüentemente,  $\hat{R} \approx 1$ , indicando a convergência.

Além disso, realizamos inspeção visual dos *traceplots* das cadeias para os parâmetros, para observar se há algum problema de convergência.

## 2.3 Métodos de avaliação da qualidade de ajuste

Para avaliar e validar o desempenho preditivo dos modelos, utilizamos uma abordagem baseada em análises gráficas e tabelas que sintetizam a comparação entre os valores preditos e os observados. De maneira geral, os gráficos foram elaborados para comparar a evolução dos pontos acumulados das equipes, a distribuição do número total de gols simulados (tanto de forma global quanto segregada por mando de campo) e as discrepâncias entre os valores preditos e observados, tanto por jogo quanto por equipe. Assim é possível identificar vieses e verificar se as simulações reproduzem adequadamente a dinâmica dos resultados reais. A comparação entre o modelo Poisson Independente e o modelo Hierárquico baseia-se na verificação de que os valores observados se enquadrem dentro das distribuições preditivas geradas, tanto em termos de totais (gols e pontos) quanto na evolução ao longo das rodadas, fornecendo uma avaliação abrangente da calibração e da precisão preditiva dos modelos.

## 2.4 Métodos de avaliação da qualidade preditiva

### 2.4.1 Regras de escore (*Scoring Rules*)

A avaliação do desempenho fora da amostra (*out-of-sample predictions*) seguiu um esquema de atualização sequencial. O processo começou com o ajuste do modelo aos dados das 19 primeiras rodadas para prever a 20<sup>a</sup>. A partir daí, a cada nova rodada, os resultados observados eram incorporados ao conjunto de dados, e o modelo era reajustado para prever a rodada subsequente. Este procedimento foi repetido até a 38<sup>a</sup> rodada, garantindo que cada previsão utilizasse apenas informações até a rodada imediatamente anterior àquela que se deseja prever..

Depois de obter as previsões, é necessário avaliar a qualidade das probabilidades geradas. Para isso, utilizam-se as chamadas regras de escore (*scoring rules*), que medem a proximidade entre uma distribuição de probabilidades prevista e o resultado que de fato ocorreu.

Considere  $P_g = (p_{g,1}, p_{g,2}, p_{g,3})$  o vetor com as probabilidades previstas para os três desfechos do jogo  $g$  (vitória do mandante, empate e vitória do visitante, respectivamente). De forma análoga, seja  $O_g = (o_{g,1}, o_{g,2}, o_{g,3})$  o vetor indicativo do resultado observado, no qual o elemento correspondente ao desfecho que ocorreu assume o valor 1 e os demais são 0.

A proximidade entre o vetor predito  $P_g$  e o resultado observado  $O_g$  é quantificada por meio das regras de escore detalhadas a seguir.

**Brier Score** Sejam  $p_{g,k}$  as probabilidades previstas para os três possíveis desfechos de uma partida  $g$ , com  $k \in \{1, 2, 3\}$  representando, respectivamente, vitória do mandante, empate e vitória do visitante. O resultado observado é codificado por um vetor indicativo  $o_{g,k}$ , em que  $o_{g,k} = 1$  se o resultado  $k$  ocorreu e 0 caso contrário.

O *Brier Score* para a partida  $g$  é dado por (Brier, 1950)

$$BS_g = \sum_{k=1}^3 (p_{g,k} - o_{g,k})^2. \quad (2.24)$$

Essa métrica corresponde ao erro quadrático entre as probabilidades previstas e o resultado observado, penalizando igualmente todas as discrepâncias. Quanto menor o valor de  $BS_g$ , melhor o ajuste das probabilidades atribuídas à partida.

**Logarithmic Score** O *Logarithmic Score*, proposto por Good (1952), é definido por

$$\text{LS}_g = - \sum_{k=1}^3 o_{g,k} \log(p_{g,k} + \varepsilon), \quad (2.25)$$

em que  $\varepsilon > 0$  é um termo de correção numérica que impede a avaliação de  $\log 0$ . Como apenas o componente associado ao resultado efetivamente observado possui  $o_{g,k} = 1$ . Valores menores indicam que o modelo conferiu maior probabilidade ao ocorrido, enquanto previsões confiantes e incorretas são penalizadas de forma mais severa.

**Spherical Score** O *Spherical Score*, detalhado em trabalhos como o de Staël von Holstein (1970), é calculado por

$$\text{SS}_g = \frac{\sum_{k=1}^3 o_{g,k} p_{g,k}}{\sqrt{\sum_{j=1}^3 p_{g,j}^2}} = \frac{p_{g,k^*}}{\|p_g\|_2}, \quad (2.26)$$

onde  $k^*$  é o índice do resultado efetivamente ocorrido e  $\|p_g\|_2$  representa a norma Euclidiana do vetor de probabilidades.

Com valores entre 0 e 1, este escore recompensa o modelo por atribuir alta probabilidade ao desfecho correto. Sua característica principal é a normalização pelo módulo de  $p_g$ , um mecanismo que premia previsões focadas na categoria certa e penaliza aquelas que se mostram excessivamente dispersas.

Além das regras de scoring probabilístico apresentadas, utilizamos também técnicas de avaliação de classificação para capturar o desempenho categórico das previsões.

## 2.4.2 Medidas de desempenho do modelo

Complementarmente às regras de escore, avaliou-se o desempenho categórico das previsões por meio da matriz de confusão. A partir dessa tabela de contingência foram calculadas as métricas de *acurácia*, *recall* (sensibilidade), *precision* (precisão) e *F1-score*, fornecendo uma síntese dos acertos e erros de classificação do modelo.

Denotando por TP, FP, FN e TN o número total de verdadeiros positivos, falsos positivos, falsos negativos e verdadeiros negativos (agregados sob o esquema *one-vs-all*, ou um-contra-todos, para o problema multiclasse), as métricas são definidas por

$$\text{Acurácia} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}. \quad (2.27)$$

$$\text{Precisão} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (2.28)$$

$$\text{F1-score} = \frac{2 \cdot \text{Precisão} \cdot \text{Recall}}{\text{Precisão} + \text{Recall}}. \quad (2.29)$$

Para ilustrar a aplicação dessas métricas, considere a seguinte matriz de confusão 2.2 hipotética para um conjunto de 100 jogos previstos.

Tabela 2.2: Exemplo de Matriz de Confusão Hipotética

		Classe Observada (Real)			Total Predito
		Mandante	Empate	Visitante	
Classe Predita	Mandante	<b>35</b>	10	15	60
	Empate	2	<b>5</b>	3	10
	Visitante	8	10	<b>12</b>	30
Total Observado		45	25	30	<b>100</b>

A partir desta tabela, a acurácia global do modelo é a soma dos acertos (a diagonal principal) dividida pelo total de jogos

$$\text{Acurácia} = \frac{35 + 5 + 12}{100} = 0,52. \quad (2.30)$$

Para calcular as outras métricas no esquema multiclasse, focamos em uma classe de interesse. Por exemplo, para a classe Mandante:

- **TP (Verdadeiros Positivos):** O modelo previu Mandante e o resultado foi Mandante.  $\text{TP} = 35$ .
- **FP (Falsos Positivos):** O modelo previu Mandante, mas o resultado foi outro (Empate ou Visitante).  $\text{FP} = 10 + 15 = 25$ .
- **FN (Falsos Negativos):** O resultado foi Mandante, mas o modelo previu outra coisa.  $\text{FN} = 2 + 8 = 10$ .

Com esses valores, calculamos as métricas para a classe Mandante

$$\text{Precisão} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{35}{35 + 25} = \frac{35}{60} \approx 0,583. \quad (2.31)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{35}{35 + 10} = \frac{35}{45} \approx 0,778. \quad (2.32)$$

$$\text{F1-score} = \frac{2 \cdot 0,583 \cdot 0,778}{0,583 + 0,778} \approx 0,666. \quad (2.33)$$

Este mesmo processo é repetido para as classes Empate e Visitante para uma avaliação completa.

### 2.4.3 Teste Kolmogorov-Smirnov

Para investigar o poder de discriminação das probabilidades previstas em cada classe (Vitória do Mandante, Empate e Vitória do Visitante), aplicamos o teste de Kolmogorov-Smirnov (KS) para duas amostras (Conover, 1999). Suponha que, para uma dada classe  $c$ ,

- $x_1 = p_{g,c}$ : probabilidades previstas de  $c$  dos jogos em que  $c$  ocorreu,
- $x_2 = p_{g,c}$ : probabilidades previstas de  $c$  dos jogos em que  $c$  não ocorreu.

Definimos as distribuições acumuladas empíricas

$$\hat{F}_1(u) = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{1}\{x_{1,i} \leq u\}, \quad \hat{F}_2(u) = \frac{1}{n_2} \sum_{j=1}^{n_2} \mathbf{1}\{x_{2,j} \leq u\}, \quad (2.34)$$

em que  $n_1$  e  $n_2$  são os respectivos tamanhos dos vetores de  $x_1$  e  $x_2$ . O teste KS calcula

$$D = \sup_{u \in [0,1]} |\hat{F}_1(u) - \hat{F}_2(u)|, \quad (2.35)$$

isto é, a maior distância vertical entre as duas curvas de FDA.

Em termos práticos, a estatística  $D$  quantifica a máxima diferença entre a fração de previsões abaixo de um corte  $u$  nos jogos positivos e nos jogos negativos. O p-valor associado ao teste KS fornece a probabilidade de observar uma distância  $D$  tão grande sob a hipótese nula de que ambas as amostras vêm da mesma distribuição. Valores de p-valor abaixo de 0,05 indicam discriminação estatisticamente significativa entre as probabilidades previstas para eventos que ocorreram contra os que não ocorreram.

Para ilustrar o cálculo, considere um exemplo simplificado. Suponha que, para a classe “Vitória do Mandante”, tenhamos um conjunto de previsões para 4 jogos em que o mandante de fato venceu ( $x_1$ ) e para 5 jogos em que ele não venceu ( $x_2$ ):

- $x_1 = \{0, 7; 0, 5; 0, 9; 0, 8\}$  (probabilidades atribuídas aos acertos)
- $x_2 = \{0, 2; 0, 6; 0, 1; 0, 3; 0, 4\}$  (probabilidades atribuídas aos erros)

O primeiro passo é ordenar os valores e calcular as funções de distribuição acumulada empírica ( $\hat{F}_1(u)$  e  $\hat{F}_2(u)$ ) para cada ponto de probabilidade  $u$  presente nos dois conjuntos. A estatística  $D$  é o valor máximo da diferença absoluta entre essas duas distribuições, como detalhado na Tabela 2.3.

Tabela 2.3: Cálculo da Estatística D do Teste KS para o Exemplo.

Valor de $u$	$\hat{F}_1(u)$ (N=4)	$\hat{F}_2(u)$ (N=5)	$ \hat{F}_1(u) - \hat{F}_2(u) $
0,1	0/4 = 0,00	1/5 = 0,20	0,20
0,2	0/4 = 0,00	2/5 = 0,40	0,40
0,3	0/4 = 0,00	3/5 = 0,60	0,60
0,4	0/4 = 0,00	4/5 = 0,80	<b>0,80</b>
0,5	1/4 = 0,25	4/5 = 0,80	0,55
0,6	1/4 = 0,25	5/5 = 1,00	0,75
0,7	2/4 = 0,50	5/5 = 1,00	0,50
0,8	3/4 = 0,75	5/5 = 1,00	0,25
0,9	4/4 = 1,00	5/5 = 1,00	0,00

Analisando a última coluna da Tabela 2.3, a maior diferença encontrada é de 0,80. Portanto, para este exemplo, a estatística do teste é  $D = 0,80$ . Este valor alto indica uma grande separação entre as distribuições de probabilidade, sugerindo que o modelo possui um bom poder de discriminação para a classe analisada.

#### 2.4.4 Medidas de Concordância

Além de avaliar separadamente a probabilidade atribuída a cada classe ou a contagem de acertos, é útil medir o grau de concordância entre as previsões do modelo e os resultados observados, levando em conta a concordância que ocorreria apenas ao acaso. Duas métricas clássicas para esse fim são o coeficiente  $k$  de Cohen (Cohen, 1960) e o Coeficiente de Correlação de Matthews (MCC) (Matthews, 1975).

**Coeficiente  $k$  de Cohen** Seja  $\mathbf{M} = (m_{ij})$  a matriz de confusão, em que  $m_{ij}$  indica o número de partidas em que a classe verdadeira é  $i$  e a previsão do modelo é  $j$ . Definem-se

$$p_o = \frac{\sum_i m_{ii}}{\sum_{i,j} m_{ij}}, \quad p_e = \frac{\sum_i (\sum_j m_{ij}) (\sum_j m_{ji})}{(\sum_{i,j} m_{ij})^2}, \quad (2.36)$$

em que  $p_o$  é a proporção de concordância observada e  $p_e$  a concordância esperada ao acaso.

O coeficiente é então

$$\kappa = \frac{p_o - p_e}{1 - p_e}. \quad (2.37)$$

Valores próximos de 1 indicam forte concordância além do acaso, enquanto valores em torno de 0 sugerem que o desempenho não difere do puro palpite.

**Coeficiente de Correlação de Matthews (MCC)** Para o caso multiclasse, seja  $\mathbf{M} = (m_{ij})_{3 \times 3}$  a matriz de confusão, em que  $m_{ij}$  conta as instâncias da classe verdadeira  $i$  preditas como  $j$ . Defina

$$c = m_{11} + m_{22} + m_{33}, \quad s = \sum_{i=1}^3 \sum_{j=1}^3 m_{ij}, \quad p_i = \sum_{j=1}^3 m_{ij}, \quad t_i = \sum_{j=1}^3 m_{ji}. \quad (2.38)$$

O Coeficiente de Correlação de Matthews para três classes é então

$$\text{MCC} = \frac{c s - \sum_{i=1}^3 p_i t_i}{\sqrt{(s^2 - \sum_{i=1}^3 p_i^2)(s^2 - \sum_{i=1}^3 t_i^2)}}. \quad (2.39)$$

Essa expressão é a especialização para  $K = 3$  da formulação geral proposta por (Gorodkin, 2004).

Essas duas métricas complementam as regras de escore ao quantificar o quanto o modelo acerta acima do acaso, oferecendo uma visão mais robusta sobre a confiabilidade das previsões categóricas.

Utilizando novamente a matriz de confusão hipotética da Tabela 2.2, podemos ilustrar o cálculo de ambas as medidas de concordância.

### Cálculo do Coeficiente $\kappa$ de Cohen

Primeiro, calculamos a proporção de concordância observada ( $p_o$ ), que é a acurácia global

$$p_o = \frac{35 + 5 + 12}{100} = 0,52. \quad (2.40)$$

Em seguida, a concordância esperada ao acaso ( $p_e$ ) é calculada usando os totais de cada linha (predições) e coluna (observações):

$$p_e = \frac{(60 \times 45) + (10 \times 25) + (30 \times 30)}{100^2} = \frac{2700 + 250 + 900}{10000} = \frac{3850}{10000} = 0,385. \quad (2.41)$$

Finalmente, o coeficiente Kappa é

$$\kappa = \frac{p_o - p_e}{1 - p_e} = \frac{0,52 - 0,385}{1 - 0,385} = \frac{0,135}{0,615} \approx 0,220. \quad (2.42)$$

### Cálculo do Coeficiente de Correlação de Matthews (MCC)

Para o MCC, calculamos os componentes a partir da matriz:

- Total de acertos:  $c = 35 + 5 + 12 = 52$ .
- Total de observações:  $s = 100$ .
- Soma dos produtos das predições e observações por classe:  $\sum p_i t_i = (60 \cdot 45) + (10 \cdot 25) + (30 \cdot 30) = 3850$ .
- Soma das predições ao quadrado:  $\sum p_i^2 = 60^2 + 10^2 + 30^2 = 4600$ .
- Soma das observações ao quadrado:  $\sum t_i^2 = 45^2 + 25^2 + 30^2 = 3550$ .

Substituindo na fórmula do MCC

$$\text{MCC} = \frac{52 \cdot 100 - 3850}{\sqrt{(100^2 - 4600)(100^2 - 3550)}} = \frac{1350}{\sqrt{5400 \cdot 6450}} \approx \frac{1350}{5901,69} \approx 0,229.$$

Ambos os valores,  $\kappa \approx 0,220$  e  $\text{MCC} \approx 0,229$ , indicam uma concordância fraca, porém positiva, entre as previsões do modelo e os resultados observados, após a correção para o acaso.

### 2.4.5 Diagrama de Calibração

Além das métricas já citadas, é fundamental verificar se as probabilidades do modelo são literalmente interpretáveis. Ou seja, se uma previsão de “20% de chance” corresponde de fato a uma frequência de 20% de ocorrências. No contexto do Brasileirão, isso significa que, ao prever 20% de chance de vitória para o visitante em um conjunto de partidas, esperamos que esse time realmente vença perto de 20% desses confrontos.

O diagrama de calibração (*reliability diagram*) expõe visualmente essa relação. A calibração perfeita ocorre sobre a reta de identidade ( $y = x$ ), e desvios sistemáticos da curva do modelo em relação a essa reta indicam vieses de superestimação (curva abaixo) ou subestimação (curva acima). Uma má calibração tem implicações práticas, podendo levar a decisões equivocadas em apostas ou na análise de risco de um clube, que envolve desde o planejamento financeiro e contratações até a gestão do risco de rebaixamento.

Para estimar a calibração, reunimos todos os pares  $(x_i, y_i)$ , com  $x_i$  sendo a probabilidade prevista do evento de interesse e

$$y_i = \begin{cases} 1, & \text{se o evento ocorreu,} \\ 0, & \text{caso contrário.} \end{cases} \quad (2.43)$$

Em seguida, aplicamos uma regressão local, especificamente o método LOESS (Cleveland e Devlin, 1988), que consiste em um procedimento de suavização de dados. Em cada ponto  $x_0$  da grade, seleciona uma vizinhança contendo uma fração fixa (*span*%) dos dados mais próximos de  $x_0$  e atribui a cada vizinho  $x_i$  um peso

$$w_i = \left(1 - (|x_i - x_0|/d)^3\right)^3, \quad (2.44)$$

onde  $d$  é a maior distância entre  $x_0$  e esses vizinhos. Sobre esses pontos ajusta-se um polinômio de baixo grau ponderado pelos  $w_i$

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg \min_{\beta_0, \beta_1} \sum_i w_i (y_i - (\beta_0 + \beta_1(x_i - x_0)))^2, \quad (2.45)$$

e a predição local em  $x_0$  é

$$\hat{f}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0. \quad (2.46)$$

Esse processo, ao ser repetido para cada ponto da grade, resulta na curva suave  $\hat{f}(x)$ . No diagrama, essa curva (em azul) é plotada junto à reta de identidade ( $y = x$ ), que representa o cenário de calibração ideal. Um modelo bem calibrado terá sua curva muito próxima a essa reta, indicando que uma probabilidade prevista de  $x\%$  corresponde, na prática, a uma frequência de ocorrência de aproximadamente  $x\%$ .

Esse diagrama permite avaliar não apenas se o modelo classifica corretamente, mas também se podemos confiar no valor absoluto das probabilidades que ele fornece, funda-

mental para decisões baseadas em risco.

### 2.4.6 Análise de Cenários para Desfechos Finais

Além da avaliação da qualidade preditiva jogo a jogo, este trabalho também avalia a capacidade dos modelos em projetar os desfechos finais do campeonato em diferentes estágios da competição. Para isso, foi desenvolvida uma análise de cenários na qual os modelos são ajustados em três momentos da reta final do torneio: faltando 15, 10 e 5 rodadas para o término.

A partir de cada um desses pontos, são simuladas finalizações completas do campeonato. Em cada simulação, os resultados das partidas restantes são previstos, os pontos (3 para vitória, 1 para empate, 0 para derrota) são computados e a tabela de classificação final é gerada. Esse processo permite estimar as probabilidades de cada equipe atingir posições-chave, como ser campeã, terminar entre os quatro primeiros (G4) ou na zona de rebaixamento (Z4). Os resultados detalhados desta análise são apresentados e discutidos no Capítulo 4.



## Capítulo 3

# Análise de qualidade de ajuste dos modelos

Neste capítulo, apresentamos a análise prática dos dados do Campeonato Brasileiro, com o objetivo de ajustar os modelos propostos e avaliar sua aderência à realidade da competição. Iniciamos a análise com uma exploração das características dos dados (Seção 3.1), estabelecendo o contexto para a estimação dos parâmetros de cada modelo. Em seguida, com as forças de ataque, defesa e o efeito de mando de campo quantificados, conduzimos simulações completas do campeonato para os modelos Hierárquico e Poisson Independente (Seções 3.3 e 3.4). A análise dos resultados simulados, comparando gols e pontos previstos com os valores reais, conclui com um confronto direto entre as duas abordagens (Seção 3.5), permitindo identificar as forças e limitações de cada uma na tarefa de modelar a dinâmica do futebol.

### 3.1 Análise Descritiva

A Figura 3.1 apresenta a distribuição de gols marcados pelos times mandantes, enquanto a Figura 3.2 mostra a distribuição de gols marcados pelos times visitantes.

Os gols dos mandantes apresentam uma distribuição assimétrica positiva, com maior concentração em valores 0, 1 e 2 gols, sendo bem ajustados por uma distribuição de Poisson com parâmetro  $\lambda = 1,42$ . Já os gols dos visitantes exibem assimetria positiva, com predominância de 0 e 1 gols, e ajuste adequado por uma distribuição de Poisson com  $\lambda = 1,07$ . Em ambos os casos, observa-se que os mandantes tendem a marcar mais gols, em média, do que os visitantes, refletindo o impacto do mando de campo.

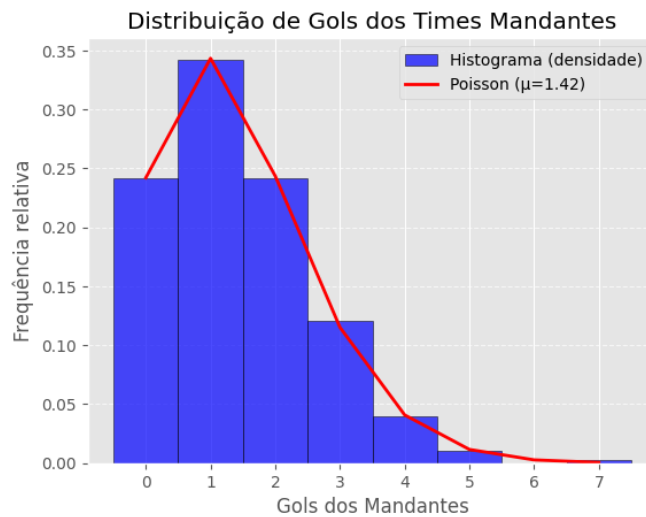


Figura 3.1: Distribuição dos gols marcados e ajuste por uma distribuição de Poisson

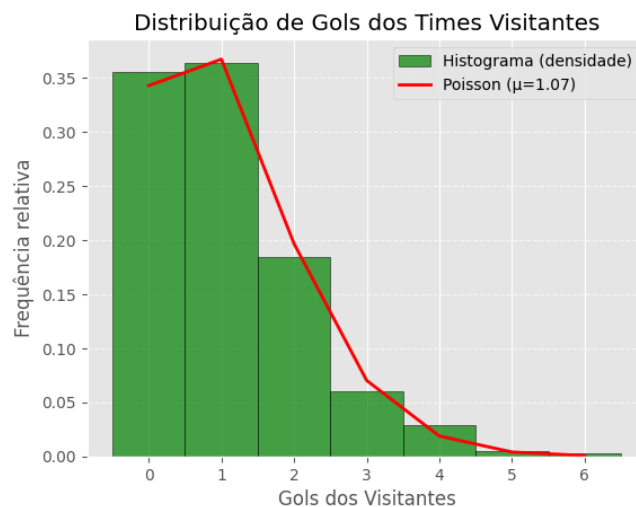


Figura 3.2: Distribuição dos gols marcados e ajuste por uma distribuição de Poisson

Além da distribuição geral, a Tabela 3.1 apresenta as médias de gols por time, se-

paradas entre mandantes e visitantes, permitindo uma visão detalhada do desempenho ofensivo das equipes.

Tabela 3.1: Médias de gols por time, ordenadas pela classificação do campeonato.

<b>Times</b>	<b>Média de Gols Mandante</b>	<b>Média de Gols Visitante</b>
Palmeiras	1,84	1,53
Grêmio	1,79	1,53
Atlético-MG	1,47	1,26
Flamengo	1,37	1,58
Botafogo-RJ	1,89	1,16
Bragantino	1,47	1,11
Fluminense	1,74	0,95
Athletico-PR	1,74	0,95
Internacional	1,58	0,84
Fortaleza	1,47	0,89
São Paulo	1,53	0,58
Cuiabá	1,21	0,89
Corinthians	1,37	1,11
Cruzeiro	0,74	1,11
Vasco da Gama	1,26	0,89
Bahia	1,53	1,11
Santos	1,37	0,68
Goiás	1,00	0,89
Coritiba	0,89	1,26
América-MG	1,11	1,11

Observamos que os times com melhor desempenho no campeonato, como Palmeiras e Grêmio, possuem médias de gols como mandantes e visitantes acima de 1,5, o que indica uma consistência ofensiva tanto em casa quanto fora. O Botafogo-RJ apresenta a maior média de gols como mandante (1,89), enquanto o Flamengo possui a maior média como visitante (1,58).

Por outro lado, os times com menor desempenho, como Coritiba e Goiás, apresentam médias inferiores a 1 gol como mandantes. Observa-se também que o mando de campo tem influência significativa, com a maioria dos times marcando mais gols como mandantes do que como visitantes. Isso está alinhado com a expectativa de vantagem para o time da casa.

Algumas exceções, no entanto, são notáveis. O Flamengo, por exemplo, possui uma média de gols ligeiramente maior como visitante (1,58) do que como mandante (1,37), indicando um bom desempenho fora de casa.

Essa análise reforça a relevância do mando de campo e das forças ofensivas e defensivas dos times, aspectos que detalhamos na modelagem estatística subsequente.

A Tabela 3.2 apresenta a classificação final do campeonato. As colunas detalham, para cada clube, a Posição final (Pos), o número de Vitórias (V), Empates (E), Derrotas (D), Gols Feitos (GF), Gols Contra (GC), o Saldo de Gols (SG) e o total de Pontos.

Tabela 3.2: Classificação final do campeonato.

Pos	Clube	V	E	D	GF	GC	SG	Pontos
1	Palmeiras	20	10	8	64	33	31	70
2	Grêmio	21	5	12	63	56	7	68
3	Atlético Mineiro	19	9	10	52	32	20	66
4	Flamengo	19	9	10	56	42	14	66
5	Botafogo	18	10	10	58	37	21	64
6	Bragantino	17	11	10	49	35	14	62
7	Fluminense	16	8	14	51	47	4	56
8	Athletico-PR	14	14	10	51	43	8	56
9	Internacional	15	10	13	46	45	1	55
10	Fortaleza	15	9	14	45	44	1	54
11	São Paulo	14	11	13	43	40	3	53
12	Cuiabá	14	9	15	40	39	1	51
13	Corinthians	12	14	12	47	48	-1	50
14	Cruzeiro	11	14	13	35	36	-1	47
15	Vasco da Gama	12	9	17	41	51	-10	45
16	Bahia	12	8	18	50	53	-3	44
17	Santos	11	10	17	39	64	-25	43
18	Goiás	9	11	18	36	53	-17	38
19	Coritiba	8	6	24	41	83	-42	30
20	América-MG	5	9	24	42	81	-39	24

O Palmeiras terminou como campeão, somando 70 pontos, com 20 vitórias, um saldo de gols de 31 e um ataque eficiente que marcou 64 gols. O Grêmio, vice-campeão, alcançou 68 pontos com 21 vitórias, apesar de ter um saldo de gols menor (7), evidenciando fragilidades defensivas, dado o elevado número de gols sofridos (56).

Atlético Mineiro e Flamengo finalizaram empatados com 66 pontos, mas o saldo de gols favoreceu o Atlético Mineiro (20 contra 14). O Botafogo, que liderou grande parte da competição, encerrou na quinta posição com 64 pontos, apresentando um saldo positivo de 21 gols.

Na parte inferior da tabela, América-MG e Coritiba ocuparam as últimas posições, com 24 e 30 pontos, respectivamente, sendo prejudicados por defesas frágeis que sofreram, em média, mais de 2 gols por jogo. O Goiás também não conseguiu escapar do rebaixamento, acumulando apenas 38 pontos.

De forma geral, a tabela reflete a importância de um equilíbrio entre ataque e defesa. Equipes como o Palmeiras e Atlético Mineiro demonstraram consistência em ambas as

frentes, enquanto equipes como o Fluminense apresentaram maior vulnerabilidade defensiva, o que afetou sua competitividade. O impacto do mando de campo e das forças ofensivas e defensivas será mais explorado na modelagem estatística subsequente.

## 3.2 Análise da Cadeia

Para ambos os modelos, geramos 3 cadeias com 3000 iterações e descartamos as primeiras 1000, de modo a eliminar a influência dos valores iniciais. As estimativas finais, incluindo as médias da distribuição posterior e os intervalos de credibilidade, foram derivadas das amostras obtidas após o *burn-in*.

Observamos *traceplots* das cadeias (ver Apêndice A) e que a estatística de Gelman-Rubin ( $\hat{R}$ ), indicaram que todos os parâmetros tiveram  $\hat{R} \leq 1.01$ , sugerindo que as cadeias alcançaram convergência adequada.

## 3.3 Simulações Modelo Hierárquico

### 3.3.1 Estimação dos Parâmetros

Os parâmetros estimados para os efeitos de mando de campo, ataque e defesa são apresentados nas Tabelas 3.3, 3.4 e 3.5, respectivamente.

Tabela 3.3: Efeito de Mando de Campo (*home*).

Parâmetro	Média	DP	IC 95%
Efeito Casa	0,316	0,044	[0,229; 0,401]

Legenda: DP = *Desvio Padrão*; IC 95% = *Intervalo de Credibilidade de 95%*.

A Tabela 3.3 mostra que o efeito de mando de campo (*home*) tem uma média estimada de 0,316, com um intervalo de credibilidade de 95% [0,229; 0,401]. Esse valor positivo reforça a hipótese de que jogar em casa confere uma vantagem significativa para os times, o que é consistente com o maior número médio de gols marcados pelos mandantes, observado na análise descritiva.

Na Tabela 3.4, são apresentados os efeitos de ataque (*att*) para cada time. Times como Palmeiras, Grêmio e Botafogo-RJ apresentam valores médios positivos, indicando uma capacidade ofensiva acima da média. Por outro lado, times como Cruzeiro, Santos e

Tabela 3.4: Efeitos de Ataque (*att*).

Time	Média	DP	IC 95%
Palmeiras	0,14	0,10	[-0,04; 0,34]
Grêmio	0,14	0,10	[-0,04; 0,35]
Atlético-MG	0,04	0,09	[-0,14; 0,21]
Flamengo	0,07	0,09	[-0,10; 0,26]
Botafogo-RJ	0,09	0,09	[-0,08; 0,29]
Bragantino	0,01	0,09	[-0,18; 0,19]
Fluminense	0,03	0,09	[-0,14; 0,22]
Athletico-PR	0,03	0,09	[-0,15; 0,21]
Internacional	-0,01	0,09	[-0,19; 0,17]
Fortaleza	-0,02	0,09	[-0,21; 0,16]
São Paulo	-0,06	0,09	[-0,26; 0,11]
Cuiabá	-0,07	0,09	[-0,27; 0,11]
Corinthians	-0,00	0,09	[-0,19; 0,18]
Cruzeiro	-0,11	0,10	[-0,32; 0,07]
Vasco	-0,05	0,10	[-0,26; 0,12]
Bahia	0,03	0,09	[-0,15; 0,21]
Santos	-0,07	0,10	[-0,28; 0,11]
Goiás	-0,10	0,10	[-0,32; 0,08]
Coritiba	-0,05	0,09	[-0,25; 0,13]
América-MG	-0,04	0,09	[-0,23; 0,14]

Legenda:  $DP = \text{Desvio Padrão}$ ;  $IC 95\% = \text{Intervalo de Credibilidade de 95\%}$ .

Goiás têm efeitos de ataque negativos ou próximos de zero, evidenciando dificuldades no setor ofensivo.

A Tabela 3.5 apresenta os efeitos de defesa (*def*). Valores negativos indicam uma boa capacidade defensiva, enquanto valores positivos apontam maior vulnerabilidade. Times como Palmeiras e Atlético-MG possuem efeitos defensivos negativos, indicando solidez defensiva. Em contrapartida, Coritiba e América-MG têm os piores efeitos defensivos, refletindo as dificuldades enfrentadas por essas equipes na temporada, que resultaram em altos números de gols sofridos.

A relação entre os efeitos de ataque e defesa evidencia o equilíbrio necessário para um bom desempenho no campeonato. Times como Palmeiras e Botafogo, que combinam bons efeitos ofensivos e defensivos, conseguiram alcançar as primeiras posições na tabela final. Já equipes com defesas vulneráveis, como Coritiba e América-MG, terminaram nas últimas colocações, mesmo que algumas apresentassem ataques razoáveis.

Esses resultados reforçam a importância de considerar tanto os aspectos ofensivos quanto defensivos no modelo estatístico, além do impacto do mando de campo, que contribui para um melhor entendimento dos fatores que influenciam o desempenho das equipes.

Tabela 3.5: Efeitos de Defesa (*def*).

Time	Média	DP	IC 95%
Palmeiras	-0,22	0,13	[-0,49; 0,03]
Grêmio	0,15	0,12	[-0,08; 0,38]
Atlético-MG	-0,24	0,13	[-0,51; 0,01]
Flamengo	-0,07	0,12	[-0,32; 0,17]
Botafogo-RJ	-0,15	0,13	[-0,41; 0,09]
Bragantino	-0,19	0,13	[-0,45; 0,06]
Fluminense	0,01	0,12	[-0,23; 0,25]
Athletico-PR	-0,05	0,12	[-0,31; 0,18]
Internacional	-0,02	0,12	[-0,27; 0,21]
Fortaleza	-0,04	0,12	[-0,28; 0,19]
São Paulo	-0,14	0,13	[-0,40; 0,11]
Cuiabá	-0,12	0,13	[-0,38; 0,11]
Corinthians	0,02	0,12	[-0,22; 0,26]
Cruzeiro	-0,25	0,14	[-0,52; 0,01]
Vasco	0,07	0,12	[-0,16; 0,29]
Bahia	0,10	0,12	[-0,13; 0,33]
Santos	0,25	0,11	[0,02; 0,47]
Goiás	0,10	0,12	[-0,14; 0,33]
Coritiba	0,36	0,11	[0,13; 0,58]
América-MG	0,45	0,11	[0,22; 0,67]

Legenda: DP = *Desvio Padrão*; IC 95% = *Intervalo de Credibilidade de 95%*.

As Figuras 3.3 e 3.4 apresentam os intervalos de credibilidade de 95% e as medianas estimadas para os parâmetros de ataque e defesa, respectivamente, de cada time.

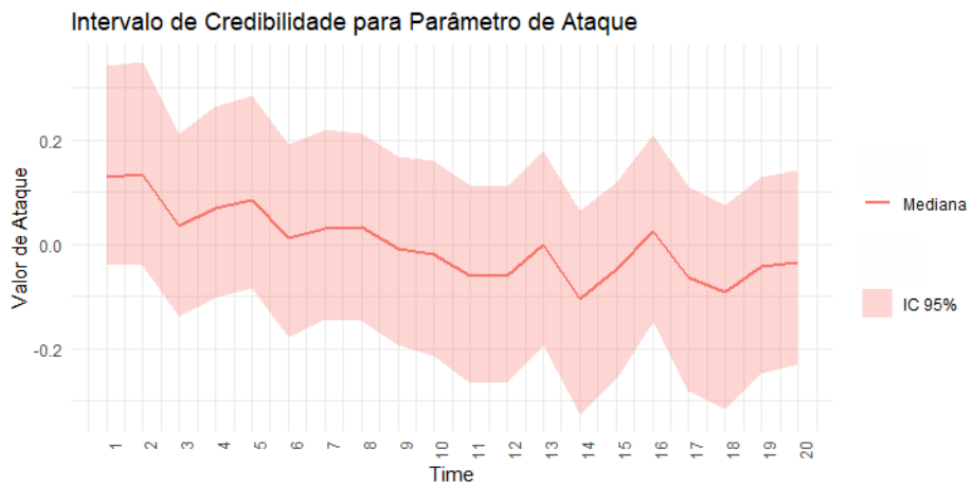


Figura 3.3: Intervalo de Credibilidade do (*att*) para cada time

Na Figura 3.3, observa-se que, em geral, os valores medianos de ataque (*att*) tendem a diminuir conforme a colocação dos times na tabela, indicando menor eficiência ofensiva para as equipes de posições inferiores.

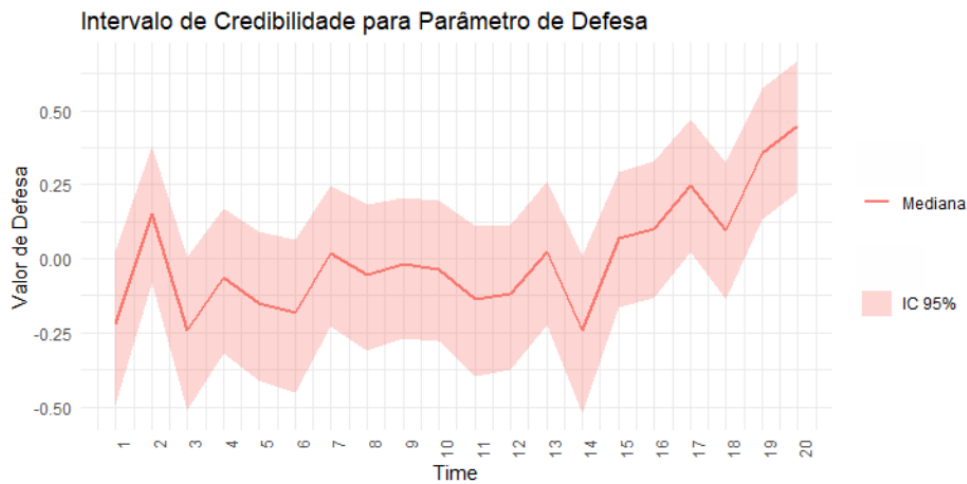


Figura 3.4: Intervalo de Credibilidade da ( $def$ ) para cada time

Por outro lado, a Figura 3.4 mostra que os valores medianos de defesa ( $def$ ) tendem a aumentar com a piora na colocação, refletindo maior vulnerabilidade defensiva das equipes nas últimas posições do campeonato.

A Figura 3.5 apresenta a relação entre os efeitos médios de ataque e defesa das equipes do Campeonato Brasileiro. No eixo  $x$ , valores positivos indicam maior força ofensiva, enquanto valores negativos representam ataques menos efetivos. Já no eixo  $y$ , valores positivos indicam defesas mais vulneráveis, enquanto valores negativos representam defesas mais sólidas.

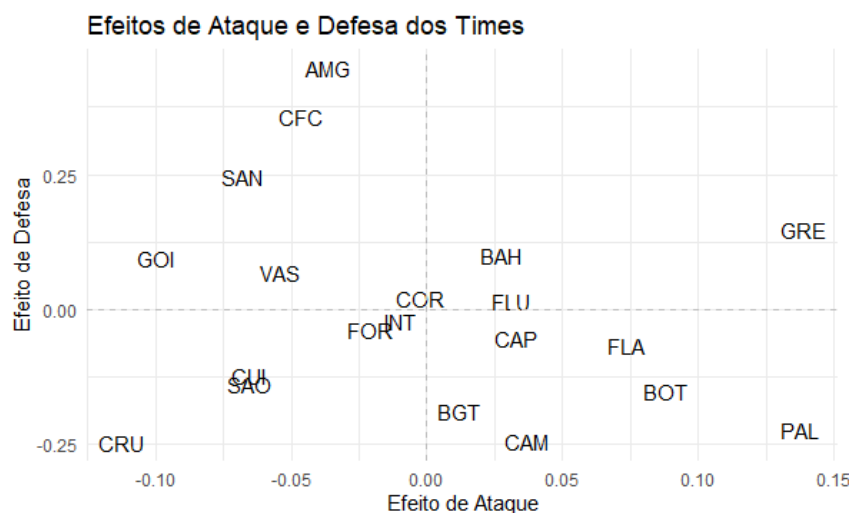


Figura 3.5: Efeitos médios de ataque e defesa

Ainda na Figura 3.5, observa-se que times como Palmeiras, Botafogo, Flamengo combinam bons efeitos ofensivos (positivos no eixo  $x$ ) e bons efeitos defensivos (valores negativos no eixo  $y$ ), o que explicam seus desempenhos superiores no campeonato. Por outro lado,

equipes como América-MG, Coritiba, Santos, Goiás apresentam altos valores de efeito defensivo (positivos no eixo  $y$ ), e indicando vulnerabilidade e valores negativos em efeitos ofensivos, enquanto equipes como Cruzeiro e São Paulo apresentam baixos valores de efeito ofensivo, refletindo uma menor capacidade de marcar gols.

Vale destacar o caso do Grêmio, que mesmo possuindo um efeito defensivo semelhante ao do Goiás, que terminou na 18<sup>a</sup> colocação, conseguiu ficar em segundo lugar no campeonato devido ao seu alto efeito ofensivo. Esse resultado indica que, em competições como o Campeonato Brasileiro, a capacidade de marcar gols pode ter um impacto maior na classificação final do que a solidez defensiva.

### 3.3.2 Predição de Jogos e Probabilidades de Vitória

A Tabela 4.10 apresenta as probabilidades estimadas de vitória do time A ( $PV_A$ ), empate ( $PE$ ) e vitória do time B ( $PV_B$ ) para as primeiras 20 partidas do Campeonato Brasileiro.

Tabela 3.6: Probabilidades de Vitória do Time A ( $PV_A$ ), Empate ( $PE$ ) e Vitória do Time B ( $PV_B$ ) para cada jogo, com os resultados observados.

Jogo	Time A	Time B	PV A	PE	PV B	Resultado Obs
1	Palmeiras	Cuiaba	0,54	0,26	0,20	A
2	America-MG	Fluminense	0,33	0,23	0,44	B
3	Bragantino	Bahia	0,52	0,26	0,22	A
4	Athletico-PR	Goiás	0,54	0,25	0,21	A
5	Fortaleza	Internacional	0,45	0,29	0,26	E
6	Botafogo-RJ	Sao Paulo	0,48	0,28	0,23	A
7	Aletico-MG	Vasco	0,56	0,26	0,19	B
8	Flamengo	Coritiba	0,65	0,19	0,15	A
9	Corinthians	Cruzeiro	0,39	0,31	0,31	A
10	Gremio	Santos	0,59	0,21	0,20	A
11	Fluminense	Athletico-PR	0,44	0,27	0,29	A
12	Cuiaba	Bragantino	0,39	0,30	0,31	E
13	Sao Paulo	America-MG	0,64	0,20	0,16	A
14	Cruzeiro	Gremio	0,49	0,27	0,24	A
15	Internacional	Flamengo	0,42	0,28	0,31	A
16	Santos	Athletico-MG	0,29	0,27	0,44	E
17	Vasco	Palmeiras	0,32	0,28	0,40	E
18	Coritiba	Fortaleza	0,34	0,26	0,40	B
19	Goiás	Corinthians	0,40	0,27	0,33	A
20	Bahia	Botafogo-RJ	0,36	0,28	0,36	B

Essas probabilidades foram geradas a partir do modelo hierárquico bayesiano, levando

em consideração os efeitos ofensivos, defensivos e o mando de campo de cada equipe. Além disso, a tabela exibe os resultados observados (*Resultado Obs*), onde “A” indica a vitória do time A, “B” a vitória do time B e “E” um empate.

Observa-se que, na maioria dos jogos, o time A (mandante) apresenta maior probabilidade de vitória ( $PV_A$ ) em comparação com o time B ( $PV_B$ ), refletindo a vantagem de jogar em casa. Um exemplo claro é o jogo 8, entre Flamengo e Coritiba, onde a probabilidade de vitória do Flamengo ( $PV_A = 0,65$ ) é consideravelmente superior à de empate ( $PE = 0,19$ ) e vitória do Coritiba ( $PV_B = 0,15$ ). Esse comportamento também é observado em outras partidas como a 1 ou 13.

No entanto, há casos onde o time visitante tem maior probabilidade de vencer. Isso ocorre, por exemplo, no confronto 2, entre América-MG e Fluminense, onde  $PV_B = 0,44$  é superior a  $PV_A = 0,33$ , indicando que o modelo considerou o Fluminense favorito mesmo jogando fora de casa. Da mesma forma, em 17, Vasco contra Palmeiras também apresenta maior probabilidade de vitória para o visitante ( $PV_B = 0,40$ ).

A análise dos resultados observados permite verificar a precisão das probabilidades estimadas. Por exemplo, no jogo Vasco x Palmeiras, o modelo indicou maior probabilidade de vitória do Palmeiras ( $PV_B = 0,40$ ) em comparação ao Vasco ( $PV_A = 0,32$ ), mas o jogo terminou empatado ( $PE = 0,28$ ), mostrando que as estimativas do modelo reconheciam um certo equilíbrio na partida.

### 3.3.3 Comparação entre Pontos Preditos e Observados

A Tabela 3.7 compara a pontuação esperada, obtida a partir do modelo estatístico, com a pontuação observada no campeonato.

De maneira geral, o modelo conseguiu capturar bem a tendência geral da competição, com a maioria das equipes apresentando valores esperados próximos aos observados. Contudo, identificamos algumas discrepâncias notáveis.

Nos extremos, o Grêmio foi claramente subestimado pelo modelo, com uma pontuação esperada de 52,23 pontos, enquanto na realidade alcançou 66 pontos. Esse erro pode ser atribuído à vulnerabilidade defensiva da equipe, que sofreu um número elevado de gols, impactando negativamente as projeções do modelo. Por outro lado, o América-MG foi superestimado, com o modelo atribuindo 43,58 pontos esperados, mas a equipe terminou o campeonato com apenas 24 pontos, indicando que o modelo não capturou adequadamente seu desempenho abaixo da média.

Tabela 3.7: Comparação entre Pontuação Esperada e Observada por Equipe.

Times	Pontos esperados	Pontos observados
Palmeiras	63,67	70
Grêmio	52,23	66
Atlético-MG	60,64	65
Botafogo-RJ	56,25	63
Bragantino	55,27	63
Flamengo	54,59	63
Athletico-PR	53,49	56
Fluminense	52,77	55
São Paulo	53,07	53
Cuiabá	52,82	53
Fortaleza	52,28	50
Internacional	52,51	50
Corinthians	51,64	50
Cruzeiro	53,92	49
Vasco	49,80	44
Santos	46,48	44
Bahia	49,97	41
Goiás	48,55	41
Coritiba	44,92	37
América-MG	43,58	24

A maior precisão das previsões ocorreu na parte intermediária da tabela, em que equipes como Athletico-PR, Fluminense, São Paulo e Cuiabá apresentaram diferenças pequenas entre pontos esperados e observados. Isso sugere que o modelo conseguiu capturar com maior assertividade o desempenho de equipes de rendimento estável ao longo da competição.

Esses resultados indicam que, apesar da boa aderência geral do modelo, algumas equipes foram afetadas por fatores não capturados nas estimativas, como variações táticas, lesões, momentos de crise ou fases positivas inesperadas. Melhorias futuras podem incluir ajustes para capturar melhor desempenhos extremos e considerar fatores adicionais que possam influenciar a pontuação final.

### 3.3.4 Avaliação do Modelo Hierárquico

Como forma de validar a coerência dos parâmetros estimados pelo modelo, realizamos simulações do campeonato utilizando os valores estimados. A Figura 3.6 apresenta o histograma do número total de gols simulados pelo modelo, enquanto a Figura 3.7 mostra a separação dos gols entre mandantes e visitantes.

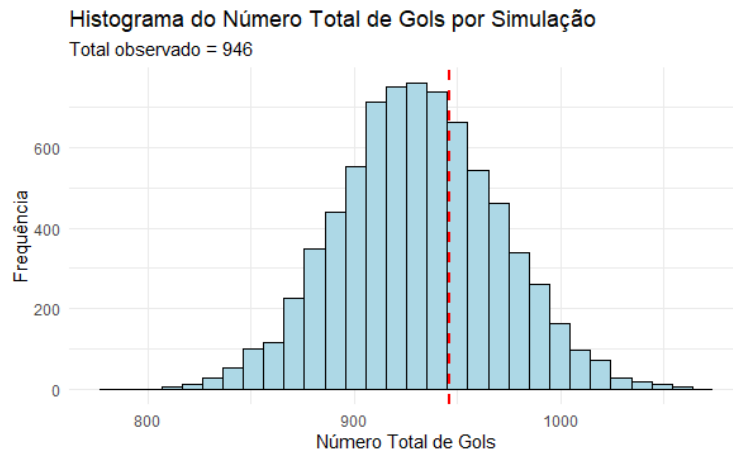


Figura 3.6: Histograma do número total de gols por simulação.

Com o histograma da Figura 3.6, nota-se que a distribuição simulada é aproximadamente normal, com a maior parte das simulações concentrada em torno do valor observado, indicando que o modelo conseguiu capturar bem a média do número total de gols.

Além disso, o fato de o valor observado estar próximo do centro da distribuição indica que o modelo está bem calibrado e não apresenta viés significativo. Caso o número real de gols estivesse nas extremidades da distribuição, isso poderia indicar um ajuste inadequado do modelo. Assim, os resultados demonstram que o modelo fornece uma boa estimativa para o total de gols no campeonato.

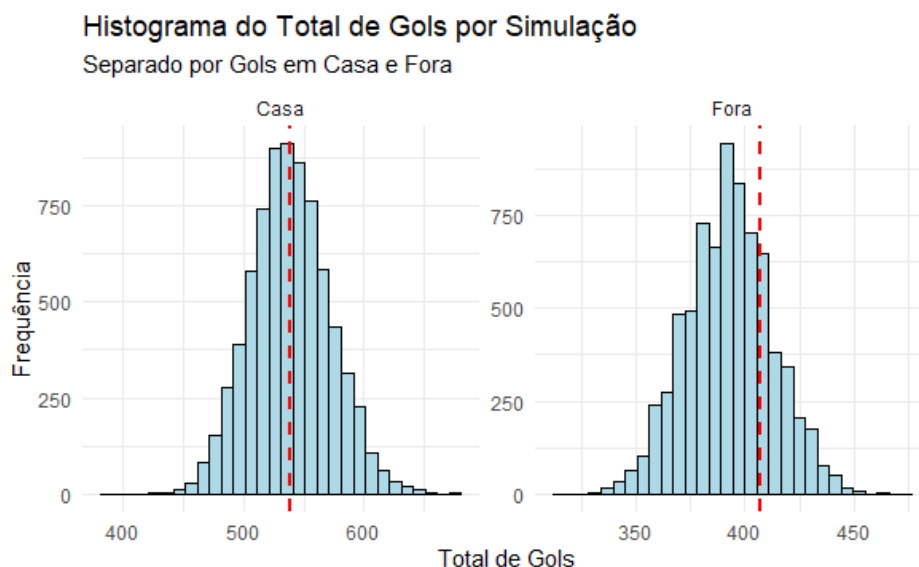


Figura 3.7: Histograma do total de gols por simulação, separado por gols marcados em casa e fora.

A Figura 3.7 apresenta a distribuição do número total de gols simulados, separados entre gols marcados pelos times mandantes (à esquerda) e visitantes (à direita). Observa-

se que ambas as distribuições seguem um formato aproximadamente normal, com a maior parte das simulações concentrada em torno dos valores observados.

Nota-se que os times mandantes tendem a marcar mais gols em comparação aos visitantes. Além disso, a variabilidade das distribuições sugere que, embora a média dos gols marcados em casa seja maior, há sobreposição nos intervalos de possíveis valores entre mandantes e visitantes. Isso indica que, apesar da vantagem média dos times da casa, há casos em que os visitantes conseguem marcar uma quantidade similar de gols.

O fato de os valores observados estarem bem posicionados dentro das distribuições simuladas indica que o modelo conseguiu representar adequadamente a dinâmica dos gols no campeonato, refletindo a diferença esperada entre desempenho em casa e fora.

Além da análise de histogramas totais de gols, avaliamos a precisão das previsões do modelo em relação ao número de gols marcados em cada jogo. A Figura 3.8 apresenta a diferença entre os gols previstos e os observados ao longo das partidas, enquanto a Figura 3.9 exibe essa diferença ao longo das rodadas, separada por equipe.

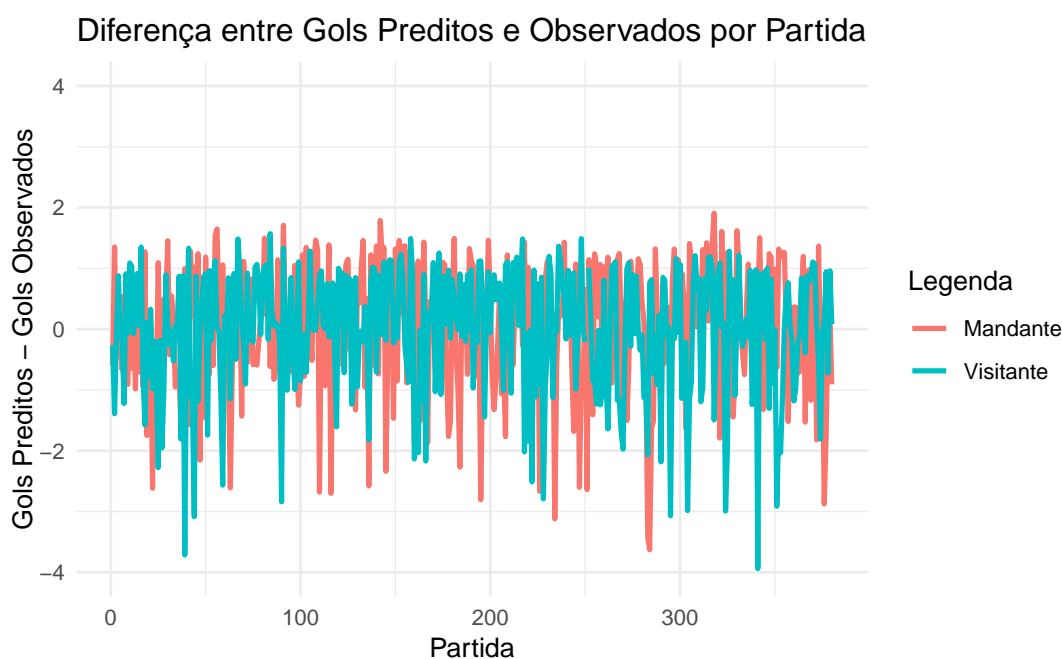


Figura 3.8: Diferença entre gols preditos e observados por jogo.

Observa-se que as diferenças oscilam em torno de zero, indicando que o modelo não apresenta um viés sistemático significativo, mas ainda possui variações individuais em cada jogo. Há momentos em que o modelo superestima (valores positivos) ou subestima (valores negativos) o número de gols marcados. Essas variações podem ser atribuídas à imprevisibilidade inerente ao futebol, como variações de desempenho das equipes, decisões

táticas e eventos inesperados, como expulsões e pênaltis.

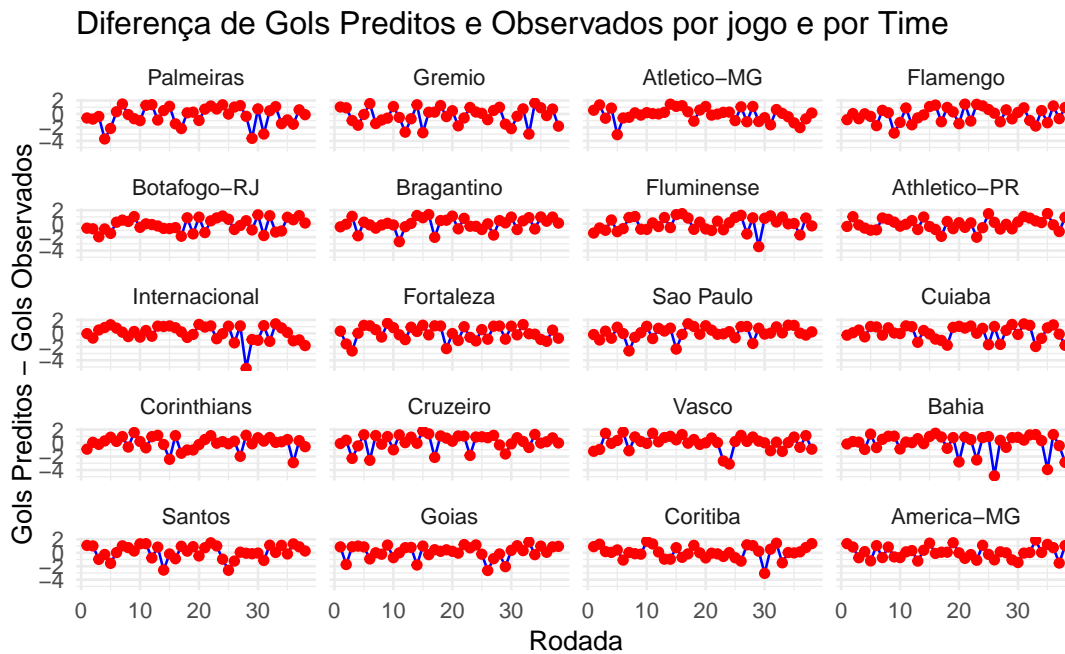


Figura 3.9: Diferença entre gols preditos e observados por equipe.

A Figura 3.9 apresenta a diferença entre os gols preditos pelo modelo e os gols observados ao longo das rodadas para cada equipe do campeonato. O eixo  $x$  representa as rodadas disputadas, enquanto o eixo  $y$  indica a diferença entre o número de gols previstos e o número real de gols marcados.

Observa-se que, para a maioria das equipes, as diferenças oscilam ao longo da competição, com momentos de superestimação e subestimação. Algumas equipes apresentam maior variabilidade nas previsões, indicando que o modelo pode ter tido dificuldades em capturar com precisão o desempenho ofensivo dessas equipes. Times como América-MG, Coritiba, Goiás e Santos parecem apresentar diferenças mais acentuadas em relação às previsões, sugerindo que o modelo pode ter subestimado ou superestimado significativamente seus desempenhos ofensivos. Já equipes como Cuiaba, São Paulo e Flamengo mostram menor dispersão, indicando que suas previsões foram mais alinhadas com os resultados reais.

A Figura 3.10 apresenta a diferença entre os pontos preditos pelo modelo e os pontos observados para cada equipe. Valores negativos indicam que o modelo subestimou o desempenho da equipe, enquanto valores positivos sugerem uma superestimação.

Observa-se que o modelo tende a puxar os times do topo da tabela para baixo e os times da parte inferior para cima, enquanto as equipes de meio de tabela apresentam uma

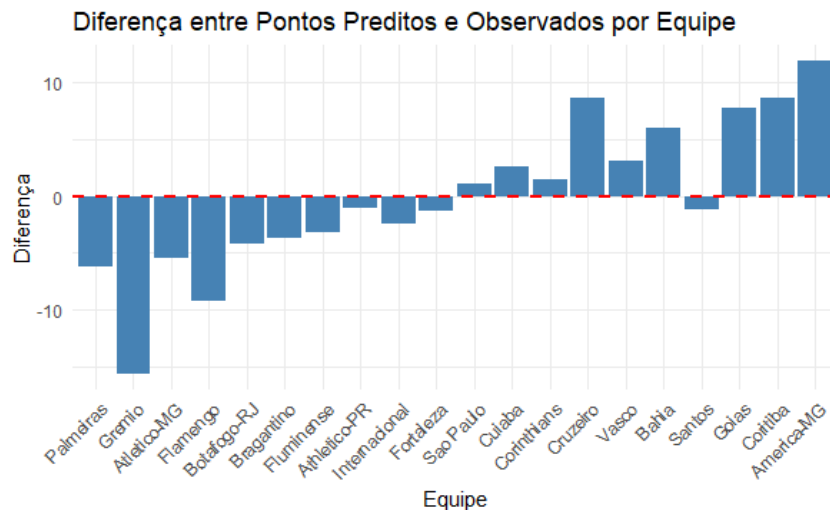


Figura 3.10: Diferença entre pontos preditos e observados por equipe.

diferença mais estável, próxima de zero. Isso sugere que o modelo pode estar suavizando excessivamente as previsões, reduzindo a discrepância entre os extremos da classificação. O Grêmio, por exemplo, foi superestimado pelo modelo, enquanto América-MG e Coritiba foram subestimados, indicando que o modelo não capturou completamente suas performances reais.

Essa tendência pode estar relacionada à estrutura hierárquica do modelo, que, ao estimar forças ofensivas e defensivas, pode apresentar um efeito de regressão à média, necessitando ajustes para capturar melhor desempenhos extremos.

A Figura 3.11 apresenta a evolução das pontuações acumuladas observadas e previstas pelo modelo ao longo das rodadas para cada equipe do campeonato. A linha sólida representa os pontos observados, enquanto a linha tracejada corresponde às expectativas do modelo.

Observa-se que, para a maioria das equipes, as trajetórias previstas acompanham razoavelmente bem as pontuações reais, demonstrando que o modelo conseguiu capturar a tendência geral do desempenho das equipes ao longo da temporada. No entanto, algumas discrepâncias são visíveis. Equipes como Botafogo-RJ e Grêmio apresentaram um desempenho inferior ao esperado pelo modelo nas rodadas finais, enquanto América-MG e Coritiba superaram as previsões.

Além disso, verifica-se que as equipes do meio da tabela tendem a apresentar menor discrepância entre a pontuação esperada e a real, indicando uma maior estabilidade na previsão para esses times. Já as equipes que terminaram em posições mais extremas (topo ou fundo da tabela) demonstram maiores desvios, o que pode estar relacionado à

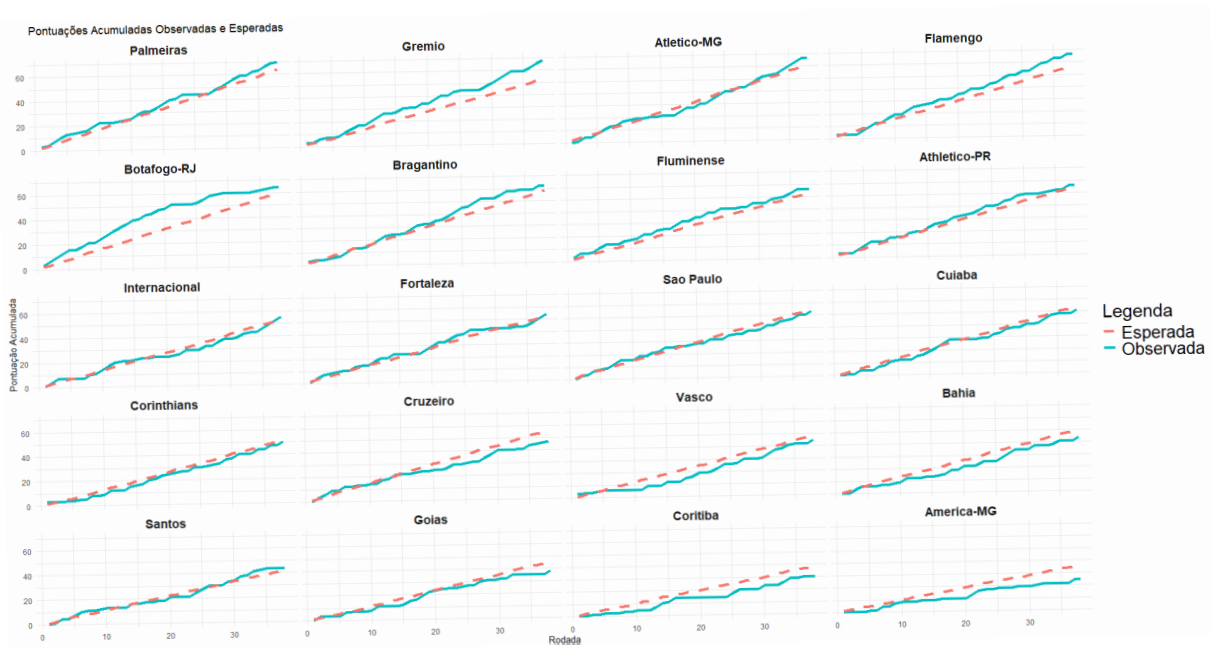


Figura 3.11: Pontuações acumuladas observadas e esperadas ao longo das rodadas para cada equipe.

dificuldade do modelo em capturar desempenhos extremos ao longo do campeonato.

A Figura 3.12 apresenta um mapa de calor com as probabilidades de cada equipe terminar em uma determinada posição na classificação final do campeonato. O eixo vertical representa os times, enquanto o eixo horizontal indica as posições na tabela. A intensidade da cor azul reflete a probabilidade associada a cada combinação de equipe e posição, conforme indicado na escala à direita.

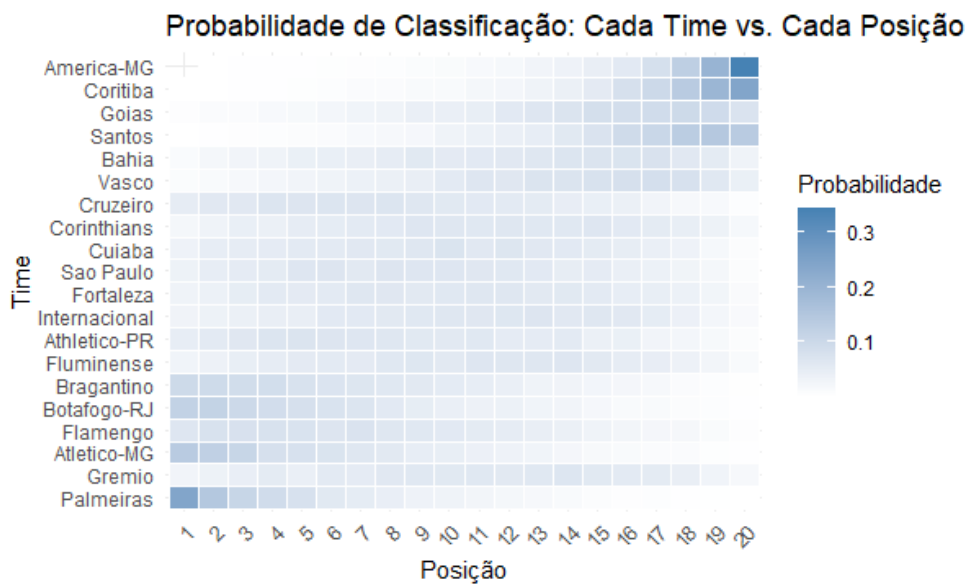


Figura 3.12: Mapa de calor das probabilidades de classificação final de cada equipe.

Observa-se que equipes nas extremidades da tabela, como Palmeiras e América-MG, apresentam uma maior concentração de probabilidades nas primeiras e últimas posições, respectivamente, refletindo as expectativas do modelo sobre seus desempenhos. Enquanto o Palmeiras era projetado para figurar entre os primeiros colocados, o América-MG tinha alta probabilidade de terminar nas últimas posições. Por outro lado, o Grêmio, apesar de sua boa campanha, foi subestimado pelo modelo, que previa uma posição inferior à que realmente alcançou, possivelmente devido ao seu baixo saldo de gols.

Além disso, percebe-se que para equipes do meio da tabela, como Fortaleza, Internacional e São Paulo, as probabilidades estão mais distribuídas ao longo de diversas posições, sugerindo uma maior incerteza na previsão de seus desempenhos finais. Essa variabilidade pode indicar que essas equipes tiveram desempenhos irregulares ao longo da temporada, tornando mais difícil uma previsão precisa de sua classificação final.

## 3.4 Simulações Modelo Poisson Independente

As análises para o modelo Poisson independente seguem a mesma estrutura da seção anterior, de modo que as interpretações já discutidas não serão repetidas aqui. Em vez disso, focaremos nos principais aspectos observados nas simulações deste modelo.

### 3.4.1 Estimação dos Parâmetros

Os parâmetros estimados para o modelo Poisson Independente são apresentados nas Tabelas 3.8, 3.9 e 3.10.

Tabela 3.8: Efeito de Mando de Campo (*home*).

Parâmetro	Média	DP	IC 95%
Efeito Casa	0,28	0,04	[0,19; 0,36]

A Tabela 3.8 mostra que o efeito de mando de campo (*home*) continua positivo, com uma média de 0,28, indicando vantagem para os times mandantes, embora ligeiramente inferior ao observado no modelo hierárquico.

A Tabela 3.9 apresenta os efeitos de ataque (*att*) das equipes. Palmeiras e Grêmio mantêm os maiores valores médios, enquanto times como Cruzeiro e Goiás apresentam efeitos negativos.

Tabela 3.9: Efeitos de Ataque (*att*).

Time	Média	DP	IC 95%
Palmeiras	0,30	0,12	[0,05; 0,54]
Grêmio	0,31	0,12	[0,06; 0,55]
Atlético-MG	0,10	0,14	[-0,18; 0,35]
Flamengo	0,18	0,13	[-0,08; 0,43]
Botafogo-RJ	0,21	0,13	[-0,06; 0,46]
Bragantino	0,04	0,14	[-0,25; 0,31]
Fluminense	0,09	0,14	[-0,19; 0,35]
Athletico-PR	0,08	0,14	[-0,19; 0,34]
Internacional	-0,02	0,15	[-0,31; 0,26]
Fortaleza	-0,04	0,14	[-0,33; 0,24]
São Paulo	-0,17	0,16	[-0,48; 0,13]
Cuiabá	-0,16	0,16	[-0,49; 0,13]
Corinthians	0,01	0,14	[-0,28; 0,28]
Cruzeiro	-0,31	0,17	[-0,65; 0,00]
Vasco	-0,13	0,15	[-0,44; 0,15]
Bahia	0,08	0,14	[-0,21; 0,34]
Santos	-0,16	0,16	[-0,48; 0,14]
Goiás	-0,26	0,16	[-0,58; 0,05]
Coritiba	-0,10	0,15	[-0,40; 0,19]
América-MG	-0,07	0,15	[-0,37; 0,22]

A Tabela 3.10 exibe os efeitos de defesa (*def*). Coritiba e América-MG continuam com os piores valores defensivos, indicando vulnerabilidade, enquanto Palmeiras e Atlético-MG demonstram maior solidez.

Os resultados sugerem que o modelo Poisson Independente captura tendências semelhantes ao modelo hierárquico, mas com variações expressivas. No modelo Poisson Independente, algumas equipes receberam efeitos ofensivos consideravelmente mais altos, como o Palmeiras, cujo efeito de ataque passou de 0,14 no modelo hierárquico para 0,30 no Poisson Independente. Esse aumento indica que o modelo Poisson Independente tende a enfatizar diferenças entre os times, enquanto o modelo hierárquico reduz variações extremas, regularizando os parâmetros.

As Figuras 3.13 e 3.14 apresentam os intervalos de credibilidade de 95% e as medianas estimadas para os parâmetros de ataque e defesa no modelo Poisson Independente.

A Figura 3.13 mantém a tendência de queda nos valores de ataque (*att*), enquanto a Figura 3.14 mantém a tendência de aumentar conforme a equipe ocupa posições inferiores.

A Figura 3.15 apresenta que a distribuição geral das equipes se mantém semelhante à observada no modelo hierárquico, mas com pequenas variações nos posicionamentos.

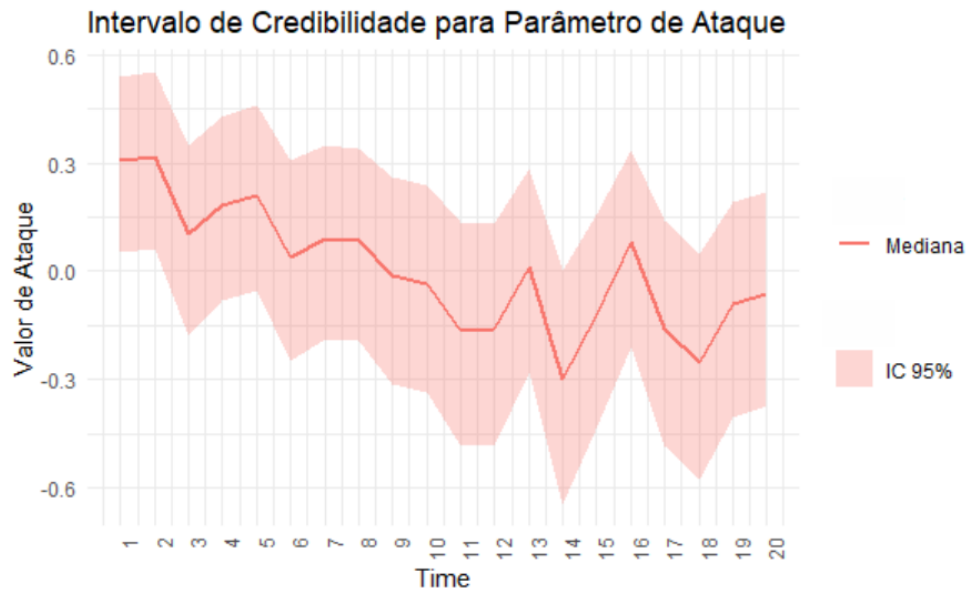


Figura 3.13: Intervalo de Credibilidade do (*att*) para cada time

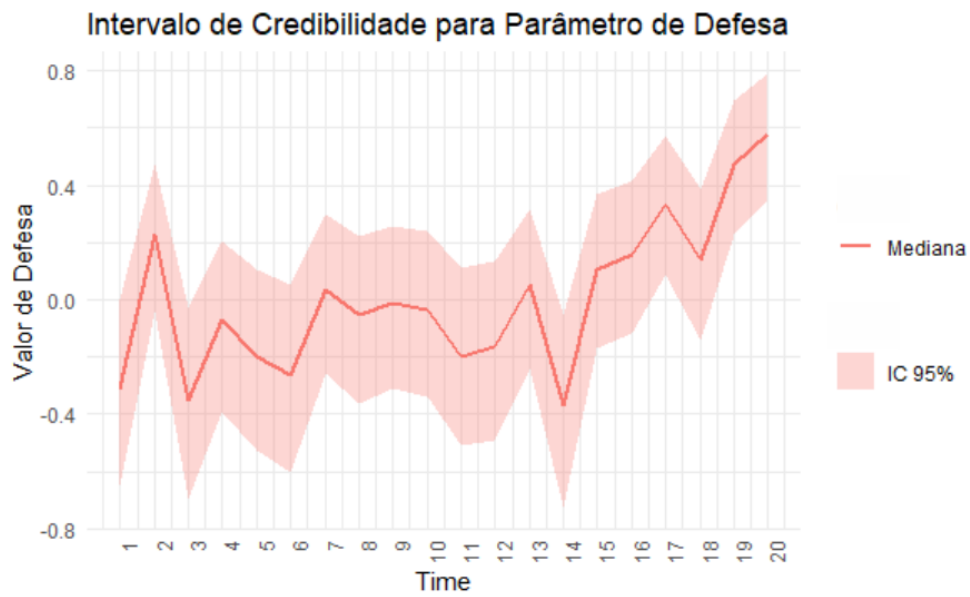


Figura 3.14: Intervalo de Credibilidade da (*def*) para cada time

Tabela 3.10: Efeitos de Defesa (*def*).

Time	Média	DP	IC 95%
Palmeiras	-0,31	0,17	[-0,65; 0,00]
Grêmio	0,22	0,13	[-0,04; 0,47]
Atlético-MG	-0,36	0,17	[-0,70; -0,03]
Flamengo	-0,08	0,15	[-0,39; 0,20]
Botafogo-RJ	-0,20	0,16	[-0,53; 0,11]
Bragantino	-0,27	0,17	[-0,60; 0,06]
Fluminense	0,03	0,14	[-0,26; 0,30]
Athletico-PR	-0,06	0,15	[-0,37; 0,22]
Internacional	-0,02	0,14	[-0,31; 0,26]
Fortaleza	-0,04	0,15	[-0,34; 0,24]
São Paulo	-0,20	0,16	[-0,51; 0,11]
Cuiabá	-0,17	0,16	[-0,49; 0,13]
Corinthians	0,05	0,14	[-0,24; 0,32]
Cruzeiro	-0,38	0,17	[-0,73; -0,05]
Vasco	0,10	0,14	[-0,17; 0,37]
Bahia	0,15	0,13	[-0,12; 0,41]
Santos	0,33	0,12	[0,09; 0,57]
Goiás	0,14	0,14	[-0,14; 0,39]
Coritiba	0,47	0,12	[0,23; 0,70]
América-MG	0,57	0,11	[0,35; 0,79]

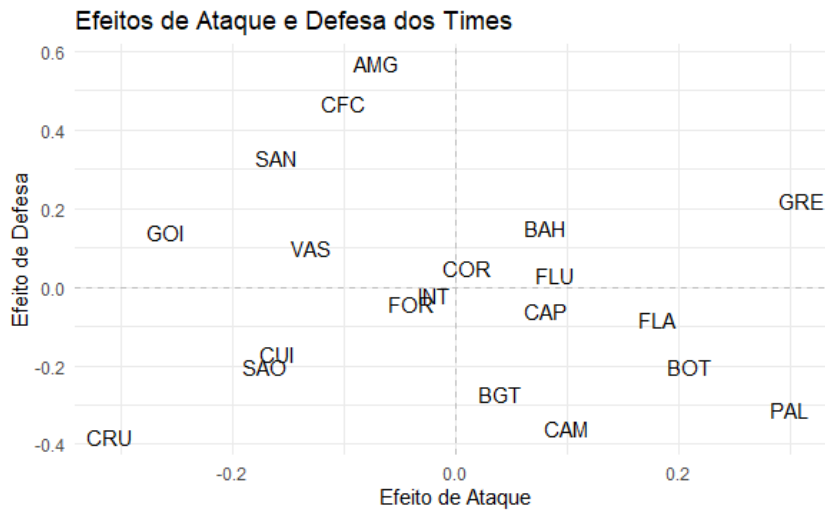


Figura 3.15: Efeitos médios de ataque e defesa

Comparado ao modelo hierárquico, observa-se que o modelo Poisson Independente diferencia melhor os times, evitando a suavização excessiva dos efeitos. Isso sugere que ele é mais eficiente na captação das variações reais entre as equipes, o que pode impactar positivamente as previsões de resultados.

### 3.4.2 Predição de Jogos e Probabilidades de Vitória

A Tabela 3.11 apresenta as probabilidades estimadas de vitória do time A, empate e vitória do time B, considerando o modelo Poisson Independente. Assim como no modelo hierárquico, a vantagem do mando de campo continua evidente, com os mandantes apresentando maior probabilidade de vitória na maioria dos jogos.

Nota-se, no entanto, que este modelo gera probabilidades um pouco mais acentuadas para os times favoritos. Um exemplo disso ocorre no jogo entre Flamengo e Coritiba, onde a probabilidade de vitória do Flamengo ( $PV_A = 0,74$ ) é ainda maior do que no modelo hierárquico. De forma semelhante, Palmeiras contra Cuiabá apresenta um  $PV_A$  superior, reforçando a confiança nas equipes de melhor desempenho.

Além disso, em confrontos equilibrados, como Internacional x Flamengo e Vasco x Palmeiras, as diferenças entre as probabilidades são ligeiramente maiores, o que pode indicar um ajuste mais sensível às diferenças ofensivas e defensivas das equipes.

Tabela 3.11: Probabilidades de Vitória do Time A (PV A), Empate (PE) e Vitória do Time B (PV B) para cada jogo, com os resultados observados.

Jogo	Time A	Time B	PV A	PE	PV B	Resultado Obs
1	Palmeiras	Cuiaba	0,59	0,26	0,15	A
2	America-MG	Fluminense	0,25	0,22	0,52	B
3	Bragantino	Bahia	0,56	0,24	0,20	A
4	Athletico-PR	Goias	0,59	0,24	0,17	A
5	Fortaleza	Internacional	0,44	0,27	0,29	E
6	Botafogo-RJ	Sao Paulo	0,53	0,29	0,19	A
7	Athletico-MG	Vasco	0,60	0,25	0,15	B
8	Flamengo	Coritiba	0,74	0,15	0,11	A
9	Corinthians	Cruzeiro	0,37	0,33	0,30	A
10	Gremio	Santos	0,69	0,17	0,14	A
11	Fluminense	Athletico-PR	0,43	0,26	0,31	A
12	Cuiaba	Bragantino	0,34	0,32	0,35	E
13	Sao Paulo	America-MG	0,66	0,20	0,14	A
14	Cruzeiro	Gremio	0,43	0,28	0,29	A
15	Internacional	Flamengo	0,37	0,27	0,36	A
16	Santos	Athletico-MG	0,20	0,26	0,55	E
17	Vasco	Palmeiras	0,22	0,26	0,52	E
18	Coritiba	Fortaleza	0,29	0,24	0,47	B
19	Goias	Corinthians	0,33	0,29	0,38	A
20	Bahia	Botafogo-RJ	0,32	0,25	0,42	B

### 3.4.3 Comparação entre Pontos Preditos e Observados

A Tabela 3.12 apresenta a comparação entre os pontos esperados e observados no modelo Poisson Independente. Assim como no modelo hierárquico, as previsões são razoavelmente próximas dos valores reais para a maioria das equipes, porém com algumas diferenças notáveis.

Neste modelo, o Palmeiras teve uma expectativa de pontuação ainda maior (71,87) do que no modelo hierárquico, enquanto times como Atlético-MG e Botafogo-RJ também tiveram valores esperados próximos da realidade. Já o Grêmio, que no modelo hierárquico teve um desempenho subestimado, continua superando a pontuação esperada, mas com uma diferença menos acentuada.

Por outro lado, América-MG, Bahia e Cruzeiro seguem sendo os times mais superestimados pelo modelo, enquanto Grêmio, Santos e Flamengos os mais subestimados, sugerindo que este modelo ainda apresenta dificuldades em capturar o desempenho de algumas equipes em situações específicas.

Avaliando a distribuição geral dos erros, o modelo Poisson Independente parece produzir previsões mais ajustadas para os times do topo da tabela, mas ainda tende a suavizar a pontuação de equipes da parte inferior. Uma análise mais detalhada da precisão preditiva dos modelos será realizada na seção comparativa.

### 3.4.4 Avaliação do Modelo Poisson Independente

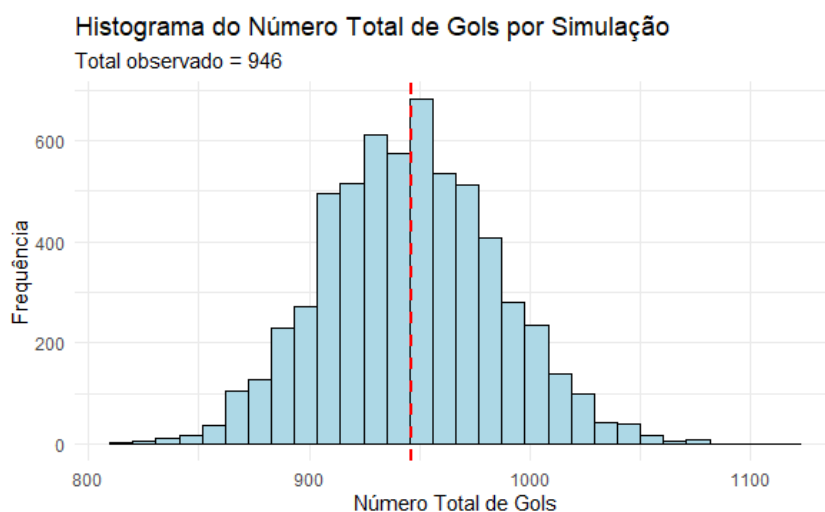


Figura 3.16: Histograma do número total de gols por simulação.

A Figura 3.16 exhibe a distribuição do número total de gols simulados pelo modelo. O

Tabela 3.12: Tabela ordenada por Pontos Observados.

Times	Pontos esperados	Pontos observados
Palmeiras	71,87	70
Grêmio	56,33	66
Atlético-MG	65,61	65
Botafogo-RJ	65,48	63
Bragantino	61,10	63
Flamengo	60,99	63
Fluminense	54,24	55
São Paulo	52,51	53
Cuiabá	51,74	53
Athletico-PR	56,89	56
Fortaleza	52,19	50
Internacional	52,07	50
Corinthians	51,17	50
Cruzeiro	53,00	49
Santos	36,38	44
Vasco	44,82	44
Bahia	50,01	41
Goiás	40,29	41
Coritiba	33,80	37
América-MG	30,94	24

### Histograma do Total de Gols por Simulação

Separado por Gols em Casa e Fora

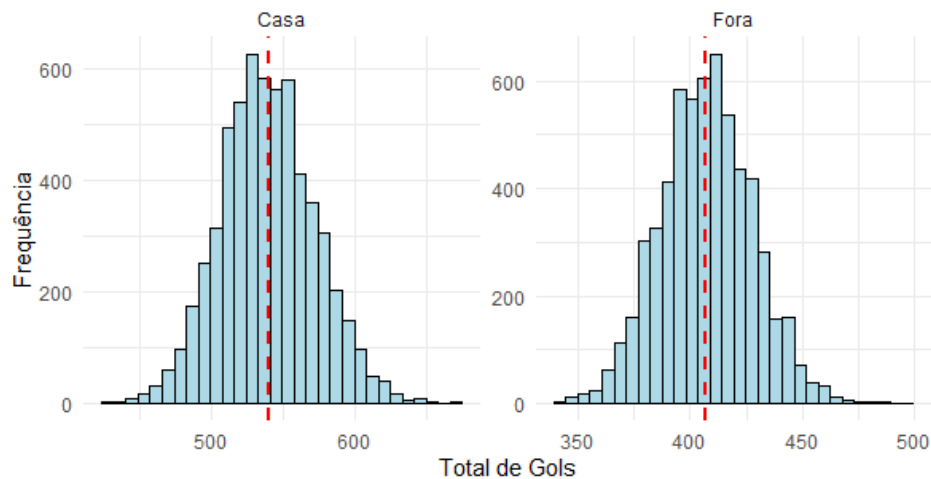


Figura 3.17: Histograma do total de gols por simulação, separado por gols marcados em casa e fora.

número real de gols está bem posicionado dentro da distribuição simulada, indicando que o modelo consegue capturar adequadamente a média total de gols do campeonato.

A Figura 3.17 apresenta a separação dos gols simulados entre mandantes e visitantes. Mantém-se a tendência de os mandantes marcarem mais gols, refletindo a influência do

mando de campo. A dispersão das distribuições sugere uma leve variação na simulação dos gols, cujos impactos serão analisados nas próximas etapas.

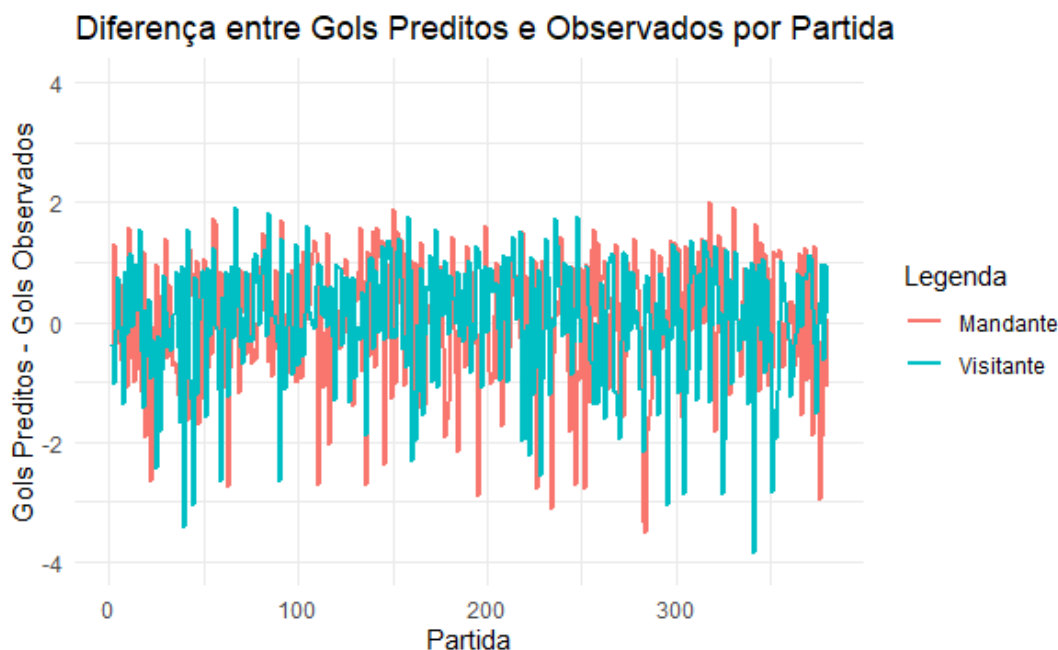


Figura 3.18: Diferença entre gols preditos e observados por jogo.

A Figura 3.18 exibe a diferença entre os gols preditos e observados para cada partida. Assim como no modelo hierárquico, as diferenças oscilam em torno de zero, sugerindo que o modelo não apresenta viés sistemático significativo.

Ainda assim, há partidas em que o modelo subestima ou superestima o número de gols, o que é esperado devido à variabilidade inerente do futebol. A comparação final entre os modelos será analisada na próxima seção.

A Figura 3.19 apresenta as diferenças entre os gols preditos e observados ao longo das rodadas para cada equipe.

Assim como no modelo hierárquico, observa-se variação entre partidas subestimadas e superestimadas, mas com menor amplitude para a maioria das equipes. Isso sugere que o modelo Poisson Independente apresenta uma menor suavização dos resultados, permitindo capturar melhor os desempenhos extremos.

Notamos que times como América-MG, Coritiba e Santos continuam apresentando diferenças mais acentuadas, sugerindo que suas performances ao longo do campeonato foram mais imprevisíveis. Em contrapartida, equipes como Palmeiras, Flamengo e Grêmio mostram previsões mais próximas dos valores reais, indicando um ajuste mais consistente para essas equipes.

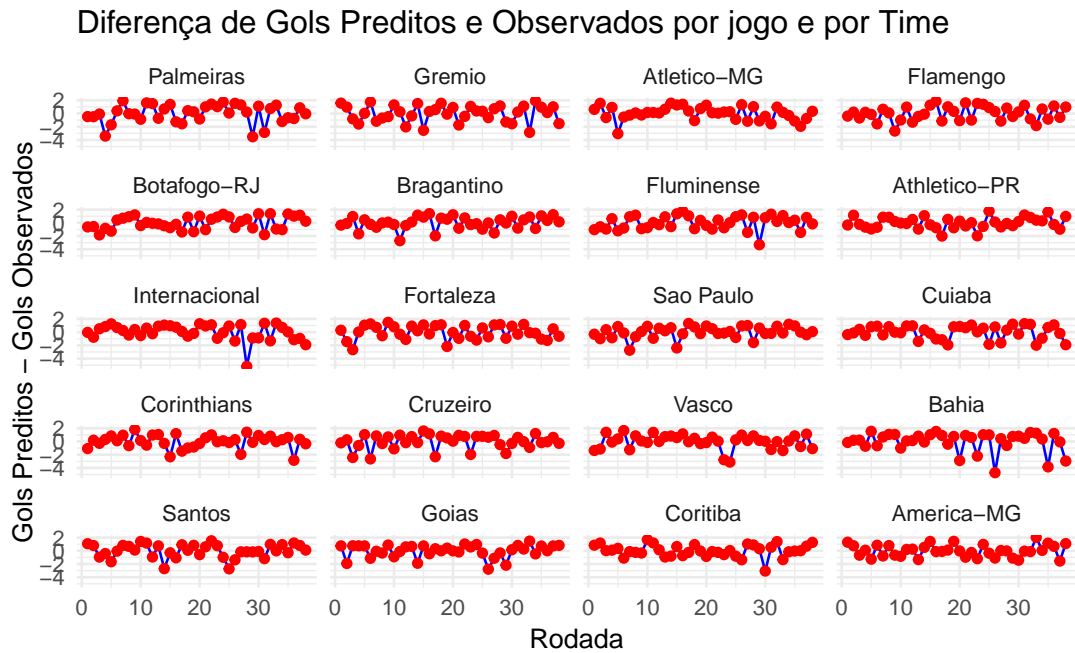


Figura 3.19: Diferença entre gols preditos e observados por equipe.

A estabilidade das previsões ao longo das rodadas reforça que este modelo pode ter um melhor ajuste para capturar tendências ofensivas e defensivas sem a regressão excessiva à média observada no modelo hierárquico.

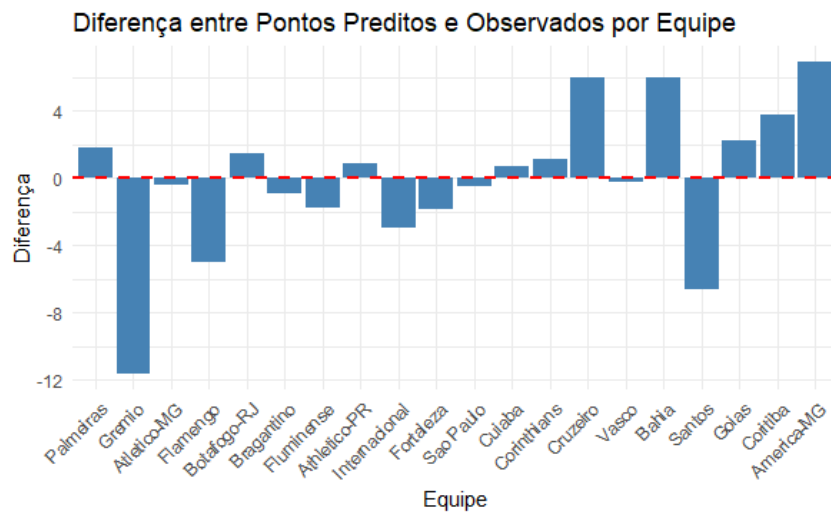


Figura 3.20: Diferença entre pontos preditos e observados por equipe.

A Figura 3.20 apresenta a diferença entre os pontos preditos e observados para cada equipe.

Diferente do modelo hierárquico, que suavizava as previsões, aqui as diferenças são mais variadas. O modelo ainda superestima o desempenho de equipes como América-MG e Coritiba, mas em menor intensidade. Da mesma forma, a subestimação de times do

topo, como Grêmio e Flamengo.

A distribuição mais balanceada sugere que o modelo Poisson Independente pode capturar melhor os desempenhos extremos, evitando a regressão excessiva à média observada no modelo hierárquico. Ainda assim, algumas discrepâncias permanecem, indicando que fatores externos ao modelo podem influenciar os resultados reais.

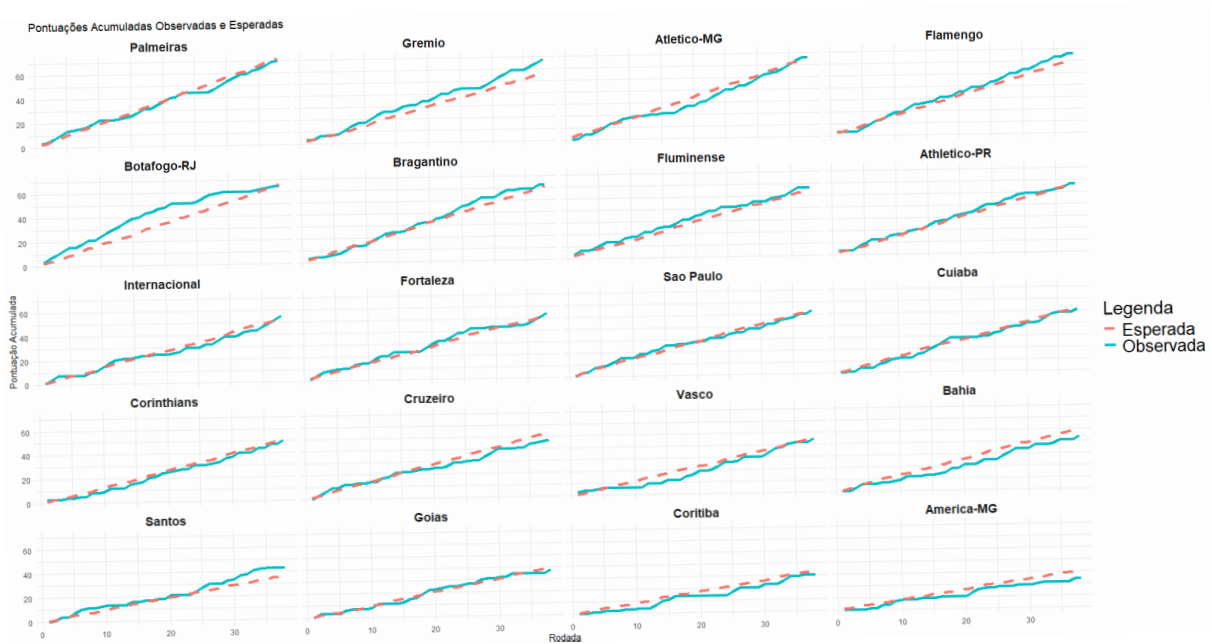


Figura 3.21: Pontuações acumuladas observadas e esperadas ao longo das rodadas.

A Figura 3.21 exibe a evolução das pontuações acumuladas observadas e esperadas ao longo das rodadas, utilizando o modelo Poisson Independente.

Ainda há discrepâncias relevantes, como a subestimação do Botafogo-RJ e a superestimação do Bahia. Ademais, as equipes do meio da tabela continuam apresentando previsões mais alinhadas com a realidade, enquanto os extremos da tabela sofrem maiores desvios.

A Figura 3.22 exibe o mapa de calor das probabilidades de cada equipe terminar em determinada posição na tabela, utilizando o modelo Poisson Independente.

Igualmente ao modelo hierárquico, observa-se que o Poisson Independente apresenta maior dispersão nas probabilidades para times intermediários, como Fortaleza e Internacional, que possuem distribuições mais espalhadas.

Além disso, o Grêmio, que foi também subestimado pelo modelo. Por outro lado, equipes como América-MG e Coritiba continuam sendo projetadas para posições inferiores, o que está alinhado com seus desempenhos no campeonato.

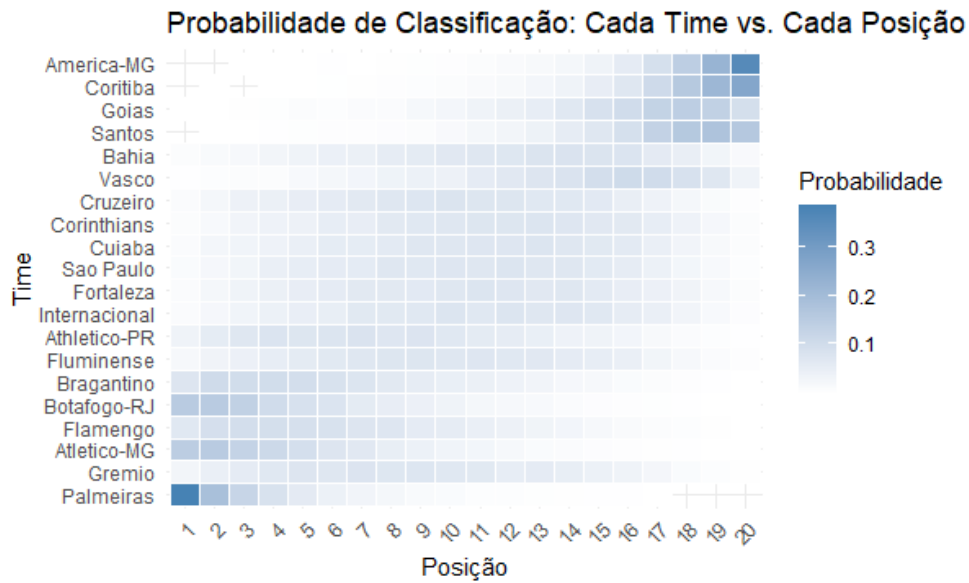


Figura 3.22: Mapa de calor das probabilidades de classificação final de cada equipe.

## 3.5 Comparação Modelo Poisson Independente e Modelo Hierárquico

Nesta seção, apresentamos uma comparação detalhada dos desempenhos preditivos e ajustes estatísticos. O objetivo é avaliar qual abordagem melhor representa o comportamento das equipes no Campeonato Brasileiro.

### 3.5.1 Avaliação do Ajuste

Na Tabela 3.13, apresentamos a comparação entre os pontos esperados pelos modelos Hierárquico e Poisson Independente e os pontos observados ao final do campeonato.

Os resultados mostram que o modelo Poisson Independente geralmente prevê pontuações mais altas para as equipes do topo da tabela, enquanto o modelo hierárquico apresenta previsões mais conservadoras, reduzindo discrepâncias extremas. Essa diferença pode ser observada claramente no caso do Palmeiras, cuja pontuação predita pelo modelo Independente (71,87) foi superior tanto à predição do modelo Hierárquico (63,67) quanto à pontuação real (70).

Entretanto, algumas discrepâncias são notáveis em ambos os modelos. O Grêmio, por exemplo, terminou o campeonato com 66 pontos, mas teve sua pontuação predita de forma mais conservadora pelo modelo hierárquico (52,23), enquanto o modelo independente chegou mais próximo (56,33), mas ainda abaixo do real. Isso sugere que ambos os modelos

Tabela 3.13: Pontos finais observados e preditos pelos modelos.

Times	Hierárquico	Independente	Observado
Palmeiras	63,67	71,87	70
Grêmio	52,23	56,33	66
Atlético-MG	60,64	65,61	65
Botafogo-RJ	56,25	65,48	63
Bragantino	55,27	61,10	63
Flamengo	54,59	60,99	63
Athletico-PR	53,49	56,89	56
Fluminense	52,77	54,24	55
São Paulo	53,07	52,51	53
Cuiabá	52,82	51,74	53
Fortaleza	52,28	52,19	50
Internacional	52,51	52,07	50
Corinthians	51,64	51,17	50
Cruzeiro	53,92	53,00	49
Santos	46,48	36,38	44
Vasco	49,80	44,82	44
Bahia	49,97	50,01	41
Goiás	48,55	40,29	41
Coritiba	44,92	33,80	37
América-MG	43,58	30,94	24

tiveram dificuldades em capturar a performance do time ao longo da temporada.

Outro caso de destaque é o América-MG, que foi amplamente superestimado por ambos os modelos. O modelo hierárquico previu 43,58 pontos, enquanto o modelo Poisson Independente estimou 30,94, quando, na realidade, o time obteve apenas 24 pontos. Esse resultado sugere que o modelo não capturou bem as fragilidades defensivas do time, que foram determinantes para sua má campanha.

### 3.5.2 Teste Qui-Quadrado para Avaliação do Ajuste

Para quantificar a adequação dos modelos às pontuações observadas, realizamos um teste qui-quadrado, comparando os valores previstos por cada modelo com os pontos efetivamente obtidos pelas equipes ao final do campeonato. O teste avalia se há diferenças estatisticamente significativas entre as distribuições preditas e os valores reais.

Os resultados obtidos no teste qui-quadrado foram de  $\chi^2 = 22,60$  ( $p$ -valor = 0,25) para o modelo hierárquico e de  $\chi^2 = 7,62$  ( $p$ -valor = 0,99) para o modelo Poisson Independente.

A análise desses resultados indica que o modelo hierárquico apresenta um valor de  $\chi^2$

relativamente alto, embora o  $p$ -valor = 0,25 não forneça evidências estatísticas fortes para rejeitar a hipótese de que os valores previstos por esse modelo sejam compatíveis com os valores observados. No entanto, observa-se que algumas discrepâncias perceptíveis persistem, especialmente em times como Grêmio e América-MG. Já o modelo Poisson Independente apresentou um valor de  $\chi^2$  consideravelmente menor, com um  $p$ -valor = 0,99, sugerindo que suas previsões estão alinhadas com os valores observados. Isso indica que, estatisticamente, esse modelo oferece um ajuste melhor em relação ao teste qui-quadrado.

Dessa forma, os resultados sugerem que, sob o critério do qui-quadrado, o modelo Poisson Independente apresenta previsões mais próximas dos pontos observados. No entanto, é importante ressaltar que essa métrica não avalia a capacidade preditiva do modelo rodada a rodada, mas sim a proximidade geral entre os valores previstos e os valores reais ao final da temporada.

### 3.5.3 Comparação dos Pontos Acumulados

A Figura 3.23 apresenta a evolução da pontuação acumulada ao longo das rodadas para cada equipe, comparando os valores observados com as previsões dos modelos Hierárquico e Poisson Independente.

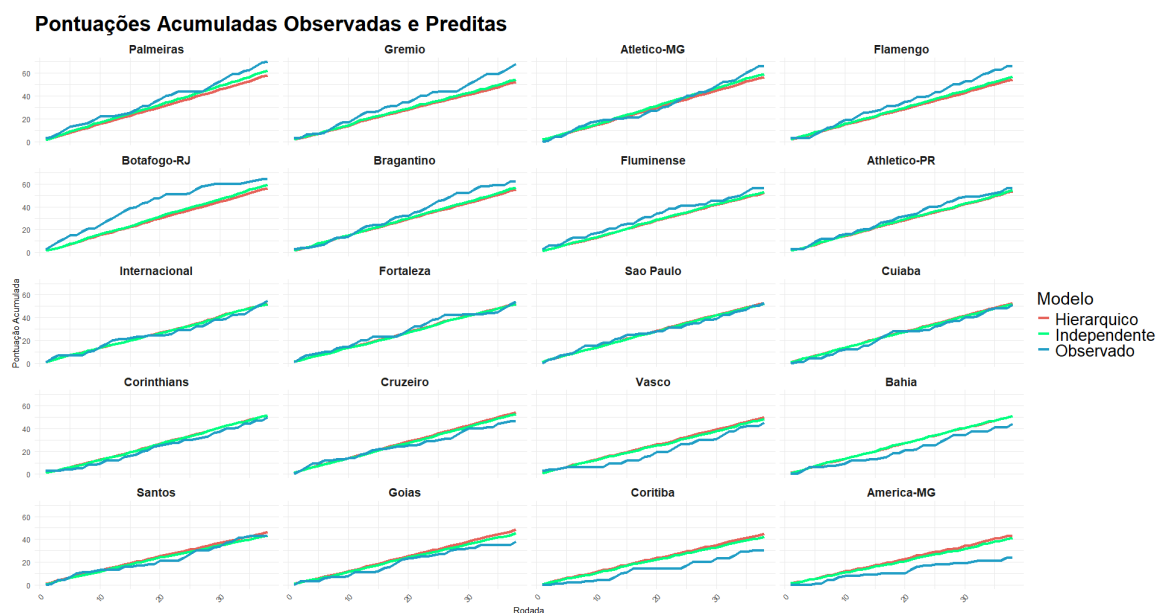


Figura 3.23: Pontuações acumuladas observadas e esperadas para os modelos.

Observa-se que ambos os modelos acompanham majoritariamente bem a trajetória da pontuação real ao longo das rodadas, com pequenas diferenças em algumas equipes. O

modelo Poisson Independente tende a seguir mais de perto a curva observada em algumas situações, enquanto o modelo Hierárquico apresenta um comportamento mais suavizado.

Para algumas equipes, o modelo Poisson Independente parece fornecer um ajuste ligeiramente melhor. Isso pode ser observado em Grêmio, Botafogo-RJ e Bragantino, onde a linha verde (Independente) acompanha mais de perto a trajetória real do que a linha vermelha (Hierárquico). Essa diferença pode estar relacionada ao fato de que o modelo Poisson Independente não impõe uma regularização tão forte nos parâmetros de ataque e defesa, permitindo maior flexibilidade na modelagem do desempenho das equipes.

Por outro lado, há casos em que ambos os modelos apresentam dificuldades em capturar a trajetória real da pontuação, como ocorre com América-MG, Coritiba e Botafogo-RJ. Isso indica que, independentemente da abordagem utilizada, ainda há desafios na modelagem de equipes que apresentaram desempenhos inconsistentes ao longo da competição.

De forma geral, o modelo Poisson Independente demonstra um ajuste ligeiramente superior ao modelo Hierárquico, especialmente para times com desempenhos mais estáveis ao longo da temporada. No entanto, para equipes de desempenho irregular, ambos os modelos apresentam dificuldades semelhantes.

A Figura 3.24 apresenta os boxplots das diferenças entre os pontos acumulados previstos e os pontos observados para os modelos Hierárquico e Poisson Independente.

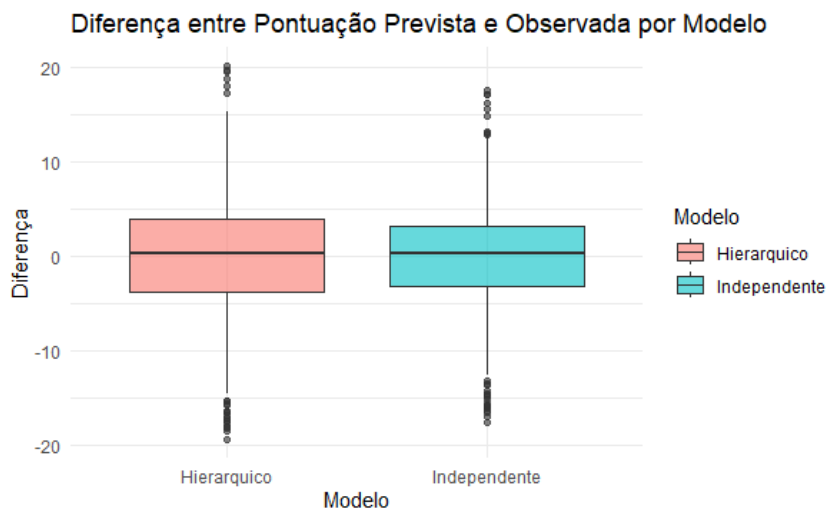


Figura 3.24: Boxplot das Diferenças entre Pontuação Prevista e Observada por Modelo.

Os resultados mostram que ambos os modelos apresentam uma dispersão considerável para as diferenças entre valores previstos e reais, com medianas próximas de zero, mas com uma variabilidade significativa. Os resíduos variam de aproximadamente -20 a +20 pontos, indicando que os modelos frequentemente erram a pontuação acumulada das equipes por

margens consideráveis. Esse nível de erro sugere que, embora os modelos sejam capazes de capturar tendências gerais do campeonato, eles possuem limitações substanciais para prever a pontuação exata das equipes ao longo das rodadas.

O modelo Poisson Independente apresenta uma distribuição de resíduos ligeiramente menos dispersa, sugerindo previsões mais estáveis e menos propensas a erros extremos em comparação com o modelo hierárquico. Em contrapartida, o modelo Hierárquico tende a suavizar as previsões, reduzindo discrepâncias extremas, mas, ao mesmo tempo, pode resultar em previsões mais conservadoras que não refletem variações individuais significativas no desempenho das equipes. Isso ocorre porque a regularização imposta pelo modelo hierárquico limita a variação dos parâmetros estimados, reduzindo a sensibilidade a desempenhos atípicos.

Outro aspecto relevante é a presença de *outliers* em ambos os modelos, o que indica rodadas em que as previsões se distanciaram substancialmente dos valores observados. Esses desvios podem estar relacionados a eventos inesperados, como mudanças táticas, desfalques, expulsões ou até fatores psicológicos, que não foram diretamente considerados nos modelos.

A análise dos resultados demonstra a eficácia dos modelos propostos como ferramentas para capturar a dinâmica probabilística do campeonato, com sua principal força residindo na capacidade de estimar as probabilidades de vitória, empate e derrota em jogos individuais e identificar as tendências de desempenho das equipes. Contudo, a análise dos resíduos revelou uma dispersão considerável na predição da pontuação final acumulada, o que não invalida os modelos, mas delinea seu escopo de aplicação. Essa variabilidade indica que, embora sejam razoáveis para a avaliação de confrontos pontuais, a previsão como a pontuação exata é um desafio mais complexo, sujeito à propagação de incertezas ao longo das 38 rodadas.

### 3.5.4 Análise dos Gols Previstos e Observados

A Figura 3.25 e a Figura 4.1 apresentam os *boxplots* das diferenças entre os gols previstos pelos modelos e os gols observados para times mandantes e visitantes, respectivamente.

No gráfico 3.25, ambos os modelos apresentam medianas próximas de zero, indicando ausência de viés sistemático na previsão dos gols dos mandantes. No entanto, a dispersão considerável dos resíduos sugere erros significativos em partidas específicas. *Outliers* negativos indicam que os modelos frequentemente subestimam os gols dos mandantes, com

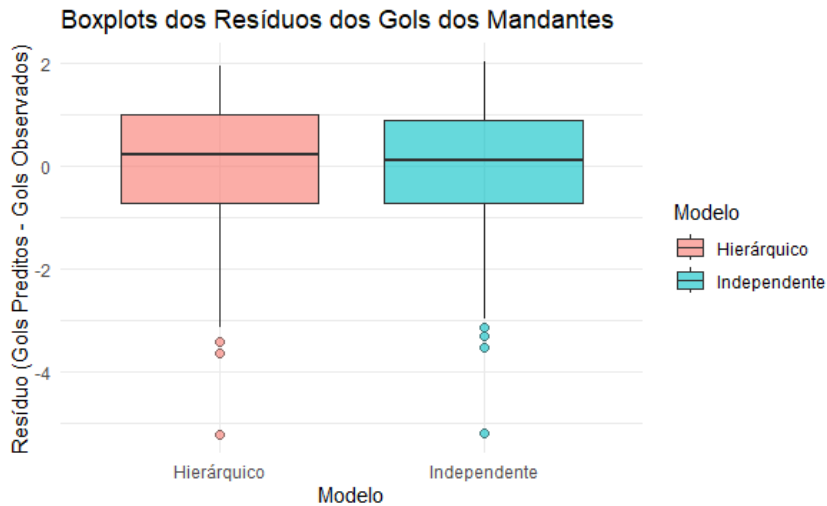


Figura 3.25: Boxplot das Diferenças entre Gols Previstos e Observados para Times Mandantes.

o modelo Poisson Independente mostrando maior estabilidade.

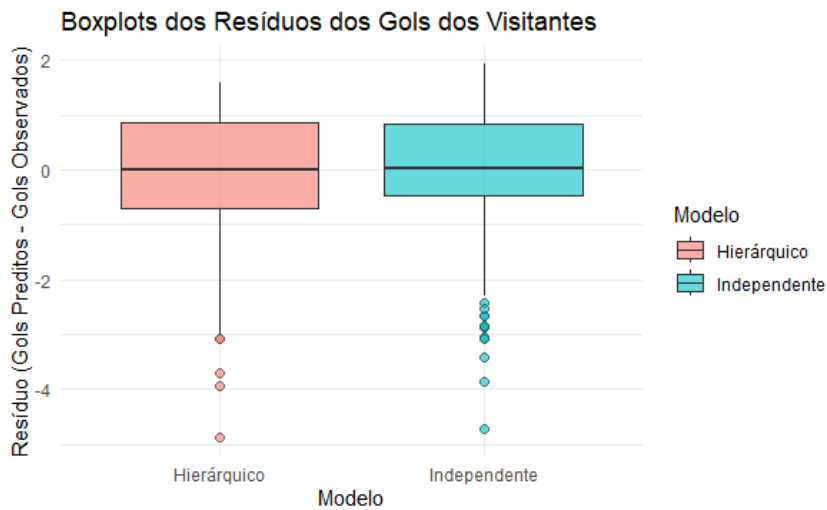


Figura 3.26: Boxplot das Diferenças entre Gols Previstos e Observados para Times Visitantes.

No gráfico 4.1, a dispersão dos resíduos dos visitantes é semelhante à dos mandantes, com ampla variabilidade e *outliers* indicando erros extremos. O modelo Poisson Independente apresenta maior frequência de outliers associados a subestimações dos gols dos visitantes.

Ambos os modelos apresentam erros consideráveis ao prever o número exato de gols tanto em casa quanto fora, demonstrando que a variabilidade dos jogos de futebol ainda é um desafio significativo para modelagem estatística. A presença de outliers indica que existem partidas em que os modelos erram substancialmente, o que pode ser resultado de fatores não modelados, como eventos inesperados durante o jogo (expulsões, pênaltis,

lesões) ou variações táticas das equipes.

Embora os modelos sejam capazes de capturar tendências gerais e oferecer previsões razoáveis para a média de gols, suas limitações na previsão precisa do número de gols reforçam a ideia de que são mais adequados para estimar probabilidades de vitória, empate e derrota, em vez de fornecer previsões exatas do placar final.



# Capítulo 4

## Análise da qualidade preditiva dos modelos

Neste capítulo, dedicamo-nos a uma avaliação da capacidade preditiva fora da amostra. Estabelecemos o protocolo de validação sequencial (Seção 4.1), avaliamos a qualidade das probabilidades com regras de escore e diagramas de calibração (Seções 4.2 e 4.3), investigamos a performance de classificação categórica (Seções 4.5 e 4.6) e, por fim, apresentamos uma análise de cenários (Seção 4.7).

## 4.1 O Conjunto de Dados para Avaliação

Realizamos a avaliação do desempenho preditivo por meio de um esquema de atualização sequencial. Para a primeira previsão (20<sup>a</sup> rodada), ajustamos o modelo com os dados das 19 primeiras. A partir daí, a cada rodada, incorporamos os resultados observados e reajustamos o modelo. Repetimos esse processo até a 38<sup>a</sup> rodada.

Para contextualizar e validar esta abordagem, é útil comparar a distribuição dos resultados no período de teste (rodadas 20-38) com a do período que serviu como base inicial para o treinamento (rodadas 1-19). As Tabelas 4.1 e 4.2 apresentam as frequências observadas em cada uma dessas metades do campeonato.

Tabela 4.1: Frequências observadas nas rodadas 1 a 19.

Resultado	Frequência	Proporção
Mandante	91	0,479
Empate	51	0,268
Visitante	48	0,253

Tabela 4.2: Frequências observadas nas rodadas 20 a 38.

Resultado	Frequência	Proporção
Mandante	87	0,457
Empate	47	0,247
Visitante	56	0,294

A comparação entre as tabelas revela que as proporções de resultados são bastante similares nos dois períodos. Essa estabilidade na distribuição dos desfechos é importante, pois sugere que os modelos foram testados em um ambiente com características semelhantes às do seu treinamento inicial, o que confere maior validade à avaliação preditiva. Nota-se também que, em ambas as metades, os resultados que não são a vitória do mandante (empate ou vitória do visitante) somam mais de 50% dos jogos, reforçando o desafio inerente à previsão no futebol.

## 4.2 Análise das Cadeias

Para a estimação dos parâmetros de ambos os modelos, utilizamos 3 cadeias de Markov (MCMC) paralelas. Executamos cada cadeia por 1.500 iterações e descartamos as 500

iniciais, resultando em um total de 3.000 amostras da distribuição a posteriori para a inferência de cada parâmetro.

No modelo de Poisson Independente, iniciamos os parâmetros (*home*) e os efeitos das equipes (*att.star*, *def.star*) com valores centrados em 0, 1 e -1 para cada cadeia (ver Apêndice B).

No modelo Hierárquico, uma estratégia análoga foi aplicada. Os parâmetros de efeito (*home*, *att.star*, *def.star*) e os hiperparâmetros de localização (*mu.att*, *mu.def*) foram iniciados em valores dispersos (0, 1, -1). Adicionalmente, iniciamos os hiperparâmetros de precisão (*tau.att*, *tau.def*) em valores distintos (1,0; 0,5; 0,01) para assegurar que as cadeias partissem de regiões com diferentes variâncias (ver Apêndice B).

A verificamos a convergência para todos os parâmetros relevantes em ambos os modelos. Em todas as simulações realizadas para as previsões, o critério de Gelman-Rubin apresentou valores próximos de 1, indicando que as cadeias convergiram para a mesma distribuição estacionária.

### 4.3 Regras de escore (*Scoring Rules*)

Nesta seção, comparamos os modelos Independente e Hierárquico por meio de boxplots e estatísticas descritivas de cada uma das três regras de escore (Brier, logarítmica e esférica), conforme ilustrado na Figura 4.1 e resumido nas Tabelas 4.3 a 4.5.

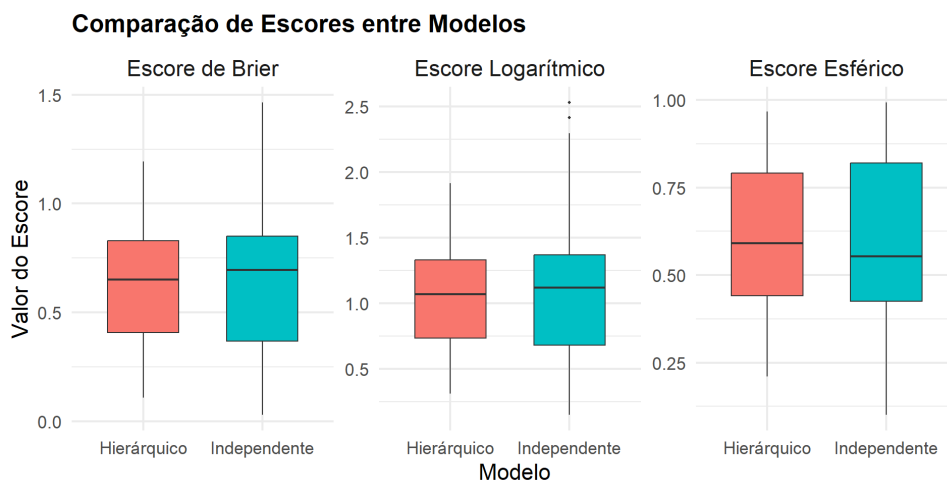


Figura 4.1: Boxplot dos Escores.

Do boxplot geral, nota-se que o modelo Hierárquico apresenta sutilmente menor variabilidade nas três regras, com quartis mais concentrados em torno de valores inferiores

para *Brier* e *Logarithmic* e valores mais altos para *Spherical*.

Tabela 4.3: Estatísticas descritivas do *Brier Score*

Modelo	Mínimo	Q1	Mediana	Média	Q3	Máximo	DP
Independente	<b>0,03</b>	<b>0,37</b>	0,69	0,64	0,85	1,46	0,30
Hierárquico	0,10	0,40	<b>0,65</b>	<b>0,63</b>	<b>0,83</b>	<b>1,19</b>	<b>0,24</b>

Na Tabela 4.3, o modelo Independente atinge o menor valor mínimo de Brier (0,03), mas também o maior máximo (1,46) e maior desvio-padrão (0,30), indicando maior instabilidade; o Hierárquico, embora não alcance o Mínimo global, tem Q1, Mediana, Média, Q3 e DP inferiores, sinalizando previsões de erro mais consistentes.

Tabela 4.4: Estatísticas descritivas do *Logarithmic Score*

Modelo	Mínimo	Q1	Mediana	Média	Q3	Máximo	DP
Independente	<b>0,15</b>	<b>0,68</b>	1,12	1,07	1,37	2,53	0,44
Hierárquico	0,31	0,73	<b>1,07</b>	<b>1,05</b>	<b>1,33</b>	<b>1,92</b>	<b>0,34</b>

Tabela 4.5: Estatísticas descritivas do *Spherical Score*

Modelo	Mínimo	Q1	Mediana	Média	Q3	Máximo	DP
Independente	0,10	0,42	0,55	0,59	<b>0,82</b>	<b>0,99</b>	<b>0,22</b>
Hierárquico	<b>0,21</b>	<b>0,44</b>	<b>0,59</b>	<b>0,60</b>	0,79	0,96	0,19

O padrão se repete para o *Logarithmic Score* (Tabela 4.4). O modelo Independente novamente alcança o menor valor mínimo (0,15), mas mostra maior instabilidade com valores extremos que chegam a 2,53. Por outro lado, o modelo Hierárquico, apesar de ter quartis ligeiramente maiores, comprova sua maior estabilidade com um valor máximo mais contido (1,92) e um desvio-padrão reduzido (0,34 contra 0,44).

Já no *Spherical Score* (Tabela 4.5), onde valores mais altos implicam melhor desempenho, o Independente atinge os picos máximos de 0,99 e 0,82 no Q3, mas novamente com maior dispersão (DP = 0,22). O Hierárquico, embora tenha Mínimo e Q1 menores, apresenta Mediana e Média superiores (0,59 e 0,60 versus 0,55 e 0,59), o que sugere que, no geral, suas probabilidades estão mais próximas de 1 (melhor calibração e discriminação).

Em resumo, os boxplots e as estatísticas descritivas indicam que o modelo Hierárquico é mais estável e consistente em todas as regras de escore, enquanto o modelo Independente consegue extremos pontuais ligeiramente melhores em alguns casos, mas paga esse desempenho com variabilidade muito maior. “

Para observar qual modelo apresenta desempenho superior ao longo das rodadas, calculamos a diferença entre as curvas LOESS suavizadas dos escores do modelo Hierárquico e do modelo Independente (Hierárquico menos Independente jogo a jogo). As Figuras 4.2, 4.3 e 4.4 mostram essa diferença para o Escore de Brier, o Escore Logarítmico e o Escore Esférico, respectivamente. Nos dois primeiros casos valores acima de zero indicam desempenho melhor do modelo Independente, enquanto valores abaixo de zero indicam desempenho melhor do modelo Hierárquico. Para o Escore Esférico ocorre o contrário, pois nesse escore valores maiores correspondem a previsões mais precisas.

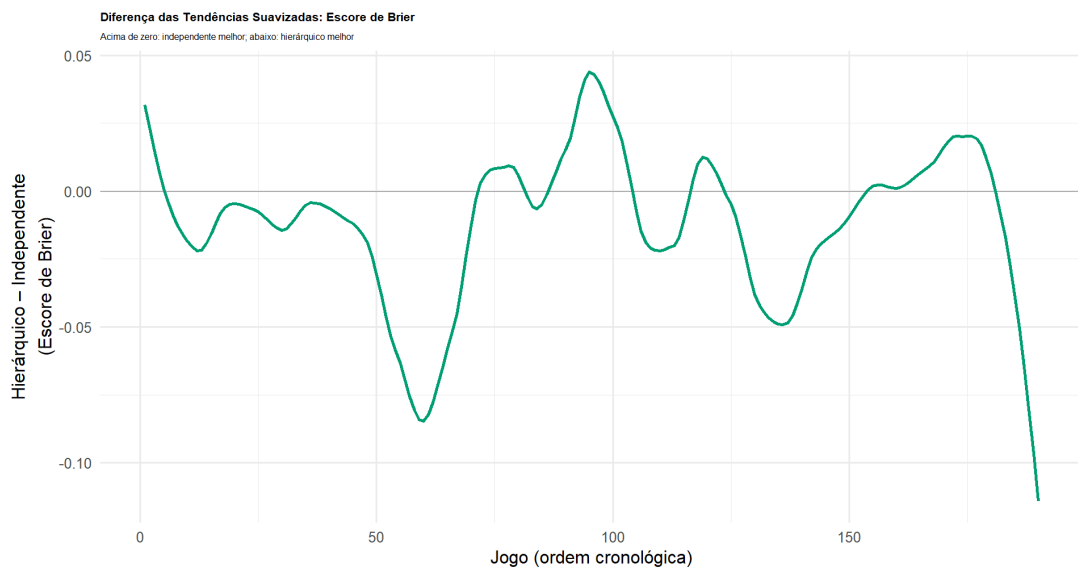


Figura 4.2: Escore de Brier jogo a jogo.

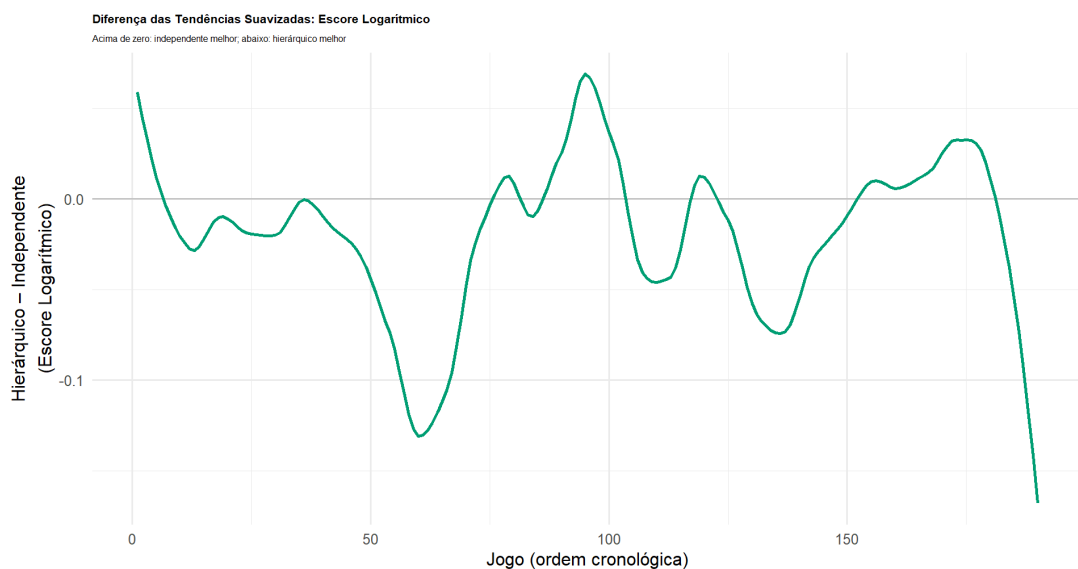


Figura 4.3: Escore Logarítmico jogo a jogo.

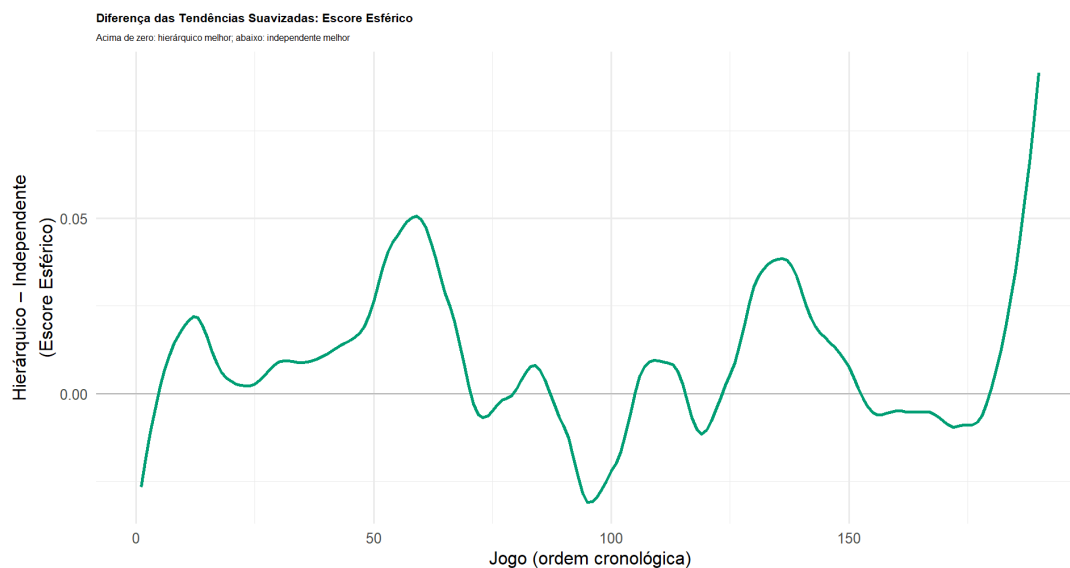


Figura 4.4: Escore Esférico jogo a jogo.

Na Figura 4.2, que detalha o Escore de Brier, nota-se que ambos os modelos apresentam forte variabilidade jogo a jogo, reflexo da imprevisibilidade inerente ao futebol. No entanto, o modelo Hierárquico suavizado quase sempre mostra uma tendência menor que a do modelo Independente, indicando Escore de Brier ligeiramente menor em grande parte das rodadas, com picos de pior desempenho menos pronunciados.

O Escore Logarítmico da Figura 4.3 tem um padrão semelhante com o do Escore de Brier. Tendo o modelo Hierárquico com valores quase sempre abaixo que o Independente e picos ligeiramente menores.

Na Figura 4.4, nas primeiras rodadas o valor fica abaixo de zero, indicando que o modelo Independente apresentava leve vantagem nesse início. A partir do jogo vinte até aproximadamente o 60 a diferença permanece positiva. Há uma breve oscilação para valores negativos perto do jogo 80, mas durante a maior parte dos 190 jogos a vantagem do Hierárquico se mantém, com picos positivos que chegam a cerca de 0,05. No final do campeonato, mais precisamente nas últimas dez partidas, nota-se um salto pronunciado acima de zero, evidenciando que o modelo Hierárquico alcançou escores Esféricos superiores ao modelo Independente, o que também foi observado nos escores de Brier e Logarítmico.

## 4.4 Análise de desempenho do modelo

Para complementar as regras de escore probabilístico, avaliamos também a capacidade dos modelos em classificar corretamente cada partida em Vitória do Mandante, Empate

ou Vitória do Visitante. Para isso, construímos a matriz de confusão, que registra o número de acertos e erros em cada combinação de classe prevista versus observada. A partir dessa matriz, calculamos métricas de desempenho para problemas de classificação multiclasse.

Para aprofundar a análise de como os modelos atribuem probabilidades, as Figuras 4.5 e 4.6 visualizam as distribuições das probabilidades previstas para cada classe, separadas por seu status como verdadeiro positivo, falso positivo, falso negativo ou verdadeiro negativo.

Uma análise visual dos gráficos revela nuances sobre o poder de discriminação dos modelos. O principal sinal de que os modelos aprenderam um padrão útil está na separação entre os acertos. Para as classes Mandante e Visitante, as distribuições de Verdadeiros Positivos estão visivelmente deslocadas para a direita em relação às de Verdadeiros Negativos. No entanto, a dificuldade na discriminação é evidenciada pela sobreposição dessas distribuições com as dos erros (Falsos Positivos e Falsos Negativos). A maior limitação, contudo, reside na classe Empate, onde todas as distribuições se comprimem em valores baixos, resultando em uma sobreposição generalizada e baixo poder preditivo.

Ao comparar os dois modelos, nota-se uma diferença clara na dispersão das previsões. As distribuições de probabilidade do modelo Hierárquico (Figura 4.6) são consistentemente mais concentradas e menos dispersas do que as do modelo Independente (Figura 4.5), que por sua vez gera previsões mais espalhadas e com caudas mais longas.

Essa diferença de comportamento é um reflexo direto da estrutura de cada modelo. A maior concentração do modelo Hierárquico é resultado do seu efeito de regularização, ou encolhimento, que suaviza as estimativas.

Nas Figuras 4.7 e 4.8 percebe-se que ambos os modelos nunca chegaram a prever empates, pois a linha correspondente encontra-se completamente vazia em ambas as matrizes. Todo o esforço de classificação concentrou-se, portanto, na distinção entre “Mandante” e “Visitante”. No modelo Independente (Figura 4.7), de 87 jogos vencidos pelo mandante, apenas 61 foram corretamente identificados, ao passo que o modelo Hierárquico (Figura 4.8) acertou 74 dessas 87 partidas.

Esse comportamento se reflete nas métricas apresentadas nas Tabelas 4.6 a 4.8. Para a classe Mandante, o modelo Hierárquico alcança *recall* de 0,85 (contra 0,70 do Independente) e  $F_1$  de 0,61 (ante 0,57), embora sua precisão seja ligeiramente inferior (0,47 vs. 0,48). Já para a classe Visitante, o Independente obtém *recall* de 0,39, superior ao 0,25

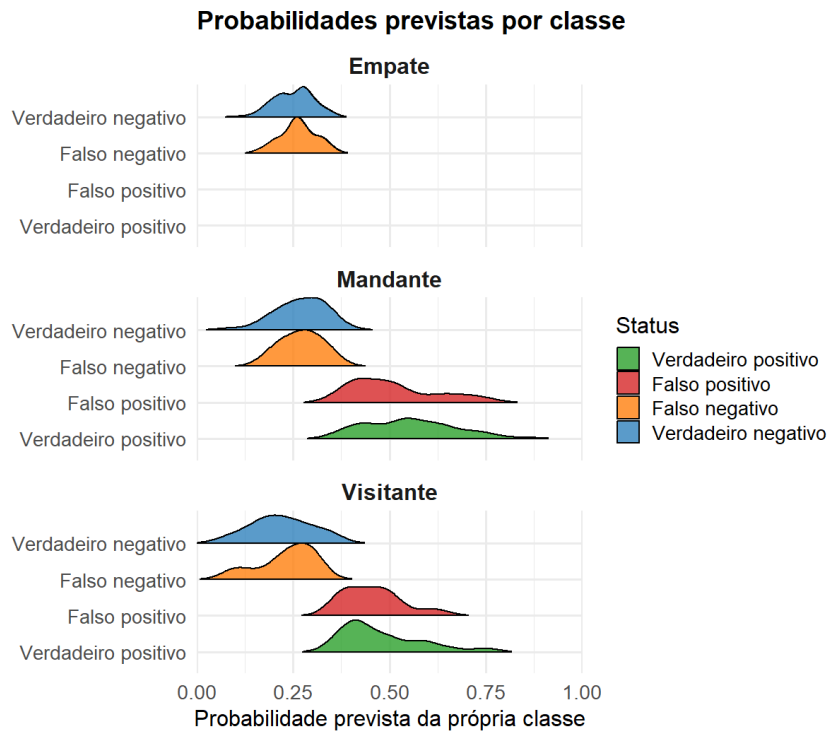


Figura 4.5: Distribuição das probabilidades previstas por classe para o modelo Independente.

do Hierárquico, e  $F_1$  de 0,37 contra 0,31, enquanto o Hierárquico exibe precisão um pouco melhor (0,41 vs. 0,36). Em termos de acurácia global (Tabela 4.8), o modelo Hierárquico também apresenta leve vantagem, com 46% de acertos frente a 43% do Independente.

Tabela 4.6: Métricas para a classe Mandante

Modelo	<i>Recall</i>	Precisão	$F_1$
Independente	0,70	<b>0,48</b>	0,57
Hierárquico	<b>0,85</b>	0,47	<b>0,61</b>

Tabela 4.7: Métricas para a classe Visitante

Modelo	<i>Recall</i>	Precisão	$F_1$
Independente	<b>0,39</b>	0,36	<b>0,37</b>
Hierárquico	0,25	<b>0,41</b>	0,31

Tabela 4.8: Acurácia global dos modelos

Modelo	Acurácia
Independente	0,43
Hierárquico	<b>0,46</b>

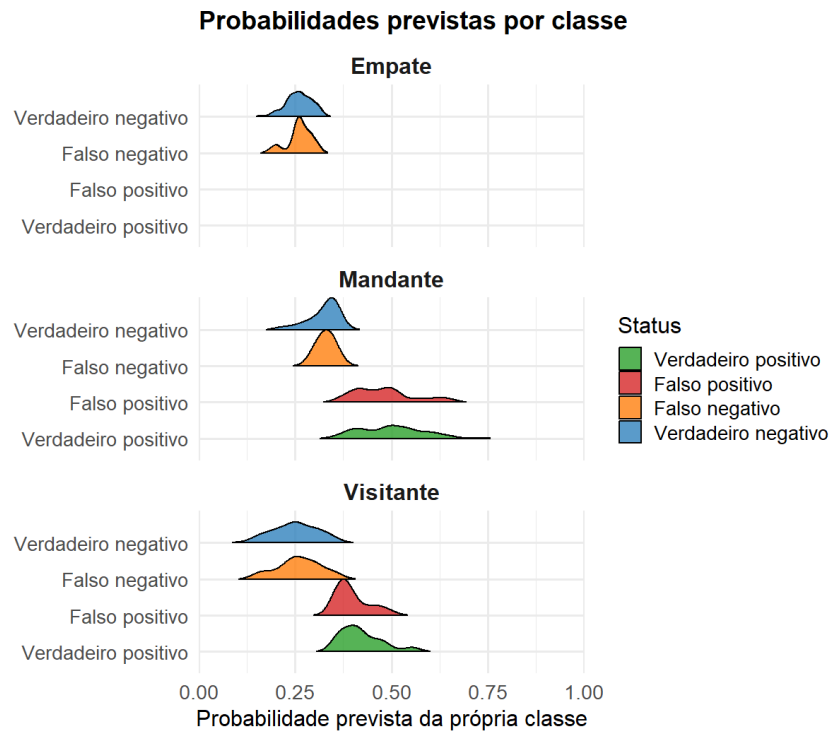


Figura 4.6: Distribuição das probabilidades previstas por classe para o modelo Hierárquico.

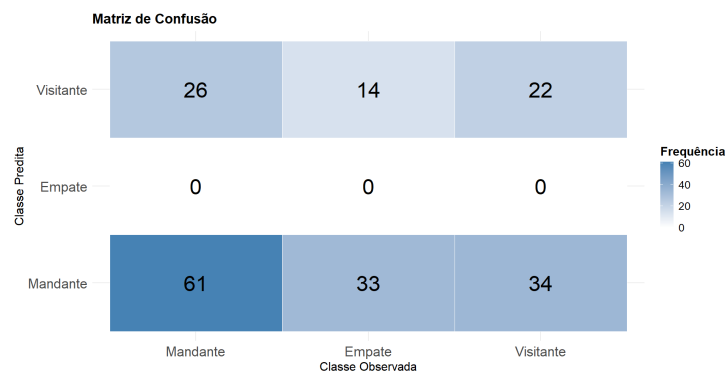


Figura 4.7: Tabela de Confusão para modelo Independente

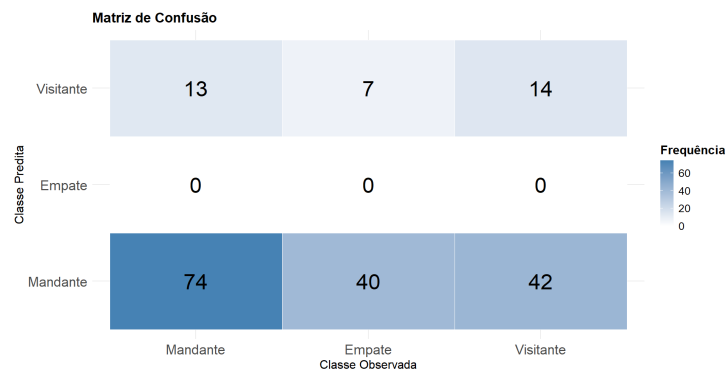


Figura 4.8: Tabela de Confusão para modelo Hierárquico

Em síntese, o modelo Hierárquico mostra-se mais eficaz na identificação de vitórias do mandante e apresenta maior acurácia geral, ao passo que o modelo Independente se sai um pouco melhor ao capturar vitórias do visitante.

Na Figura 4.9 observa-se que a acurácia dos dois modelos flutua bastante de rodada a rodada, refletindo a imprevisibilidade intrínseca de cada confronto. No início do segundo turno (rodadas 20–24), ambos alcançam níveis moderados de acerto, em torno de 40–50%, com pequenas alternâncias de vantagem entre Independente e Hierárquico. A rodada 26 marca uma queda mais pronunciada, sobretudo para o modelo Independente, enquanto o Hierárquico mantém desempenho perto de 30%. Já nas rodadas 27 a 29, verifica-se um período de bom acerto (acima de 60%), momento em que o Hierárquico atinge duas rodadas consecutivas de 60% e o Independente chega a 70% na rodada 29.

A virada de turno (rodada 30) traz novo declínio e, especialmente na rodada 33, o modelo Independente registra acurácia zero, enquanto o Hierárquico, apesar de também baixo, não chega a falhar em todas as partidas. O desempenho de ambos se recupera nas últimas cinco rodadas, com o pico de performance ocorrendo na rodada 37, quando a acurácia se aproximou de 70%. A temporada se encerra com os modelos apresentando uma acurácia entre 40% e 50% na rodada 38.

Em síntese, embora nenhum dos modelos entregue acurácia estável ao longo de toda a sequência, o Hierárquico exibe quedas menos extremas em pontos críticos e uma recuperação mais homogênea no fim do campeonato, enfatizando sua maior robustez frente a oscilações bruscas.

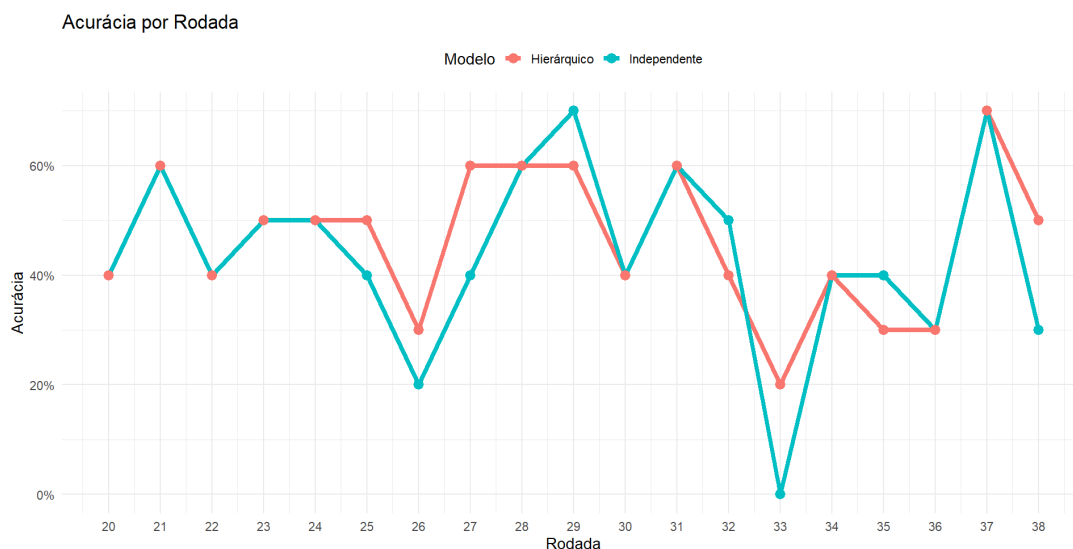


Figura 4.9: Acurácia rodada a rodada

## 4.5 Teste KS

Para avaliar a capacidade dos modelos em discriminar entre os resultados de uma partida (vitória do mandante, empate ou vitória do visitante), empregamos o teste de Kolmogorov-Smirnov (KS). Este teste é usado para verificar se as distribuições de probabilidade previstas pelo modelo para os eventos que de fato ocorreram são significativamente diferentes daquelas para os eventos que não ocorreram.

As Figuras 4.10 e 4.11 são o fundamento visual desta análise. Elas apresentam as Funções de Distribuição Acumulada (FDA) das probabilidades previstas, comparando duas amostras:

- Amostra Positiva (curva azul, “Verdadeiro”): As probabilidades que o modelo atribuiu aos resultados que de fato aconteceram (por exemplo, as probabilidades de vitória do mandante em todos os jogos que o mandante venceu).
- Amostra Negativa (curva vermelha, “Falso”): As probabilidades que o modelo atribuiu para o mesmo resultado, mas nos jogos em que ele não ocorreu (por exemplo, as probabilidades de vitória do mandante em todos os jogos que terminaram em empate ou vitória do visitante).

Uma boa capacidade discriminativa se traduz em uma separação vertical clara entre as duas curvas. Idealmente, um bom modelo atribui probabilidades baixas aos casos “Falsos” e probabilidades altas aos “Verdadeiros”. Isso faz com que a curva vermelha (Falso) suba rapidamente no início (acumulando os casos de baixa probabilidade) e a curva azul (Verdadeiro) suba de forma mais lenta, concentrando seu crescimento em valores de probabilidade mais altos. Visualmente, portanto, a interpretação correta de um bom poder discriminatório é observar a curva vermelha posicionada consistentemente acima da curva azul. O teste KS quantifica a maior distância vertical entre essas curvas (a estatística  $D$ ), e a linha tracejada vertical nos gráficos indica o exato ponto de corte da probabilidade onde essa separação máxima ocorre.

No modelo Independente, a separação mais expressiva dá-se para a classe Mandante em  $u = 0,53$  com  $D = 0,212$  e  $p\text{-valor} = 0,023$ , indicando discriminação estatisticamente significativa. Para Empate ( $u = 0,24$ ,  $D = 0,171$ ,  $p = 0,209$ ) e Visitante ( $u = 0,26$ ,  $D = 0,201$ ,  $p = 0,067$ ), embora ainda perceptível, a distinção não atinge o nível de 5

No modelo Hierárquico, os valores de  $D$  são ligeiramente menores e mais homogêneos: Mandante em  $u = 0,52$  ( $D = 0,189$ ,  $p = 0,055$ ), Empate em  $u = 0,25$  ( $D = 0,193$ ,  $p$

= 0,108) e Visitante em  $u = 0,34$  ( $D = 0,160$ ,  $p = 0,225$ ). Isso sugere que o Hierárquico mantém capacidade discriminativa equilibrada, ainda que com distinções menos extremas do que o Independente.

Em síntese, o modelo Independente demonstra maior poder de separação para vitórias do mandante, mas perde robustez nas demais classes. O modelo Hierárquico, por sua vez, exibe valores de KS mais homogêneos e p-valores menos extremos, refletindo uma capacidade discriminativa equilibrada e menos sujeita a flutuações.

Tabela 4.9: Pontos de corte  $u$ , estatística  $D$  e p-valor do teste KS por modelo e classe

Modelo	Classe	Corte $u$	$D$	p-valor	$n_{\text{classe}}$	$n_{\text{outros}}$
Independente	Mandante	0,53	0,212	0,023	87	103
	Empate	0,24	0,171	0,209	47	143
	Visitante	0,26	0,201	0,067	56	134
Hierárquico	Mandante	0,52	0,189	0,055	87	103
	Empate	0,25	0,193	0,108	47	143
	Visitante	0,34	0,160	0,225	56	134

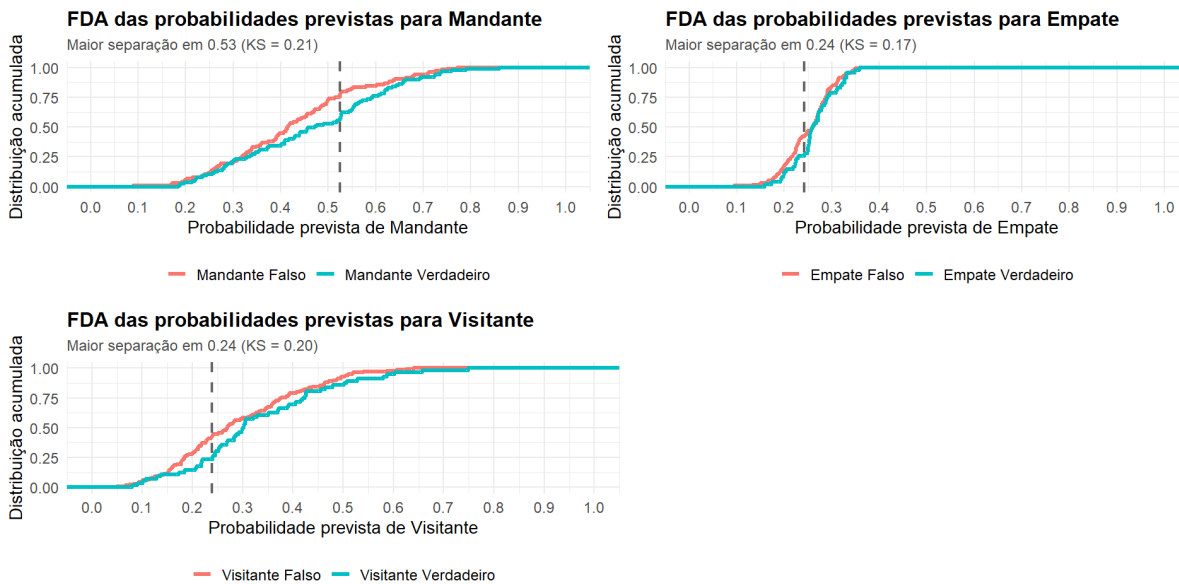


Figura 4.10: FDA das probabilidades previstas pelo modelo Independente.

## 4.6 Medidas de Concordância

Complementarmente às métricas já apresentadas, empregamos duas medidas de concordância para avaliar a qualidade das classes previstas: o *Cohen's kappa* e o coeficiente de correlação de Matthews (*MCC*). O *Cohen's kappa* ajusta o índice de acerto pela

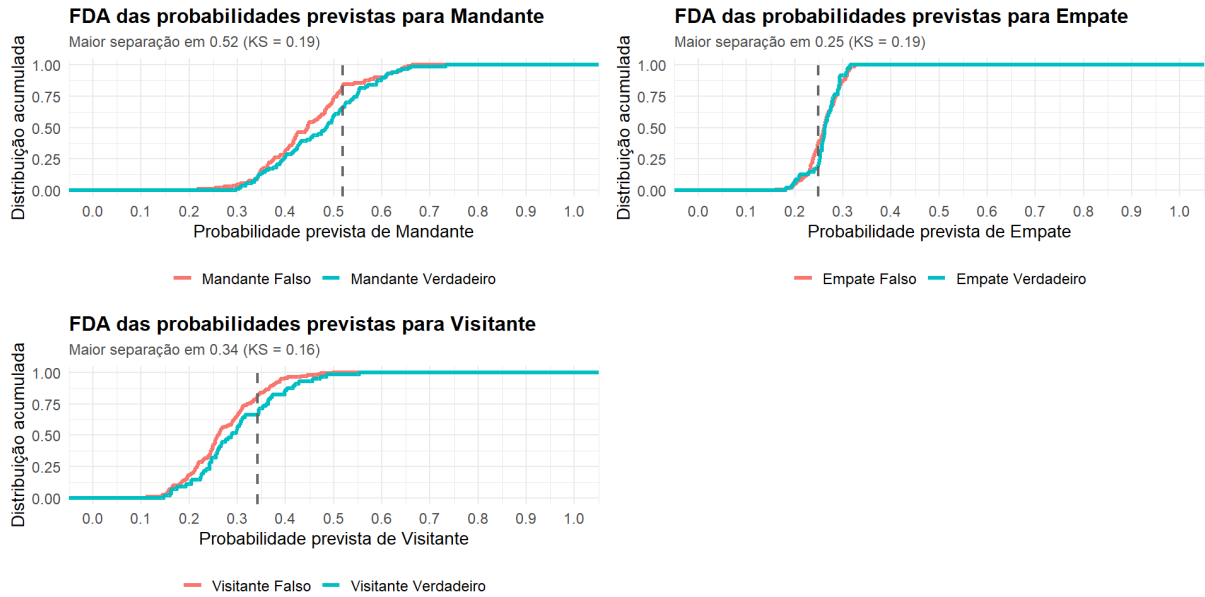


Figura 4.11: FDA das probabilidades previstas pelo modelo Hierárquico.

concordância esperada ao acaso, variando de  $-1$  (discordância total) a  $1$  (concordância perfeita). O  $MCC$  também varia de  $-1$  a  $1$  e é especialmente indicado em cenários com classes desbalanceadas, pois leva em conta verdadeiros positivos (TP), verdadeiros negativos (TN), falsos positivos (FP) e falsos negativos (FN):

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad \text{e} \quad MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \quad (4.1)$$

Para o modelo Hierárquico obtivemos

$$\kappa = 0,0603, \quad MCC = 0,0793, \quad (4.2)$$

enquanto o modelo Independente apresentou

$$\kappa = 0,0541, \quad MCC = 0,0606. \quad (4.3)$$

Os valores, embora baixos, revelam que o modelo Hierárquico alcança uma concordância ligeiramente superior à do modelo Independente, reforçando sua capacidade de alinhamento entre previsões e resultados observados.“

## 4.7 Diagrama de Calibração

Para verificar se as probabilidades previstas podem ser interpretadas literalmente, ou seja, se uma chance de 40% corresponde a uma frequência de ocorrência de aproximadamente 40%, construímos os diagramas de calibração mostrados nas Figuras 4.12 e 4.13. A reta tracejada  $y = x$  representa calibração perfeita, enquanto a curva azul, ajustada por LOESS, revela os desvios sistemáticos do modelo.

Na Figura 4.12, observamos que o modelo Independente tende a subestimar levemente as probabilidades mais baixas (a curva azul fica abaixo de  $y = x$  para  $p \lesssim 0,3$ ), atinge boa aderência no intervalo intermediário ( $0,3 \lesssim p \lesssim 0,6$ ) e volta a subestimar em probabilidades altas ( $p \gtrsim 0,7$ ), embora dentro das bandas de confiança. Já na Figura 4.13 o modelo Hierárquico apresenta comportamento similar, mas com desvios menores na região intermediária, indicando calibração ligeiramente superior entre 0,4 e 0,6. Em ambos os casos, as maiores incertezas (bandas mais largas) ocorrem nos extremos, onde há menos observações e a regressão local torna-se menos precisa.

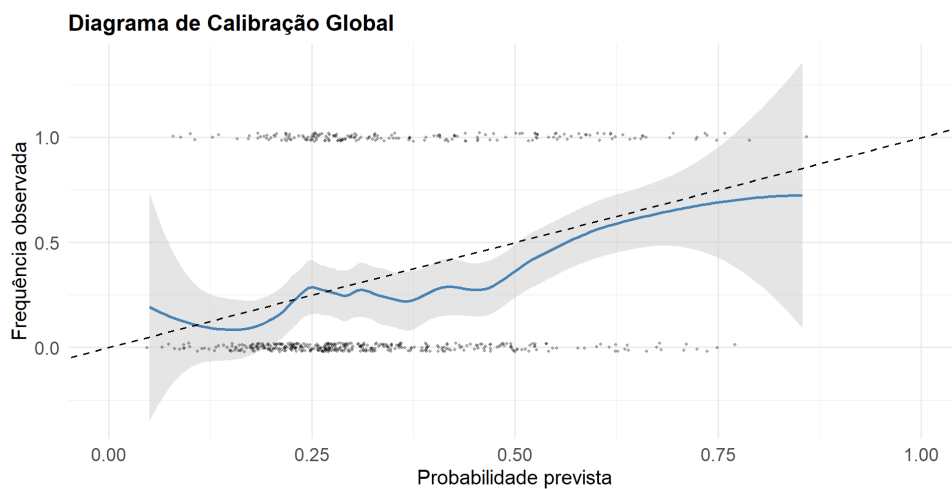


Figura 4.12: Diagrama de calibração do modelo Independente.

## 4.8 Probabilidades para a Classificação Final

Para avaliar a capacidade dos modelos em capturar a dinâmica da reta final do campeonato, conduzimos uma análise de cenários. Definimos três momentos distintos para ajustar os modelos e simular o desfecho da competição:

- **Cenário 1:** Modelos ajustados com dados até a 23<sup>a</sup> rodada (faltando 15 rodadas).

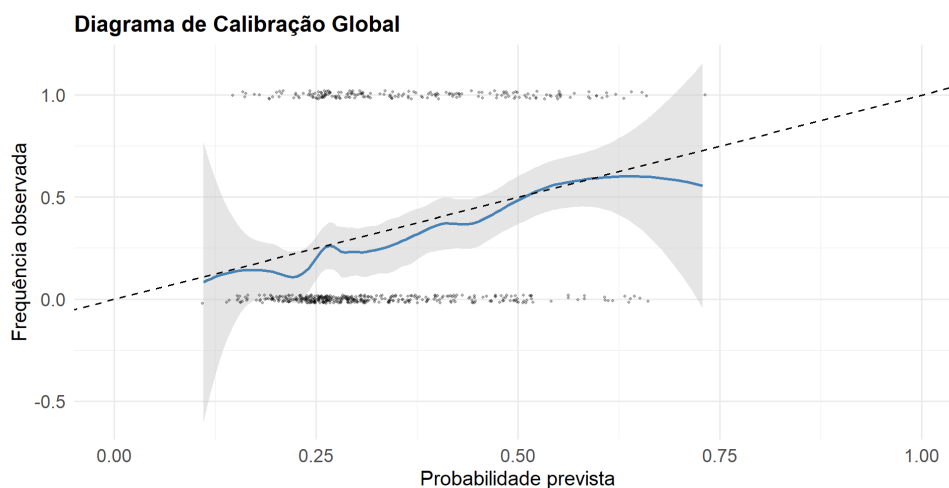


Figura 4.13: Diagrama de calibração do modelo Hierárquico.

- **Cenário 2:** Modelos ajustados com dados até a 28<sup>a</sup> rodada (faltando 10 rodadas).
- **Cenário 3:** Modelos ajustados com dados até a 33<sup>a</sup> rodada (faltando 5 rodadas).

Para cada cenário e cada modelo, simulamos 1000 finalizações completas do campeonato. Em cada simulação, os resultados para as rodadas restantes foram previstos, os pontos foram atribuídos (3 para vitória, 1 para empate e 0 para derrota) e a tabela de classificação final foi gerada. A repetição desse processo permitiu calcular a probabilidade de cada time terminar o campeonato em posições-chave.

Nesta análise, focamos em ser campeão, terminar no G4 e terminar no Z4. O termo G4 refere-se aos quatro primeiros colocados, que historicamente garantem vaga direta na fase de grupos da Copa Libertadores, a principal competição de clubes da América do Sul. O Z4, por sua vez, corresponde à zona de rebaixamento, que engloba as quatro últimas equipes da tabela, rebaixadas para a Série B.

#### 4.8.1 Análise das Probabilidades Simuladas

As Tabelas 4.10 a 4.15 resumam as probabilidades calculadas para cada cenário. A análise revela a natureza volátil do campeonato e a capacidade dos modelos de adaptar suas previsões à medida que novas informações se tornam disponíveis.

A disputa pelo título, no Cenário 1, com 15 rodadas restantes, ambos os modelos apontavam o Grêmio como o favorito ao título (com 48% de chance no modelo Independente e 43% no Hierárquico), seguido por Palmeiras e Atlético-MG. No entanto, no Cenário 2, uma mudança drástica ocorre: as chances do Grêmio despencam, e Flamengo

e Botafogo-RJ emergem como os principais candidatos, refletindo seus bons desempenhos naquele período do campeonato. Na reta final (Cenário 3), uma nova reviravolta: o Palmeiras assume o favoritismo em ambos os modelos (com 28-30% de chance), seguido de perto por Grêmio e Atlético-MG, o que se alinhou com o desfecho real do campeonato.

A disputa pelo G4 mostrou-se acirrada, com um grande grupo de times, incluindo Flamengo, Botafogo-RJ e Bragantino, mantendo altas chances em todos os cenários. É notável como a probabilidade do Bragantino de terminar no G4 cresce consistentemente, passando de 30% para 68% em ambos os modelos, que capturaram bem a arrancada da equipe.

Na parte de baixo da tabela, a análise é ainda mais conclusiva. Ambos os modelos identificaram corretamente, desde o primeiro cenário, o altíssimo risco de rebaixamento de América-MG e Coritiba, cujas probabilidades de queda se aproximam de 100% no cenário final. A luta para evitar as duas vagas restantes no Z4 foi bem capturada, com as probabilidades de equipes como Santos, Goiás, Bahia e Vasco flutuando a cada cenário, demonstrando a sensibilidade dos modelos aos resultados recentes. O modelo previu corretamente o alto risco do Santos e a situação crítica do Goiás, que de fato foram rebaixados.

De modo geral, os dois modelos contaram uma história similar sobre a evolução do campeonato. A principal diferença, consistente com as análises anteriores, é que o modelo Hierárquico tendeu a apresentar probabilidades um pouco menos extremas, um reflexo do seu efeito de regularização. Por exemplo, no favoritismo inicial do Grêmio, o modelo Independente atribuiu 47,6% de chance, enquanto o Hierárquico atribuiu 42,8%. Essa suavização, no entanto, não impediu o modelo de capturar as grandes tendências e reviravoltas da competição, mostrando que ambas as abordagens foram eficazes em traduzir o desempenho em campo em probabilidades para os desfechos da temporada.

Tabela 4.10: Probabilidade de ser campeão nos cenários 1, 2 e 3 (Modelo Independente)

<b>Time</b>	<b>1</b>	<b>2</b>	<b>3</b>
Palmeiras	<b>0,271</b>	0,006	<b>0,284</b>
Gremio	<b>0,476</b>	0,093	<b>0,226</b>
Athletico-MG	<b>0,147</b>	<b>0,184</b>	<b>0,201</b>
Flamengo	0,060	<b>0,233</b>	0,176
Botafogo-RJ	0,024	<b>0,207</b>	0,093
Bragantino	0,012	0,133	0,021
Fluminense	0,005	0,070	0,000
Athletico-PR	0,002	0,042	0,000
Internacional	0,001	0,013	0,000
Fortaleza	0,001	0,010	0,000
Sao Paulo	0,001	0,005	0,000
Cuiaba	0,000	0,000	0,000
Corinthians	0,000	0,001	0,000
Cruzeiro	0,000	0,001	0,000
Vasco	0,000	0,000	0,000
Bahia	0,000	0,000	0,000
Santos	0,000	0,000	0,000
Goias	0,000	0,000	0,000
Coritiba	0,000	0,000	0,000
America-MG	0,000	0,000	0,000

Tabela 4.11: Probabilidade de ser campeão nos cenários 1, 2 e 3 (Modelo Hierárquico)

<b>Time</b>	<b>1</b>	<b>2</b>	<b>3</b>
Palmeiras	<b>0,255</b>	0,010	<b>0,297</b>
Gremio	<b>0,428</b>	0,121	<b>0,230</b>
Athletico-MG	<b>0,172</b>	<b>0,206</b>	<b>0,209</b>
Flamengo	0,073	<b>0,248</b>	0,155
Botafogo-RJ	0,041	<b>0,176</b>	0,081
Bragantino	0,022	0,114	0,027
Fluminense	0,005	0,071	0,001
Athletico-PR	0,003	0,033	0,000
Internacional	0,000	0,011	0,000
Fortaleza	0,000	0,005	0,000
Sao Paulo	0,001	0,003	0,000
Cuiaba	0,000	0,001	0,000
Corinthians	0,000	0,000	0,000
Cruzeiro	0,000	0,001	0,000
Vasco	0,000	0,000	0,000
Bahia	0,000	0,000	0,000
Santos	0,000	0,000	0,000
Goias	0,000	0,000	0,000
Coritiba	0,000	0,000	0,000
America-MG	0,000	0,000	0,000

Tabela 4.12: Probabilidade de ficar no G4 nos cenários 1, 2 e 3 (Modelo Independente)

<b>Time</b>	<b>1</b>	<b>2</b>	<b>3</b>
Palmeiras	0,033	0,055	<b>0,477</b>
Gremio	0,179	<b>0,399</b>	<b>0,479</b>
Athletico-MG	<b>0,505</b>	0,352	0,446
Flamengo	<b>0,455</b>	0,316	0,442
Botafogo-RJ	<b>0,388</b>	0,341	0,437
Bragantino	0,301	<b>0,390</b>	<b>0,684</b>
Fluminense	0,267	<b>0,366</b>	0,031
Athletico-PR	0,206	0,301	0,003
Internacional	0,180	0,231	0,001
Fortaleza	0,137	0,127	0,000
Sao Paulo	0,111	0,064	0,000
Cuiaba	0,090	0,035	0,000
Corinthians	0,052	0,013	0,000
Cruzeiro	0,035	0,005	0,000
Vasco	0,032	0,003	0,000
Bahia	0,016	0,000	0,000
Santos	0,005	0,001	0,000
Goiias	0,004	0,000	0,000
Coritiba	0,002	0,000	0,000
America-MG	0,001	0,000	0,000

Tabela 4.13: Probabilidade de ficar no G4 nos cenários 1, 2 e 3 (Modelo Hierárquico)

<b>Time</b>	<b>1</b>	<b>2</b>	<b>3</b>
Palmeiras	0,035	0,052	0,422
Gremio	0,206	<b>0,390</b>	0,432
Athletico-MG	<b>0,431</b>	0,345	0,450
Flamengo	<b>0,418</b>	0,341	<b>0,465</b>
Botafogo-RJ	<b>0,388</b>	<b>0,368</b>	<b>0,476</b>
Bragantino	0,316	<b>0,389</b>	<b>0,686</b>
Fluminense	0,292	0,357	0,056
Athletico-PR	0,229	0,311	0,012
Internacional	0,190	0,203	0,001
Fortaleza	0,162	0,123	0,000
Sao Paulo	0,115	0,066	0,000
Cuiaba	0,076	0,029	0,000
Corinthians	0,048	0,016	0,000
Cruzeiro	0,035	0,005	0,000
Vasco	0,022	0,003	0,000
Bahia	0,022	0,001	0,000
Santos	0,007	0,000	0,000
Goiias	0,005	0,000	0,000
Coritiba	0,001	0,000	0,000
America-MG	0,000	0,000	0,000

Tabela 4.14: Probabilidade de ficar no Z4 nos cenários 1, 2 e 3 (Modelo Independente)

<b>Time</b>	<b>1</b>	<b>2</b>	<b>3</b>
Palmeiras	0,000	0,000	0,000
Gremio	0,000	0,000	0,000
Athletico-MG	0,001	0,000	0,000
Flamengo	0,003	0,000	0,000
Botafogo-RJ	0,009	0,001	0,000
Bragantino	0,017	0,001	0,000
Fluminense	0,023	0,009	0,038
Athletico-PR	0,050	0,018	0,050
Internacional	0,071	0,031	0,095
Fortaleza	0,107	0,065	0,131
Sao Paulo	0,140	0,101	0,136
Cuiaba	0,154	0,166	0,134
Corinthians	0,198	0,173	0,127
Cruzeiro	0,235	0,209	0,122
Vasco	0,272	0,232	0,106
Bahia	0,316	0,276	0,114
Santos	0,393	0,333	0,203
<b>Goias</b>	<b>0,474</b>	<b>0,461</b>	<b>0,746</b>
<b>Coritiba</b>	<b>0,665</b>	<b>0,929</b>	<b>1,000</b>
<b>America-MG</b>	<b>0,872</b>	<b>0,997</b>	<b>1,000</b>

Tabela 4.15: Probabilidade de ficar no Z4 nos cenários 1, 2 e 3 (Modelo Hierárquico)

<b>Time</b>	<b>1</b>	<b>2</b>	<b>3</b>
Palmeiras	0,000	0,000	0,000
Gremio	0,000	0,000	0,000
Athletico-MG	0,002	0,000	0,000
Flamengo	0,003	0,000	0,000
Botafogo-RJ	0,007	0,001	0,000
Bragantino	0,013	0,002	0,000
Fluminense	0,020	0,004	0,015
Athletico-PR	0,037	0,014	0,031
Internacional	0,058	0,029	0,064
Fortaleza	0,086	0,053	0,091
Sao Paulo	0,094	0,090	0,115
Cuiaba	0,142	0,141	0,115
Corinthians	0,178	0,178	0,113
Cruzeiro	0,219	0,210	0,161
Vasco	0,275	0,253	0,195
Bahia	0,296	0,284	0,170
Santos	0,392	0,355	0,250
<b>Goias</b>	<b>0,528</b>	<b>0,458</b>	<b>0,681</b>
<b>Coritiba</b>	<b>0,733</b>	<b>0,935</b>	<b>1,000</b>
<b>America-MG</b>	<b>0,918</b>	<b>0,994</b>	<b>1,000</b>



# Capítulo 5

## Conclusão

Neste trabalho, desenvolvemos e comparamos duas abordagens bayesianas, uma hierárquica e uma independente, para a previsão de resultados no Campeonato Brasileiro. A análise dos resultados revelou um claro *trade-off* entre as duas arquiteturas. Verificamos que o modelo Poisson Independente se mostrou mais sensível às particularidades de cada equipe, evitando a suavização excessiva e apresentando um melhor ajuste geral aos dados observados. Contudo, essa sensibilidade veio ao custo de uma maior instabilidade em suas previsões probabilísticas, conforme demonstramos pelas regras de escore. Em contrapartida, o modelo Hierárquico, por meio do seu efeito de regularização, produziu previsões mais robustas e consistentes, embora tenha subestimado o desempenho de equipes nos extremos da tabela.

Apesar de capturarem as tendências gerais da competição, observamos que ambos os modelos apresentaram limitações inerentes à complexidade do futebol. Uma das principais dificuldades observadas foi a incapacidade de prever empates, um desfecho comum no esporte, mas probabilisticamente complexo. Além disso, a suposição da distribuição de Poisson, embora um ponto de partida padrão, pode não ser suficiente para capturar a superdispersão frequentemente encontrada na contagem de gols.

Diante desses achados, algumas linhas de pesquisa futuras se mostram promissoras para aprimorar a capacidade preditiva. Para lidar com as limitações da distribuição, seria interessante explorar alternativas como a Binomial Negativa, que acomoda a superdispersão. Modelos mais complexos, como o Poisson Bivariado, poderiam capturar a correlação entre os gols do time da casa e do visitante, enquanto modelos com inflação de zero ou inflação na diagonal poderiam tratar diretamente da alta frequência de placares de 0 a 0 e outros empates, respectivamente. Outro refinamento seria modelar o efeito

de mando de campo individualmente para cada equipe, reconhecendo que a vantagem de jogar em casa não é uniforme entre os clubes.

Uma segunda frente de pesquisa, de grande potencial, envolve a incorporação de informações externas, especificamente do mercado de apostas. As probabilidades implícitas nas odds poderiam ser usadas para definir prioris informativas mais realistas para os parâmetros do modelo, potencialmente melhorando sua calibração. As probabilidades do mercado poderiam também servir como um modelo de benchmark, permitindo comparar o desempenho do modelo estatístico para verificar se ele oferece algum valor preditivo adicional. Por fim, uma aplicação prática seria usar o modelo para identificar apostas de valor, onde as previsões divergem das do mercado, e simular o retorno financeiro de uma estratégia baseada nessas previsões.

Por fim, neste estudo, demonstramos a utilidade e a flexibilidade da modelagem bayesiana na análise de dados esportivos. Embora a previsão de resultados no futebol continue sendo um desafio notável, os modelos que desenvolvemos foram capazes de quantificar as forças das equipes, capturar a dinâmica da competição e fornecer uma base sólida sobre a qual futuras investigações podem ser construídas. A combinação de estruturas hierárquicas, distribuições mais adequadas e a incorporação de fontes de dados externas, como as do mercado de apostas, representa um caminho fértil para o avanço da análise preditiva no esporte.

# Referências Bibliográficas

- Baio, G. e Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, **37**(2), 253–264.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, **78**(1), 1–3.
- Cleveland, W. S. e Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, **83**(403), 596–610.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, **20**(1), 37–46.
- Conover, W. J. (1999). *Practical nonparametric statistics*. John Wiley & Sons, New York, third edition.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. e Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, third edition.
- Geman, S. e Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6), 721–741.
- Gilks, W. R., Richardson, S. e Spiegelhalter, D., editors (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.
- Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society, Series B (Methodological)*, **14**(1), 107–114.
- Gorodkin, J. (2004). Comparing two k-category assignments by a k-category correlation coefficient. *Computational Biology and Chemistry*, **28**(5-6), 367–374.

- Haario, H., Saksman, E. e Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli*, **7**(2), 223–242.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, **57**(1), 97–109.
- Karlis, D. e 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.
- Lunn, D., Spiegelhalter, D., Thomas, A. e Best, N. (2009). The bugs project: Evolution, critique and future directions. *Statistics in Medicine*, **28**(25), 3049–3067.
- Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, **405**(2), 442–451.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. e Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, **21**(6), 1087–1092.
- Rue, H. e Salvesen, (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **49**(3), 399–418.
- Santos-Fernandez, E. (2019). Bayesian statistics meets sports - a review. *REVSTAT-Statistical Journal*, **17**(4), 443–467.
- Staël von Holstein, C.-A. S. (1970). *Assessment and evaluation of subjective probability distributions*. The Economic Research Institute at the Stockholm School of Economics, Stockholm.

# Apêndice A

## Cadeias de Markov e MCMC

### A.1 Traceplots das cadeias de Markov

Nesta seção, apresentamos os *traceplots* de alguns parâmetros selecionados para avaliar a convergência das cadeias MCMC. Devido ao grande número de parâmetros no modelo, optamos por exibir apenas um subconjunto representativo, incluindo parâmetros de ataque e defesa e o efeito de mando de campo. Essa seleção é suficiente para ilustrar que as cadeias oscilaram em torno de um estado estacionário sem tendências evidentes.

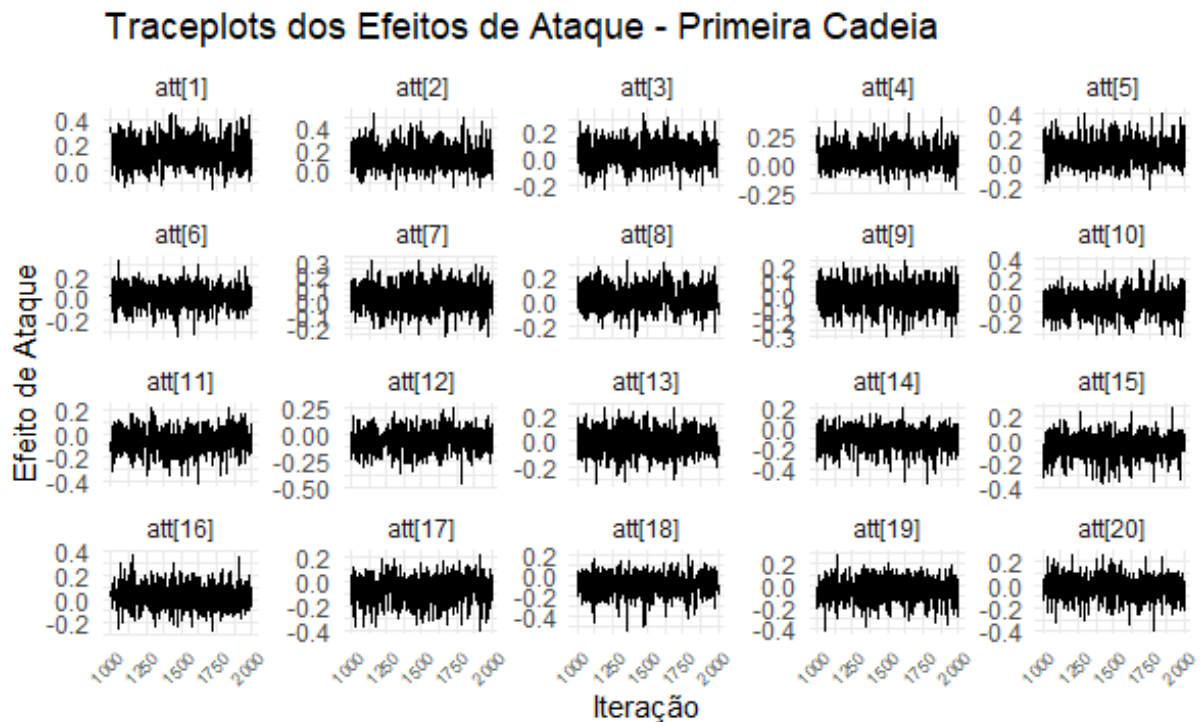


Figura A.1: Traceplots das Cadeias MCMC para os Efeitos de Ataque (*att*).

Os resultados indicam que as cadeias atingiram a convergência, conforme evidenciado

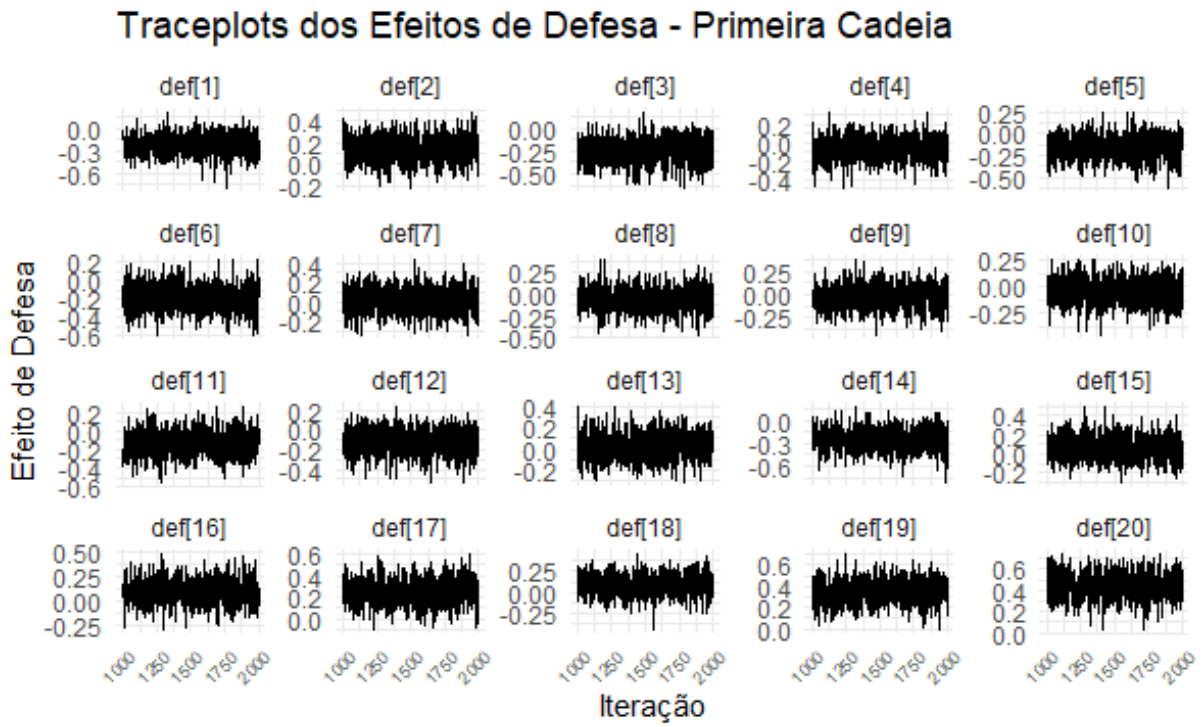


Figura A.2: Traceplots das Cadeias MCMC para os Efeitos de Defesa (*def*).

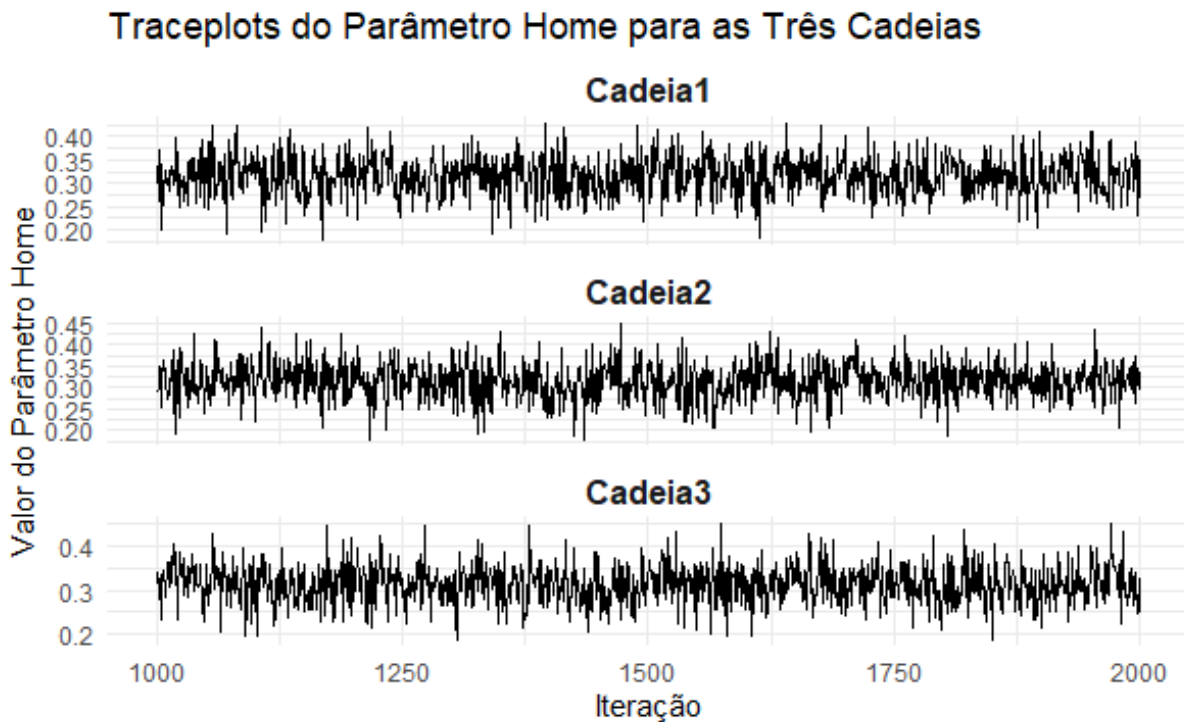


Figura A.3: Traceplots das Cadeias MCMC para o Efeito do Mando de Campo (*home*).

pela ausência de padrões sistemáticos nos gráficos e pela estabilidade das médias ao longo das iterações.



# Apêndice B

## Modelo

### 1. Modelo Poisson Hierárquica

```
modelo <- function(){
  # Verossimilhança e modelo de efeitos aleatórios
  for (g in 1:n games) {
    # Número observado de gols marcados por cada equipe no jogo g
    y1[g] ~ dpois(theta[g,1])
    y2[g] ~ dpois(theta[g,2])

    # Distribuição preditiva para o número de gols marcados no jogo g
    y.pred[g,1] ~ dpois(theta[g,1])
    y.pred[g,2] ~ dpois(theta[g,2])

    # Intensidades médias de marcação de gols
    log(theta[g,1]) <- home + att[h[g]] + def[a[g]]
    log(theta[g,2]) <- att[a[g]] + def[h[g]]
  }

  # Priori para o efeito de mando de campo
  home ~ dnorm(0, 0.0001)

  # Código para impor a restrição de soma-zero
  for (t in 1:nteam){
    att.star[t] ~ dnorm(mu.att, tau.att)
```

```

def.star[t] ~ dnorm(mu.def, tau.def)
att[t] <- att.star[t] - mean(att.star[])
def[t] <- def.star[t] - mean(def.star[])
}

# Prioris para os efeitos aleatórios (hiperparâmetros)
mu.att ~ dnorm(0, 0.0001)
mu.def ~ dnorm(0, 0.0001)
tau.att ~ dgamma(0.01, 0.01)
tau.def ~ dgamma(0.01, 0.01)
}

```

## 2. Modelo Poisson Independente

```

modelo_indep <- function(){
  for (g in 1:ngames) {
    # Número observado de gols marcados por cada equipe no jogo g
    y1[g] ~ dpois(theta[g,1])
    y2[g] ~ dpois(theta[g,2])

    # Distribuição preditiva para o número de gols marcados no jogo g
    y.pred[g,1] ~ dpois(theta[g,1])
    y.pred[g,2] ~ dpois(theta[g,2])

    # Intensidades médias de marcação de gols
    log(theta[g,1]) <- home + att[h[g]] + def[a[g]]
    log(theta[g,2]) <- att[a[g]] + def[h[g]]
  }

  # Priori para o efeito de mando de campo
  home ~ dnorm(0, 0.0001)

  # Priori para efeitos att e def
  for (t in 1:nteam) {

```

```
att.star[t] ~ dnorm(0, 0.0001)
def.star[t] ~ dnorm(0, 0.0001)
}

# Código para impor a restrição de soma-zero
mean_att <- mean(att.star[])
mean_def <- mean(def.star[])

for (t in 1:nteams) {
  att[t] <- att.star[t] - mean_att
  def[t] <- def.star[t] - mean_def
}
}
```



# Apêndice C

## Detalhamento da Distribuição a Posteriori

$$\begin{aligned}
 p(\theta | D) \propto & \underbrace{\prod_{g=1}^{380} \left[ \frac{\exp(-e^{\lambda_{g1}})(e^{\lambda_{g1}})^{y_{g1}}}{y_{g1}!} \times \frac{\exp(-e^{\lambda_{g2}})(e^{\lambda_{g2}})^{y_{g2}}}{y_{g2}!} \right]}_{\text{Verossimilhança (Poisson)}} \\
 & \times \underbrace{\left[ \prod_{t=1}^{20} \sqrt{\frac{\tau_{\text{att}}}{2\pi}} \exp\left(-\frac{\tau_{\text{att}}}{2}(\text{att}_t - \mu_{\text{att}})^2\right) \right]}_{\text{Prioris para Ataque}} \\
 & \times \underbrace{\left[ \prod_{t=1}^{20} \sqrt{\frac{\tau_{\text{def}}}{2\pi}} \exp\left(-\frac{\tau_{\text{def}}}{2}(\text{def}_t - \mu_{\text{def}})^2\right) \right]}_{\text{Prioris para Defesa}} \\
 & \times \underbrace{\sqrt{\frac{0,0001}{2\pi}} \exp\left(-\frac{0,0001}{2}\text{home}^2\right)}_{\text{Prior para Mando de Campo}} \\
 & \times \underbrace{\sqrt{\frac{0,0001}{2\pi}} \exp\left(-\frac{0,0001}{2}\mu_{\text{att}}^2\right) \times \frac{0,1^{0,1}}{\Gamma(0,1)} \tau_{\text{att}}^{0,1-1} e^{-0,1\tau_{\text{att}}}}_{\text{Hiperprioris de Ataque}} \\
 & \times \underbrace{\sqrt{\frac{0,0001}{2\pi}} \exp\left(-\frac{0,0001}{2}\mu_{\text{def}}^2\right) \times \frac{0,1^{0,1}}{\Gamma(0,1)} \tau_{\text{def}}^{0,1-1} e^{-0,1\tau_{\text{def}}}}_{\text{Hiperprioris de Defesa}}
 \end{aligned} \tag{C.1}$$

A verossimilhança marginal dos dados,  $p(D)$ , é a constante de normalização do Teorema de Bayes. Ela é calculada integrando-se o produto da verossimilhança pela priori

sobre todo o espaço de parâmetros  $\theta$ . Para o nosso modelo, isso se traduz na seguinte integral de 45 dimensões

$$p(D) = \int \cdots \int_{\Theta} \left( \begin{aligned} & \left( \prod_{g=1}^{380} \frac{\exp(-e^{\lambda_{g1}})(e^{\lambda_{g1}})^{y_{g1}} \exp(-e^{\lambda_{g2}})(e^{\lambda_{g2}})^{y_{g2}}}{y_{g1}! y_{g2}!} \right) \\ & \times \left( \prod_{t=1}^{20} \sqrt{\frac{\tau_{\text{att}}}{2\pi}} \exp\left(-\frac{\tau_{\text{att}}}{2}(\text{att}_t - \mu_{\text{att}})^2\right) \right) \\ & \times \left( \prod_{t=1}^{20} \sqrt{\frac{\tau_{\text{def}}}{2\pi}} \exp\left(-\frac{\tau_{\text{def}}}{2}(\text{def}_t - \mu_{\text{def}})^2\right) \right) \\ & \times \sqrt{\frac{0,0001}{2\pi}} \exp\left(-\frac{0,0001}{2}\text{home}^2\right) \\ & \times \sqrt{\frac{0,0001}{2\pi}} \exp\left(-\frac{0,0001}{2}\mu_{\text{att}}^2\right) \times \frac{0,1^{0,1}}{\Gamma(0,1)} \tau_{\text{att}}^{-0,9} e^{-0,1\tau_{\text{att}}} \\ & \times \sqrt{\frac{0,0001}{2\pi}} \exp\left(-\frac{0,0001}{2}\mu_{\text{def}}^2\right) \times \frac{0,1^{0,1}}{\Gamma(0,1)} \tau_{\text{def}}^{-0,9} e^{-0,1\tau_{\text{def}}} \end{aligned} \right) d\theta \quad (\text{C.2})$$

onde o termo diferencial  $d\theta$  representa a integração sobre todos os 45 parâmetros estocásticos do modelo

$$d\theta = d(\text{home}) \left( \prod_{t=1}^{20} d(\text{att}_t) \right) \left( \prod_{t=1}^{20} d(\text{def}_t) \right) d(\mu_{\text{att}}) d(\tau_{\text{att}}) d(\mu_{\text{def}}) d(\tau_{\text{def}}).$$