

**Universidade Federal de Alfenas**

**Nikolas Neves de Figueiredo**

**Modelagem Estatística e o Limite da Predição no Futebol Brasileiro: Um Estudo Comparativo de Desempenho e Parcimônia**

**Alfenas/MG**

**2025**

**Nikolas Neves de Figueiredo**

**Modelagem Estatística e o Limite da Predição no Futebol Brasileiro: Um Estudo Comparativo de Desempenho e Parcimônia**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística Aplicada e Biometria da Universidade Federal de Alfenas - MG, como parte dos requisitos para a obtenção do título de Mestre. Linha de Pesquisa: Modelagem Estatística e Estatística Computacional.  
Orientador: Prof. Dr. Eric Batista Ferreira.

**Alfenas/MG**

**2025**

Sistema de Bibliotecas da Universidade Federal de Alfenas  
Biblioteca Central

Figueiredo, Nikolas Neves de .

Modelagem estatística e o limite da predição no futebol brasileiro : um estudo comparativo de desempenho e parcimônia / Nikolas Neves de Figueiredo. - Alfenas, MG, 2025.

109 f. : il. -

Orientador(a): Eric Batista Ferreira.

Dissertação (Mestrado em Estatística Aplicada e Biometria) -  
Universidade Federal de Alfenas, Alfenas, MG, 2025.

Bibliografia.

1. Previsão esportiva. 2. Aposta esportiva. 3. Modelagem estatística. I. Ferreira, Eric Batista, orient. II. Título.

**Modelagem Estatística e o Limite da Predição no Futebol Brasileiro: Um Estudo Comparativo de Desempenho e Parcimônia**

A Banca examinadora abaixo-assinada aprova a Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre em Estatística Aplicada e Biometria pela Universidade Federal de Alfenas. Área de concentração: Estatística Aplicada e Biometria.

Aprovada em: 22 de dezembro de 2025.

Prof. Dr. Eric Batista Ferreira

Instituição: Universidade Federal de Alfenas - UNIFAL-MG

Profa. Dra. Josiane Magalhães Teixeira

Instituição: Universidade Federal dos Vales do Jequitinhonha e Mucuri - UFVJM - Campus JK

Prof. Dr. Julio Sílvio de Sousa Bueno Filho

Instituição: Universidade Federal de Lavras - UFLA



Documento assinado eletronicamente por **Eric Batista Ferreira, Presidente**, em 22/12/2025, às 11:04, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://sei.unifal-mg.edu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **1697449** e o código CRC **D85B513C**.

---

Dedico este trabalho à Clara, que me fortaleceu  
e me permitiu trilhar este caminho.

## AGRADECIMENTOS

Gostaria de agradecer a todas as pessoas que colaboraram e me apoiaram na realização deste trabalho. Sem o apoio dessas pessoas, eu não teria alcançado êxito em sua conclusão. Agradeço à minha família, especialmente aos meus pais, por todo o suporte e dedicação que sempre tiveram com a minha educação. Agradeço também à minha esposa, que compartilha comigo o mérito deste trabalho e tudo o que ele representa. Não poderia deixar de agradecer à minha filha, por ser o bebê mais tranquilo deste mundo e por ter me permitido trabalhar nesta pesquisa.

Agradeço imensamente ao meu querido amigo e orientador, Professor Eric, que confiou em mim, me ensinou e sempre me apoiou desde nossas primeiras pesquisas, iniciadas em 2020. Agradeço, de forma especial, aos professores Fabrício, Denismar, Gilberto e Luiz Alberto, do programa, por todos os ensinamentos, momentos e experiências compartilhadas, bem como ao professor visitante Cléber, que trouxe excelente contribuição acadêmica durante sua passagem pelo programa. Ainda dentro dos agradecimentos institucionais, registro minha gratidão à secretária Martha, que, com muita paciência e dedicação, sempre me auxiliou e foi responsável por me orientar no processo de ingresso no programa.

Agradeço também ao meu amigo e parceiro de pesquisa Giovani. Apesar da distância, desejo que em breve possamos continuar nossas discussões e trabalhos sobre futebol. Estendo meus agradecimentos ao meu grande amigo Rafael, que, desde a graduação, tem sido um porto seguro dentro da universidade.

Por fim, agradeço ao Programa de Pós-Graduação em Estatística Aplicada e Biometria e à sua equipe gestora, por permanecerem ativos e por me oferecerem esta oportunidade. Agradeço à UNIFAL-MG por proporcionar ensino de qualidade e gratuito, bem como às instituições de fomento à pesquisa que financiaram meus projetos desde 2019 e contribuíram imensamente para minha formação. Meu agradecimento à CAPES, ao CNPq, à FAPEMIG e à UNIFAL-MG.

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brasil, Código de Financiamento 001.

“Ganhar é bom, mas ganhar da Argentina é muito melhor.”

(Galvão Bueno)

## RESUMO

O presente trabalho investiga a aplicação de modelos estatísticos para a previsão de resultados de partidas do Campeonato Brasileiro de Futebol Série A, com ênfase na comparação de desempenho entre diferentes abordagens metodológicas. A motivação do estudo decorre do histórico vínculo entre probabilidade e jogos, bem como do crescente interesse acadêmico e comercial impulsionado pela popularização das apostas esportivas, especialmente por meio de plataformas digitais. Embora a literatura apresente diversas propostas de modelagem para o futebol, observa-se que, na maioria dos casos, as taxas de acerto raramente ultrapassam 60% e, quando o fazem, tendem a perder consistência em aplicações de longo prazo. Foram analisados 35 modelos distintos, originados a partir de dois métodos amplamente reconhecidos na literatura e de variações metodológicas propostas neste trabalho. Os modelos foram avaliados com base em onze métricas, buscando captar as particularidades de cada abordagem. Essas métricas mensuram acurácia, erro, calibração probabilística, nível de complexidade, além de critérios adicionais destinados a aferir robustez e aplicabilidade prática. Uma atenção especial foi dada à parcimônia dos modelos e ao equilíbrio entre simplicidade e desempenho. O estudo também incorporou a análise do impacto de diferentes janelas temporais, definidas em função do número de rodadas anteriores consideradas na geração das previsões. Dessa forma, o desempenho dos modelos variou de acordo com as métricas empregadas, mas alguns se destacaram de forma clara. Entre eles, os modelos *UfmgGer* e *SdlGer* apresentaram os melhores resultados globais, enquanto o modelo *ConTeiMan* surpreendeu por sua metodologia extremamente simples aliada a uma elevada taxa de acerto.

**Palavras-chave:** Previsão esportiva, Aposta esportiva, Modelagem estatística.

## ABSTRACT

The present study investigates the application of statistical models for predicting the outcomes of matches in the Brazilian Série A Football Championship, with an emphasis on comparing performance across different methodological approaches. The motivation for this research stems from the historical link between probability theory and games, as well as from the growing academic and commercial interest driven by the popularization of sports betting, particularly through digital platforms. Although the literature presents a wide range of modeling proposals for football, it is observed that, in most cases, prediction accuracy rates rarely exceed 60%, and when they do, they tend to lose consistency in long-term applications. A total of 35 distinct models were analyzed, derived from two methods widely recognized in the literature and from methodological variations proposed in this study. The models were evaluated based on eleven metrics, aiming to capture the specific characteristics of each approach. These metrics assess accuracy, error, probabilistic calibration, and model complexity, in addition to further criteria designed to evaluate robustness and practical applicability. Special attention was given to model parsimony and to the balance between simplicity and performance. The study also incorporated an analysis of the impact of different temporal windows, defined according to the number of previous rounds considered in generating the predictions. As a result, model performance varied depending on the metrics employed, although some models clearly stood out. Among them, the UfmgGer and Sd1Ger models achieved the best overall results, while the ConTeiMan model was notable for its extremely simple methodology combined with a high prediction accuracy.

**Keywords:** Sports forecasting, Sports betting, Statistical modeling.

## LISTA DE FIGURAS

Figura 1 – Exemplo gráfico de uma distribuição Poisson Univariada com $\lambda = 3$ .	19
Figura 2 – Exemplo gráfico de uma distribuição de Poisson bivariada de $P$ e $Q$ .	20
Figura 3 – Representação gráfica da medida de De Finetti (MD)	47
Figura 4 – Representação gráfica da medida de De Finetti Detalhada (MDD)	49
Figura 5 – Contagem de vitórias por time.	56
Figura 6 – Contagem de derrotas por time.	56
Figura 7 – Contagem de empates por time.	57
Figura 8 – Gráfico composicional das vitórias, derrotas e empates ponderados pelo número de jogos que cada time participou.	59
Figura 9 – Gráfico da frequência de gols observados.	61
Figura 10 – Gráfico da frequência de gols observados comparada à distribuição de Poisson com $\lambda = 1,28$ .	62
Figura 11 – Taxa de acerto dos modelos.	65
Figura 12 – Erro preditivo médio ponderado dos modelos.	66
Figura 13 – Medida de De Finetti dos modelos.	67
Figura 14 – Medida de De Finetti detalhada dos modelos.	69
Figura 15 – Comparativo da taxa de acerto com o nível de complexidade dos modelos.	70
Figura 16 – Valores de verossimilhança do resultado dos modelos.	71
Figura 17 – Precisão dos modelos.	72
Figura 18 – Recall dos modelos.	73
Figura 19 – Valores de F1-Score dos modelos.	74
Figura 20 – Valores do coeficiente de correlação de matthews dos modelos.	75
Figura 21 – Taxa de acerto dos modelos agrupada por janelas de tempo.	76
Figura 22 – Medida de De Finetti dos modelos agrupada por janelas de tempo.	77
Figura 23 – F1-score dos modelos agrupados por janelas de tempo.	78
Figura 24 – Taxa de acerto dos modelos agrupada por família.	80
Figura 25 – Medida de De Finetti dos modelos agrupada por família.	81
Figura 26 – F1-score dos modelos agrupada por família.	82
Figura 27 – Quadro de resultados das métricas.	85

Figura 28 – Resultados comparativos da taxa de acerto para o estudo de simulação nos modelos selecionados. . . . .	87
Figura 29 – Resultados comparativos da métrica MD para o estudo de simulação nos modelos selecionados. . . . .	88
Figura 30 – Resultados comparativos da precisão para o estudo de simulação nos modelos selecionados. . . . .	89
Figura 31 – Resultados comparativos do recall para o estudo de simulação nos modelos selecionados. . . . .	90
Figura 32 – Resultados comparativos do F1-score para o estudo de simulação nos modelos selecionados. . . . .	91
Figura 33 – Resultados comparativos do coeficiente de correlação de Matthews (MCC) para o estudo de simulação nos modelos selecionados. . . . .	92

## LISTA DE TABELAS

Tabela 1 –	Retrospectiva histórica do Campeonato Brasileiro de Futebol. . . . .	17
Tabela 2 –	Quantitativo de títulos de Campeão Brasileiro de Futebol reconhecidos pela CBF. . . . .	18
Tabela 3 –	Resumo dos Modelos de Predição . . . . .	42
Tabela 4 –	Número de jogos, times, rodadas, gols e time campeão de cada ano analisado. . . . .	55
Tabela 5 –	Estatísticas das métricas dos modelos . . . . .	63
Tabela 6 –	Desempenho absoluto dos times no Campeonato Brasileiro . . . . .	100
Tabela 7 –	Desempenho dos times no Campeonato Brasileiro ponderado por jogo. . . . .	101
Tabela 8 –	Resultados Geral das Métricas de avaliação dos modelos . . . . .	102
Tabela 9 –	valores dos critérios de avaliação do Nível de Complexidade por modelo . . . . .	103
Tabela 10 –	Resultados do teste de Kruskal–Wallis para as classes de Memória . . . . .	104
Tabela 11 –	Dunn (Holm) post-hoc — TA — fator Memória . . . . .	104
Tabela 12 –	Dunn (Holm) post-hoc — MD — fator Memória. . . . .	105
Tabela 13 –	Dunn (Holm) post-hoc — MC_F1 — fator Memória. . . . .	105
Tabela 14 –	Resultados do teste de Kruskal–Wallis para as classes de Família . . . . .	106
Tabela 15 –	Dunn (Holm) post-hoc — TA — fator Família . . . . .	107
Tabela 16 –	Dunn (Holm) post-hoc — MD — fator Família . . . . .	108
Tabela 17 –	Dunn (Holm) post-hoc — MC_F1 — fator Família . . . . .	109

## SUMÁRIO

1	<b>INTRODUÇÃO</b>	14
2	<b>OBJETIVO</b>	15
2.1	OBJETIVO GERAL	15
2.2	OBJETIVOS ESPECÍFICOS	15
3	<b>REVISÃO DE LITERATURA</b>	16
3.1	O CAMPEONATO BRASILEIRO DE FUTEBOL	16
3.2	DISTRIBUIÇÕES DE PROBABILIDADE	18
3.2.1	<b>Distribuição de Poisson Univariada</b>	19
3.2.2	<b>Distribuição de Poisson Bivariada</b>	20
3.2.2.1	Método dos Momentos	23
3.3	MODELOS DE PREDIÇÃO DE RESULTADOS DE FUTEBOL	24
4	<b>MATERIAL E MÉTODOS</b>	31
4.1	MATERIAL	31
4.2	MÉTODOS	32
4.2.1	<b>Modelo <math>SD\ 0</math> de Arruda (2000)</b>	32
4.2.2	<b>Modelo <math>SD\ 1</math> de Arruda (2000)</b>	33
4.2.3	<b>Modelo <i>Chance I</i> de Arruda (2000)</b>	34
4.2.4	<b>Modelo <i>Chance II</i> de Arruda (2000)</b>	35
4.2.5	<b>Modelo <i>UFMG</i> de Lima et al. (2012)</b>	36
4.2.6	<b>Modelo <i>UFMG</i> Sem sorteio</b>	38
4.2.7	<b>Modelo <i>UFMG</i> Sorteio Normal</b>	38
4.2.8	<b>Modelo <i>Controle Teimoso</i></b>	39
4.2.9	<b>Modelo <i>Controle Média</i></b>	39
4.2.10	<b>Modelo <i>Controle Uniforme</i></b>	39
4.2.11	<b>Modelo <i>Controle Poisson</i></b>	40
4.2.12	<b>Modelo <i>Controle Implícito Média</i></b>	40
4.3	RESUMO DOS MODELOS	41
4.4	MÉTRICAS DE AVALIAÇÃO	43
4.4.1	<b>Taxa de Acerto</b>	43
4.4.2	<b>Erro Preditivo Médio Ponderado</b>	44
4.4.3	<b>Taxa de Funcionamento</b>	45
4.4.4	<b>Nível de Complexidade</b>	45
4.4.5	<b>Medida de De Finetti</b>	47
4.4.6	<b>Medida de De Finetti Detalhada</b>	48
4.4.7	VEROSSIMILHANÇA DO RESULTADO	49
4.4.8	<b>Matriz de Confusão e Métricas de Desempenho</b>	50
4.4.8.1	Precisão	51
4.4.8.2	Recall ou Sensibilidade	51
4.4.8.3	F1-Score	51
4.4.8.4	Coeficiente de Matthews Generalizado	52
4.5	ESTUDO DE SIMULAÇÃO	52
5	<b>RESULTADOS</b>	54
5.1	ANÁLISE EXPLORATÓRIO DO CAMPEONATO BRASILEIRO	54
5.2	MODELOS E SUAS MÉTRICAS	62

5.2.1	<b>Análise das Métricas Isoladas</b>	64
5.2.2	<b>Comparação dos Resultados Agrupados por Janelas de Tempo</b>	75
5.2.3	<b>Comparação de resultados agrupados por família</b>	79
5.2.4	<b>Tabela Geral de Resultados</b>	84
5.3	<b>RESULTADOS DA SIMULAÇÃO</b>	85
6	<b>CONSIDERAÇÕES FINAIS</b>	93
6.1	<b>TRABALHOS FUTUROS</b>	95
	<b>REFERÊNCIAS</b>	96
	<b>APÊNDICE</b>	100

## 1 INTRODUÇÃO

Desde o surgimento da probabilidade como área formal de estudo, seu desenvolvimento esteve profundamente associado aos jogos de azar e ao desejo humano de antecipar o futuro, estimando quais eventos seriam mais prováveis de ocorrer. Ao longo do tempo, esse campo evoluiu significativamente, incluindo novas metodologias, abordagens matemáticas e aplicações práticas, mas mantendo como propósito central a quantificação da chance de ocorrência de determinados eventos.

Recentemente, esse interesse ganhou novo impulso com a crescente popularização das apostas esportivas, especialmente devido ao avanço das plataformas digitais e à ampla difusão da internet, que ampliaram de maneira expressiva o acesso do público a essas práticas. Esse cenário de expansão comercial reacendeu o interesse acadêmico e profissional pelo estudo das probabilidades aplicadas ao esporte, estimulando o desenvolvimento de modelos estatísticos cada vez mais sofisticados e capazes de estimar resultados com maior precisão.

Entretanto, no contexto do futebol nacional, mais especificamente no que se refere às previsões de resultados de partidas do Campeonato Brasileiro de Futebol Série A, observa-se que os modelos descritos na literatura raramente alcançam taxas de acerto superiores a 60%. Mesmo quando tais índices são obtidos, eles tendem a não se manter consistentes quando aplicados a um grande número de edições ou partidas.

Diante desse cenário, este trabalho propõe um estudo comparativo envolvendo 35 modelos distintos, construídos a partir de variações metodológicas aplicadas a dois modelos amplamente reconhecidos na literatura. Esses modelos podem ser organizados segundo dois critérios principais: (i) a família à qual pertencem, refletindo a metodologia estatística adotada; e (ii) a janela temporal utilizada, que sistematiza a quantidade de partidas consideradas como base para gerar as previsões. Para a comparação entre os modelos, foram utilizadas onze métricas, com o objetivo de avaliar sua acurácia, funcionamento, robustez, parcimônia e precisão.

O desempenho desses modelos foi descrito individualmente em relação a cada métrica, mas também analisado de maneira agrupada por famílias e por janelas temporais, permitindo observar o impacto dessas variações sobre os resultados. Adicionalmente, foi conduzido um estudo complementar que examina o desempenho dos modelos, considerando diferentes janelas de tempo para suas previsões, definidas a partir do número de rodadas utilizadas como referência antes da realização de cada previsão.

## 2 OBJETIVO

### 2.1 OBJETIVO GERAL

Comparar o desempenho dos modelos existentes e propostos na previsão de resultados de partidas do Campeonato Brasileiro de Futebol, com base em diferentes critérios avaliativos.

### 2.2 OBJETIVOS ESPECÍFICOS

- I Estabelecer um banco de dados para a realização das previsões.
- II Realizar uma análise exploratória dos dados, utilizando o banco para obter informações e métricas do Campeonato Brasileiro Série A.
- III Implementar modelos preditivos para resultados de partidas de futebol.
  - a Implementar o modelo da UFMG, conforme proposto por Lima et al. (2012).
  - b Implementar os modelos SD0, SDI, Chance I, Chance II e Implícito I, originalmente propostos por Arruda (2000).
  - c Desenvolver modelos autorais baseados em metodologias intuitivas, com fins comparativos.
- IV Implementar variações metodológicas nos modelos, dando origem a novas versões.
- V Classificar os modelos com base em métricas de avaliação.
- VI Avaliar o impacto das melhorias em relação aos modelos originais.
- VII Attingir uma taxa de acerto superior a 60% na aplicação de pelo menos um modelo.
- VIII Determinar os três modelos com melhor desempenho.
- IX Realizar um estudo de simulação para determinar as melhores janelas de tempo para cada um dos modelos selecionados.

### 3 REVISÃO DE LITERATURA

Nesta seção, os conceitos, métodos e temas que foram utilizados para o desenvolvimento deste trabalho serão apresentados e detalhados.

#### 3.1 O CAMPEONATO BRASILEIRO DE FUTEBOL

O Campeonato Brasileiro de Futebol, também conhecido como Brasileirão, representa a principal competição nacional entre clubes do país. Sua trajetória é marcada por mudanças estruturais significativas, refletindo os desafios de consolidar-se como um torneio de abrangência nacional em um território vasto e futebolisticamente diverso. Além disso, o Brasileirão se destaca por sua expressiva receptividade junto aos torcedores e à mídia. Somando todas as plataformas lineares, incluindo TV aberta e canais esportivos da Globo, mais de 143 milhões de pessoas foram impactadas pelas transmissões do torneio entre abril e dezembro de 2024 (ESPORTIVO, 2023). Nas arenas, o Brasileirão de 2023 registrou média histórica de público pagante superior a 25 mil torcedores por partida, e superior a 26 mil em 2024, culminando em um total de mais de 10 milhões de espectadores ao longo de cada uma das edições (GLOBO, 2023; NASCIMENTO; BRASIL, 2024).

Em 2023, a Confederação Brasileira de Futebol (CBF) considerou o Torneio dos Campeões como o primeiro campeonato de amplitude nacional, sendo, portanto, o Atlético Mineiro considerado o primeiro campeão brasileiro de futebol (RIBEIRO, 2023). Entretanto, apenas em 1959 foi estabelecida uma regularidade para as disputas dessa competição nacional, agora chamada Taça do Brasil (Agência Estado, 2010). Assim, 22 anos depois a CBF voltou a considerar o título de campeão brasileiro, consagrando o Bahia como o segundo a receber esse título (FURTADO, 2010). A partir disso, o campeonato passou a ter diversos nomes e estruturas, como pode ser observado na Tabela 1.

A implementação de um campeonato nacional foi desafiadora devido à influência de competições estaduais e regionais, que já exerciam grande impacto. Por isso, surgiram diversos torneios de âmbito nacional que culminaram no atual campeonato 'Brasileirão'. Essas competições adotaram três principais modalidades: o mata-mata, que consiste em confrontos eliminatórios nos quais os times se enfrentam em partidas de ida e volta (ou em jogo único) e

o perdedor é eliminado; o sistema de pontos corridos, adotado de forma contínua desde 2003, que privilegia a regularidade ao longo de toda a temporada, pois cada equipe enfrenta todas as outras em turno e retorno, e o time com a maior pontuação ao final do campeonato é declarado campeão; e, por fim, o sistema misto, que combina elementos dos dois formatos, geralmente com uma fase classificatória baseada em pontos corridos seguida de confrontos eliminatórios.

Tabela 1 – Retrospectiva histórica do Campeonato Brasileiro de Futebol.

<b>Período</b>	<b>Nome do Campeonato</b>	<b>Número de Times</b>	<b>Tipo de Competição</b>
1937	Torneio dos Campeões	6	Mata-mata
1959–1968	Taça Brasil	16 até 23	Mata-mata
1967–1970	Torneio Roberto Gomes Pedrosa	15-17	Mata-mata
1971–1974	Campeonato Nacional de Clubes	20	Mata-mata
1975–1980	Copa Brasil	32	Mata-mata
1981–1983	Taça de Ouro	16	Sistema misto
1984	Copa Brasil	40	Mata-mata
1985	Taça de Ouro	16	Sistema misto
1986–1988	Copa Brasil ("Copa União")	24	Pontos corridos
1989–1999	Campeonato Brasileiro Série A	20-24	Pontos corridos
2000	Copa João Havelange	116	Mata-mata
2001–presente	Campeonato Brasileiro Série A	20-24	Pontos corridos

Fonte: Adaptada de Correio da Manhã (1937), Furtado (2010), Beting (2016), Hirata (2013) e Nogueira (2017)

Quanto ao número de títulos, destaca-se na Tabela 2 todos os times que já foram considerados campeões brasileiros. O destaque especial vai para o Palmeiras, que ostenta o maior número de títulos do Brasileirão, superando o Santos em 4 unidades.

A tabela revela a hegemonia histórica de alguns clubes. Além do Palmeiras, que lidera o quadro de títulos, clubes como Santos, Corinthians, Flamengo e São Paulo demonstram forte tradição e consistência nas conquistas nacionais, reforçando sua importância no cenário do futebol brasileiro. Enquanto alguns clubes possuem números expressivos, outros, como Guarani, Coritiba, Sport e Athletico-PR, conquistaram o título em ocasiões isoladas. Essa variedade evidencia a competitividade e a imprevisibilidade das disputas nacionais. A distribuição dos títulos ao longo dos anos reflete também a evolução do futebol brasileiro, com diferentes períodos de dominância por parte de certos clubes. Essa alternância não só enriquece a história do campeonato, mas também incentiva uma disputa acirrada a cada nova temporada.

Tabela 2 – Quantitativo de títulos de Campeão Brasileiro de Futebol reconhecidos pela CBF.

<b>Títulos</b>	<b>Time</b>	<b>Anos</b>
12	Palmeiras	1960, 1967, 1967, 1969, 1972, 1973, 1993, 1994, 2016, 2018, 2022, 2023
8	Santos	1961, 1962, 1963, 1964, 1965, 1968, 2002, 2004
7	Flamengo	1980, 1982, 1983, 1992, 2009, 2019, 2020
7	Corinthians	1990, 1998, 1999, 2005, 2011, 2015, 2017
6	São Paulo	1977, 1986, 1991, 2006, 2007, 2008
4	Cruzeiro	1966, 2003, 2013, 2014
4	Fluminense	1970, 1984, 2010, 2012
4	Vasco	1974, 1989, 1997, 2000
3	Botafogo	1968, 1995, 2024
3	Atlético-MG	1937, 1971, 2021
3	Internacional	1975, 1976, 1979
2	Grêmio	1981, 1996
2	Bahia	1959, 1988
1	Athletico-PR	2001
1	Sport	1987
1	Coritiba	1985
1	Guarani	1978

Fonte: Adaptada de Estadão (2023)

### 3.2 DISTRIBUIÇÕES DE PROBABILIDADE

Segundo Magalhães (2006), uma distribuição de probabilidade descreve como os possíveis resultados de um experimento aleatório ou evento são distribuídos em termos de suas probabilidades. Essa distribuição, em geral, está atrelada a uma função ou a uma tabela que relaciona evento e probabilidade. Vale destacar que essa probabilidade não deve assumir valores negativos, e o somatório das probabilidades de todos os eventos deve ser igual a 1.

As distribuições Normal, Binomial, Poisson, Bernoulli, Uniforme, Geométrica, Hipergeométrica, Gama, F, t e Beta estão entre as mais conhecidas e utilizadas para modelar a probabilidade de eventos e experimentos. No entanto, existe uma infinidade de outras distribuições que podem ser aplicadas em contextos gerais e específicos. A distribuição destacada neste trabalho foi a distribuição de Poisson, que foi usada para modelar a probabilidade de gols feitos por cada time em dois modelos.

### 3.2.1 Distribuição de Poisson Univariada

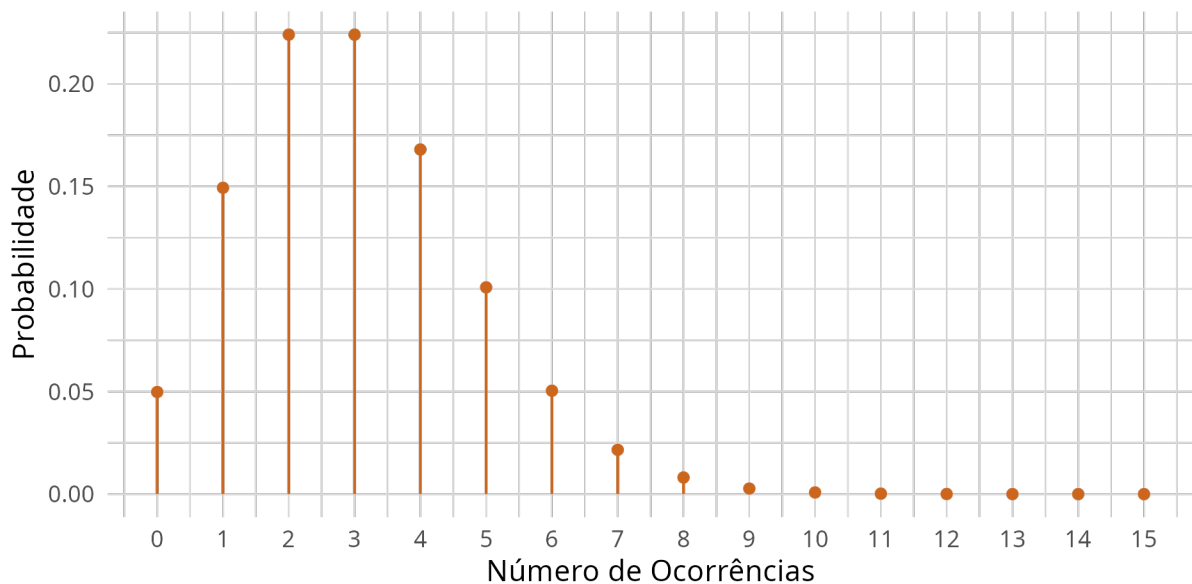
Essa distribuição, que originalmente foi introduzida em 1837 por Siméon Denis Poisson, é utilizada para modelar a probabilidade de eventos ditos raros, uma vez que o número de ocorrências não é muito alto (ROSS, 1976). Além disso, esse número é discreto, ou seja, essa variável aleatória assume apenas valores inteiros.

Essa distribuição é dita univariada já que observa-se o comportamento de apenas uma variável, como representado abaixo, a variável  $X$ . A função de probabilidade dessa distribuição é dada por:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (3.1)$$

O único parâmetro dessa distribuição é o  $\lambda$ , ele é interpretado como a taxa de ocorrência de um determinado evento em relação a uma unidade amostral. Em particular, esse valor também representa a média e a variância dessa distribuição. Observe a seguir a representação gráfica de uma distribuição de Poisson gerada a partir da expressão 3.1 e com  $\lambda = 3$

Figura 1 – Exemplo gráfico de uma distribuição Poisson Univariada com  $\lambda = 3$ .



Fonte: dos Autores

Note que a distribuição ilustrada pela Figura 1 apresenta uma assimetria à direita que se localiza próxima ao valor da média dessa distribuição. Para intervalos maiores, é possível constatar que sua distribuição se assemelha à distribuição normal fixada em torno do meu valor de  $\lambda$ .

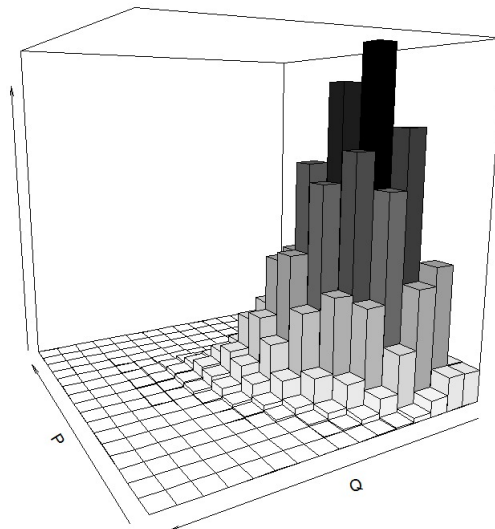
### 3.2.2 Distribuição de Poisson Bivariada

Para iniciar o estudo dessa distribuição, será apresentado primeiramente um caso particular que se relaciona mais diretamente com o caso univariado: a distribuição de Poisson bivariada independente. Ela é análoga ao caso mais simples, com apenas uma variável. Entretanto, neste cenário, observa-se o comportamento de  $X$  e  $Y$  (independentes), sendo que cada uma dessas variáveis aleatórias possui seu próprio parâmetro  $\lambda_X$  e  $\lambda_Y$ , referentes à distribuição da qual foram amostradas. Assim, sua função de probabilidade é dada por:

$$P(X = x, Y = y) = \frac{e^{-(\lambda_X + \lambda_Y)} (\lambda_X^x \lambda_Y^y)}{x! y!} \quad (3.2)$$

A partir da distribuição mostrada na Equação 3.2, é possível obter uma representação gráfica. Como ela varia em duas dimensões, sua visualização é feita no espaço tridimensional, como no exemplo da Figura 2.

Figura 2 – Exemplo gráfico de uma distribuição de Poisson bivariada de  $P$  e  $Q$ .



Fonte: Adaptado de Astivia (2020).

A distribuição de Poisson bivariada geral considera duas variáveis aleatórias  $X$  e  $Y$  tais que:  $X = X' + U$  e  $Y = Y' + U$ , em que  $X'$ ,  $Y'$  e  $U$  são variáveis aleatórias independentes com distribuição de Poisson univariada e parâmetros  $\lambda_{X'}$ ,  $\lambda_{Y'}$  e  $\lambda_U$ , respectivamente (HOLGATE, 1964). Ainda, uma vez que  $X \sim \text{Poisson}(\lambda_X)$  e  $Y \sim \text{Poisson}(\lambda_Y)$ , pode-se observar as seguintes relações:

$$\lambda_U = \text{Cov}(XY)$$

$$\lambda_X = \lambda_{X'} + \lambda_U$$

$$\lambda_Y = \lambda_{Y'} + \lambda_U$$

Portanto,  $X \sim \text{Poisson}(\lambda_{X'} + \lambda_U)$  e  $Y \sim \text{Poisson}(\lambda_{Y'} + \lambda_U)$ . Logo, sua função de probabilidade está representada na equação 3.3.

$$P(X = x, Y = y) = e^{-(\lambda_{X'} + \lambda_{Y'} + \lambda_U)} \sum_{k=0}^{\min(x,y)} \frac{\lambda_{X'}^{x-k} \lambda_{Y'}^{y-k} \lambda_U^k}{(x-k)!(y-k)!k!} \quad (3.3)$$

Essa função de distribuição pode ser construída de duas formas. A maneira mais simples foi apresentada em Arruda (2000) e está detalhada abaixo:

$$\begin{aligned} P(X = x, Y = y) &= P(X' + U = x, Y' + U = y) \\ &= \sum_{k=0}^{\infty} P(X' + U = x, Y' + U = y \mid U = k)P(U = k) \\ &= \sum_{k=0}^{\infty} P(X' = x - k, Y' = y - k)P(U = k) \\ &= \sum_{k=0}^{\min(x,y)} P(X' = x - k)P(Y' = y - k)P(U = k) \\ &= \sum_{k=0}^{\min(x,y)} \frac{e^{-\lambda_{X'}} \lambda_{X'}^{x-k}}{(x-k)!} \frac{e^{-\lambda_{Y'}} \lambda_{Y'}^{y-k}}{(y-k)!} \frac{e^{-\lambda_U} \lambda_U^k}{k!} \\ &= e^{-(\lambda_{X'} + \lambda_{Y'} + \lambda_U)} \sum_{k=0}^{\min(x,y)} \frac{\lambda_{X'}^{x-k} \lambda_{Y'}^{y-k} \lambda_U^k}{(x-k)!(y-k)!k!}. \end{aligned}$$

A outra maneira de construir essa distribuição é por meio do limite de uma distribuição binomial bivariada que foi demonstrada por Hamdan e Al-Bayyati (1969) e apresenta uma demonstração mais direta do que a proposta originalmente por Campbell (1934). Essa construção parte da observação simultânea de duas variáveis Bernoulli  $A$  e  $B$ . Utilizando a notação de Teicher (1954), define-se  $p_{11} = P(A = 1, B = 1)$ ,  $p_{01} = P(A = 0, B = 1)$ ,  $p_{10} = P(A = 1, B = 0)$  e  $p_{00} = P(A = 0, B = 0)$  com  $\sum_{i,j=0}^1 p_{ij} = 1$ . Assim temos:

$$P(A = i, B = j) = p_{11}^m p_{10}^{i-m} p_{01}^{j-m} p_{00}^{1-i-j-m},$$

em que  $i, j \in \{0,1\}$ , e  $m = \min(i, j)$ . Essa expressão é um passo a mais, expresso por Ho (1995) para favorecer o didatismo da construção. Agora, se essa observação das variáveis A e B forem repetidas um número  $n$  de vezes de maneira independente teremos:

$$P(U = n_1, X' = n_2, Y' = n_3) = \frac{n! p_{11}^{n_1} p_{10}^{n_2} p_{01}^{n_3} p_{00}^{n - \sum_{i=1}^3 n_i}}{n_1! n_2! n_3! (n - \sum_{i=1}^3 n_i)!} \quad (3.4)$$

sendo que  $U, X'$  e  $Y'$  representam o número de vezes que ocorreram os resultados  $(A = 1, B = 1)$ ,  $(A = 1, B = 0)$   $(A = 0, B = 1)$ , respectivamente. Se a equação 3.4 for considerar  $X$  e  $Y$  como definido no início dessa sessão temos:

$$P(X = x, Y = y) = \sum_{k=0}^{\min(x,y)} \frac{n! p_{11}^k p_{10}^{x-k} p_{01}^{y-k} p_{00}^{n-x-y-k}}{k!(x-k)!(y-k)!(n-x-y+k)!} \quad (3.5)$$

Segundo Marshall e Olkin (1985) a expressão 3.5 tem distribuição binomial bivariada com parâmetros  $n, p_{10}, p_{01}, p_{11}$  e  $p_{00}$ . Se for tomado o limite com  $n \rightarrow \infty$  e conseqüentemente  $p_{10}, p_{01}, p_{11} \rightarrow 0$ , mantendo  $x$  e  $y$  fixos e definindo as seguintes variáveis,  $n(p_{10} + p_{11}) = \lambda_X$ ,  $n(p_{01} + p_{11}) = \lambda_Y$  e  $np_{11} = \lambda_U$ , temos a seguinte expressão (HAMDAN; AL-BAYYATI, 1969).

$$\begin{aligned} \lim_{n \rightarrow \infty} P(X = x, Y = y) &= \sum_{k=0}^{\min(x,y)} \lim_{n \rightarrow \infty} \left[ \frac{n!}{k!(x-k)!(y-k)!(n-x-y+k)!} \right. \\ &\quad \left. \left( \frac{\lambda_U}{n} \right)^k \left( \frac{\lambda_X - \lambda_U}{n} \right)^{x-k} \left( \frac{\lambda_Y - \lambda_U}{n} \right)^{y-k} \left( 1 - \frac{\lambda_X + \lambda_Y - \lambda_U}{n} \right)^{n-x-y+k} \right] \\ &= \sum_{k=0}^{\min(x,y)} \lim_{n \rightarrow \infty} \left[ \frac{n(n-1) \cdots (n-k)(n-k-1) \cdots (n-x) \cdots (n-x-y+k+1)}{k!(x-k)!(y-k)!} \right. \\ &\quad \left. \left( \frac{\lambda_U}{n} \right)^k \left( \frac{\lambda_X - \lambda_U}{n} \right)^{x-k} \left( \frac{\lambda_Y - \lambda_U}{n} \right)^{y-k} \left( 1 - \frac{\lambda_X + \lambda_Y - \lambda_U}{n} \right)^{n-x-y+k} \right] \\ &= \sum_{k=0}^{\min(x,y)} \lim_{n \rightarrow \infty} \left[ \frac{1(1 - \frac{1}{n}) \cdots (1 - \frac{k}{n}) \cdots (1 - \frac{x}{n}) \cdots (1 - \frac{x+y-k-1}{n})}{k!(x-k)!(y-k)!} \right. \\ &\quad \left. (\lambda_U)^k (\lambda_X - \lambda_U)^{x-k} (\lambda_Y - \lambda_U)^{y-k} \left( 1 - \frac{\lambda_X + \lambda_Y - \lambda_U}{n} \right)^{n+k-x-y} \right] \end{aligned}$$

$$\begin{aligned}
&= e^{-(\lambda_X + \lambda_Y - \lambda_U)} \sum_{k=0}^{\min(x,y)} \frac{(\lambda_U)^k (\lambda_X - \lambda_U)^{x-k} (\lambda_Y - \lambda_U)^{y-k}}{k!(x-k)!(y-k)!} \\
&= e^{-(\lambda_{X'} + \lambda_{Y'} + \lambda_U)} \sum_{k=0}^{\min(x,y)} \frac{\lambda_U^k \lambda_{X'}^{x-k} \lambda_{Y'}^{y-k}}{k!(x-k)!(y-k)!}
\end{aligned}$$

Obtém-se, assim, a distribuição de Poisson bivariada apresentada na Equação 3.3. Outro ponto importante a ser destacado é a função de distribuição acumulada, expressa na Equação 3.6. Essa função descreve a probabilidade de que as variáveis aleatórias assumam valores menores ou iguais a um determinado par  $(x, y)$ , representando, portanto, a probabilidade acumulada conjunta da distribuição.

$$F(x, y) = \sum_{x'=0}^x \sum_{y'=0}^y e^{-(\lambda_{X'} + \lambda_{Y'} + \lambda_U)} \sum_{k=0}^{\min(x',y')} \frac{\lambda_{X'}^{x'-k} \lambda_{Y'}^{y'-k} \lambda_U^k}{(x'-k)!(y'-k)!k!} \quad (3.6)$$

### 3.2.2.1 Método dos Momentos

Para estimar os parâmetros da distribuição de Poisson Bivariada é possível utilizar o método de momentos que consiste em igualar os momentos populacionais aos momentos amostrais. Para a distribuição de Poisson bivariada de Holgate, os momentos populacionais são:

$$\mathbb{E}[X] = \lambda_{X'} + \lambda_U, \quad \mathbb{E}[Y] = \lambda_{Y'} + \lambda_U$$

$$\text{Var}(X) = \lambda_{X'} + \lambda_U, \quad \text{Var}(Y) = \lambda_{Y'} + \lambda_U.$$

$$\text{Cov}(X, Y) = \lambda_U$$

Dada uma amostra de  $n$  observações  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , os momentos amostrais são calculados como:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

$$S_X^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2, \quad S_Y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{Y})^2.$$

$$S_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y}).$$

Igualando os momentos populacionais aos momentos amostrais, obtemos o seguinte sistema de equações:

$$\bar{X} = \lambda_{X'} + \lambda_U, \quad \bar{Y} = \lambda_{Y'} + \lambda_U,$$

$$S_X^2 = \lambda_{X'} + \lambda_U, \quad S_Y^2 = \lambda_{Y'} + \lambda_U,$$

$$S_{XY} = \lambda_U.$$

Resolvendo o sistema, obtemos as estimativas dos parâmetros:

$$\begin{cases} \lambda_{X'} = \bar{X} - S_{XY} \\ \lambda_{Y'} = \bar{Y} - S_{XY} \\ \lambda_U = S_{XY} \end{cases}$$

### 3.3 MODELOS DE PREDIÇÃO DE RESULTADOS DE FUTEBOL

Nesta seção, busca-se apresentar um apanhado de trabalhos dedicados à elaboração de modelos de previsão de resultados de partidas de futebol, com o objetivo de fornecer ao leitor um panorama geral sobre o tema. Dessa forma, pretende-se que leitores sem experiência prévia possam conhecer os métodos mais populares, bem como os principais parâmetros e métricas utilizados na literatura. Para os leitores mais experientes, esta seção busca servir como fonte de consulta e acervo, permitindo o diálogo com resultados já consolidados.

De modo geral, três abordagens principais podem ser destacadas na construção de modelos preditivos para resultados de partidas de futebol: métodos determinísticos, probabilísticos e computacionais. Além disso, observa-se que os campeonatos mais recorrentes como objeto

de estudo incluem a *UEFA Champions League*, o Campeonato Europeu de Futebol, a *La Liga* e, no contexto brasileiro, o Campeonato Brasileiro de Futebol. Abaixo estão listados alguns trabalhos e um breve resumo sobre eles.

- Wong et al. (2025) Utilizando dados da *Premier League* (2019–2022), o estudo comparou diferentes algoritmos de aprendizado de máquina — Regressão Logística, Random Florest, SVM, XGBoost, LightGBM e CNN — na previsão de resultados. Para avaliação dos modelos foram utilizadas medidas baseadas na matriz de confusão dos resultados, e foram chamados de acurácia, precisão, recall, F1 score e Coeficiente de Correlação de Matthews Normalizado. O SVM apresentou o melhor equilíbrio entre precisão e sensibilidade (F1 = 60,6%), enquanto o XGBoost obteve a maior acurácia (62,3%). Já o LightGBM e a Regressão Logística tiveram desempenho intermediário, e o Random Forest e a CNN apresentaram os piores resultados. Ao comparar os modelos com as casas de apostas (Bet365, William Hill e Pinnacle), observou-se que estas possuem acurácia ligeiramente superior ( $\approx 66\%$ ), porém menor sensibilidade. O modelo de *voting* destacou-se por identificar melhor derrotas do time da casa (recall = 53,7%), sugerindo uma divergência preditiva e possíveis oportunidades lucrativas frente às previsões das casas.
- Stafylidis et al. (2024) Este estudo apresenta uma análise descritiva de diversas métricas observáveis em partidas de futebol e suas relações com os possíveis resultados. Foram analisadas as 36 partidas da fase de grupos da *UEFA Euro 2024*, investigando fatores que influenciam o desfecho dos jogos a partir de indicadores ofensivos, defensivos e de goleiro. Testes do qui-quadrado revelaram que marcar o primeiro gol aumenta significativamente as chances de um resultado positivo, embora não tenham sido encontradas diferenças relevantes na distribuição temporal dos gols entre os tempos ou intervalos de 15 minutos. Os testes de Kruskal–Wallis indicaram que as equipes vencedoras apresentaram melhor desempenho em indicadores ofensivos, como assistências, finalizações e entradas na grande área, além de maior eficiência defensiva, evidenciada por mais recuperações de bola e partidas sem sofrer gols. O modelo de regressão logística identificou as finalizações e os passes para a grande área como preditores positivos de vitória, enquanto as finalizações de fora da área e as tentativas de cruzamento apresentaram efeito negativo. Conclui-se que a eficiência ofensiva e o controle tático são determinantes para o sucesso, ressaltando a importância de estratégias equilibradas entre ataque e defesa no futebol de alto rendimento.

- Galvão e Bueno Filho (2020) O estudo propõe um modelo estatístico baseado na distribuição de Poisson para estimar e prever o número esperado de gols em partidas do Campeonato Brasileiro de Futebol de 2017 (Série A). O modelo considera parâmetros de força ofensiva ( $\hat{\gamma}$ ) e defensiva ( $\hat{\delta}$ ) para cada equipe, além de um parâmetro comum de vantagem de mando de campo ( $\mu$ ). Inicialmente, as estimativas foram obtidas a partir do primeiro turno e, em seguida, reestimadas rodada a rodada no segundo turno. As probabilidades de vitória, empate e derrota foram derivadas a partir dessas estimativas, permitindo prever resultados e identificar partidas com desempenho inesperado. Os resultados mostraram que equipes como Corinthians, Grêmio e Palmeiras apresentaram elevados parâmetros de ataque, enquanto clubes em posições inferiores foram caracterizados por baixa força defensiva. O parâmetro de mando de campo estimado ( $\hat{\mu} = 0,31$ ) influenciou significativamente as probabilidades de vitória das equipes mandantes. Conclui-se que as estimativas obtidas por máxima verossimilhança para parâmetros de ataque e defesa são ferramentas eficazes para descrever o desempenho das equipes e prever resultados de jogos, sendo o modelo adaptável a torneios de pontos corridos e passível de aprimoramento por meio de inferência Bayesiana.
- Mendes-Neves et al. (2025) O estudo propõe um método alternativo para a previsão de resultados de partidas de futebol utilizando aprendizado de máquina, integrando o sistema de pontuação criado por Arpad Elo, originalmente desenvolvido para o xadrez. Em vez de estimar diretamente a probabilidade de vitória ou o número de gols, o modelo prevê distribuições normais para o número de finalizações e para a qualidade média dessas finalizações de cada equipe, condicionadas aos índices ELO do time mandante e visitante. A partir dessas distribuições, simulam-se milhares de partidas, o que permite estimar as probabilidades de resultados e placares exatos. A metodologia foi avaliada com dados de partidas das ligas europeias, empregando métricas como o Escore de Probabilidade Classificada (0,201), a Raiz do Erro Quadrático Médio (1,58) e o Erro Absoluto Médio (1,65). Entre os algoritmos testados — Regressão Linear, Árvore de Decisão e Floresta Aleatória — o modelo apresentou desempenho competitivo, superando a linha de base do mercado de apostas, cuja margem média é de -5,6%.
- Rico-González et al. (2023) O artigo apresenta uma revisão sistemática focada em modelos de aprendizado de máquina aplicados ao futebol, analisando 32 estudos originais, dos quais 7 são voltados à previsão de lesões, 21 à previsão de desempenho e 5 à identificação

de talentos. Destaca-se o uso da acurácia como a principal métrica de avaliação, obtida por meio da matriz de confusão, com valores variando entre 50% e 100%, despencando principalmente nos modelos de previsão de desempenho. Quanto às fontes de dados utilizadas, elas incluem desde bancos de dados prontos até observações diretas, imagens de câmeras, informações de geolocalização, métricas de saúde e desempenho físico dos jogadores, além de combinações entre esses diferentes tipos de dados. Dos trabalhos analisados nesse estudo apenas um deles tem como objetivo prever os resultados final de uma partida, se assemelhando com a proposta da presente dissertação, assim esse trabalho foi selecionado para compor esse referencial e está descrito a abaixo.

- Bilek e Ulas (2019) O estudo teve como objetivo investigar as variáveis situacionais e indicadores de desempenho que influenciam significativamente o resultado das partidas (vitória, empate ou derrota) na Premier League inglesa durante a temporada 2017–2018, totalizando 760 partidas, considerando a qualidade do adversário. Foram utilizados 22 indicadores e variáveis situacionais, aplicando-se análises de variância (ANOVA unidirecional), teste de Tukey, agrupamento *k-means* e árvores de decisão. Os resultados indicaram que marcar o primeiro gol é o fator mais determinante, aparecendo como nó raiz em todas as árvores de decisão, elevando a probabilidade de vitória para até 86% contra adversários mais fracos. Outros indicadores relevantes incluem o número de chutes, chutes no alvo, desarmes, posse de bola e mando de campo, sendo que a influência desses fatores varia conforme a força do oponente. As árvores de decisão atingiram acurácias de 67,9%, 73,9% e 78,4% para jogos contra adversários de qualidade equilibrada, superior e inferior, respectivamente. Quando a qualidade do adversário não foi considerada, a taxa de acerto caiu para 64,8%, evidenciando a importância de incorporar esse fator nas análises.
- Bunker, Yeung e Fujii (2024) Entre os principais bancos de dados discutidos estão o *Open International Soccer Database*, com mais de 216 mil partidas em 35 países; o *European Soccer Database*, que reúne dados de jogadores, formações e probabilidades de apostas; e bases comerciais como *StatsBomb* e *Wyscout*, que fornecem informações espaço-temporais detalhadas. Fontes complementares incluem sistemas de classificação (Elo, pi-ratings, Berrar ratings e GAP ratings) e sites de apostas. Os modelos utilizados são organizados em três grupos: (i) modelos estatísticos — como distribuições de Poisson e regressões logísticas; (ii) modelos de aprendizado de máquina — incluindo

algoritmos genéticos, fuzzy, redes bayesianas e métodos probabilísticos; e (iii) sistemas de classificação — como Elo e pi-ratings. Destaca-se que modelos baseados em árvores potencializadas por gradiente, como o *CatBoost*, quando aplicados a ratings específicos do futebol (pi-ratings), apresentam desempenho superior em bases contendo apenas dados de gols. No *Football Prediction Challenge 2017*, modelos como o *XGBoost* e redes bayesianas híbridas obtiveram os melhores desempenhos, com Escore de Probabilidade Classificada (RPS) em torno de 0,20 e acurácia próxima de 52%.

- Arruda (2000) propõe seis modelos distintos baseados em metodologias próprias associadas diretamente à distribuição de Poisson bivariada, considerando ou não a correlação entre as variáveis. A principal métrica de avaliação utilizada é a Medida de DeFinetti, complementada pela Medida de Confiabilidade. Dentre os modelos analisados, destacam-se: o modelo *SD0*, pela sua aplicabilidade e fundamentação metodológica; o modelo *Chance I*, que, apesar de apresentar eventuais problemas de convergência dependendo do conjunto de observações empregado na estimação dos parâmetros, demonstra robustez e coerência conceitual; e o modelo *Implícito I*, notável pela simplicidade metodológica ao incorporar diretamente a opinião de especialistas para a definição de seus parâmetros. Os resultados evidenciam desempenhos próximos entre os modelos *Chance I* e *Implícito I*, corroborando a validade de abordagens bayesianas em contextos com informações subjetivas bem fundamentadas.
- Araújo et al. (2015) O estudo propõe um modelo de previsão de resultados no futebol brasileiro utilizando a distribuição de Poisson para o número de gols marcados e o método Soma e Diferença (*SD0*) para estimar os parâmetros. A metodologia foi aplicada ao Campeonato Brasileiro de Futebol de 2014, permitindo simular toda a segunda fase da competição e calcular probabilidades de eventos de interesse, como ser campeão, classificar para a Taça Libertadores e ser rebaixado. A avaliação do modelo com a medida de DeFinetti indicou boa capacidade preditiva, com aproximadamente 50% de acertos até a 33ª rodada.
- Santana et al. (2020) Este estudo apresenta duas abordagens complementares para previsão de resultados no futebol, aplicadas à Série A do Campeonato Brasileiro de 2017. A primeira parte consiste em uma modelagem atemporal de classificação, na qual são testadas três técnicas supervisionadas: Regressão Logística Polinômica, Máquinas de Vetores

de Suporte e Florestas Aleatórias. As duas primeiras técnicas apresentaram previsões de qualidade aceitável (Brier Score  $< 2/3$ ), enquanto o modelo de Florestas Aleatórias teve desempenho insatisfatório (Brier Score = 0,7706). A segunda parte utiliza uma modelagem temporal de regressão, com o modelo Poisson Auto-Regressivo com Covariáveis Exógenas (PARX), para estimar o número de gols marcados por cada equipe e prever o placar final das partidas, incorporando dependência serial e superdispersão, características comuns em dados de futebol. Os resultados foram satisfatórios, com diagnósticos adequados segundo os gráficos PIT e de calibração marginal, e Brier Score também inferior a  $2/3$ .

- Lima et al. (2012) O artigo apresenta um modelo matemático de previsão de resultados no futebol que combina ajuste determinístico e simulação estocástica, buscando representar a dinâmica de desempenho das equipes ao longo do tempo. O modelo define, para cada time, dois vetores de probabilidade: um para quando atua como mandante ( $PM = (pvm, pem, pdm)$ ) e outro para quando atua como visitante ( $PV = (p dv, pev, pvv)$ ). Esses vetores representam, respectivamente, as probabilidades de vitória, empate e derrota em cada condição. O modelo também considera situações especiais, como jogos “clássicos”, em que ambos os times são tratados como mandantes para efeito de atualização. Após os ajustes determinísticos, introduz-se aleatoriedade via simulação de Monte Carlo, sorteando números  $\gamma$  no intervalo  $[0,1]$  para determinar o resultado do jogo com base nas probabilidades calculadas. Cada confronto é simulado 10.000 vezes, permitindo estimar empiricamente as probabilidades finais de vitória do mandante, empate e vitória do visitante.
- Ramos, Fernandes e Batista (2021) O artigo propõe um método simples baseado na distribuição de Poisson para estimar probabilidades em partidas de futebol, aplicando-o a 250 jogos da Série A do Campeonato Brasileiro de 2018 e comparando seus resultados com o modelo desenvolvido pelo Departamento de Matemática da UFMG. O modelo assume independência entre os gols marcados por cada time, cujas médias refletem fatores de ataque e defesa. A metodologia, implementada no software R, atualiza os parâmetros rodada a rodada, aprendendo com o desempenho dos times ao longo do campeonato. O modelo apresentou média da medida de DeFinetti igual a 0,568, com cerca de 57% dos resultados abaixo de  $2/3$ , indicando boa qualidade preditiva. Foi capaz de prever corretamente o campeão (Palmeiras), além de gerar resultados coerentes para classificados à

Libertadores e times rebaixados. Apesar da boa performance nas vitórias dos mandantes, apresentou limitações na previsão de empates e vitórias dos visitantes. O estudo sugere melhorias como ponderação de gols por dificuldade do adversário, uso de médias móveis e aplicação do modelo a outros torneios de pontos corridos. Destaca-se que o método do PROFMAT é mais adequado a torneios de mata-mata, enquanto o modelo da UFMG se mostra mais estável para campeonatos de pontos corridos.

## 4 MATERIAL E MÉTODOS

No primeiro tópico desta seção, serão abordados os bancos de dados utilizados como principais fontes de informação para alimentar os modelos de produção. O segundo tópico apresenta os modelos e suas adaptações, descrevendo passo a passo as premissas e hipóteses consideradas para as modificações. Por fim, no terceiro tópico, é detalhada a metodologia utilizada para o estudo de simulação que busca definir, em rodadas, a janela de tempo ideal para as previsões de alguns modelos selecionados.

### 4.1 MATERIAL

O principal banco de dados utilizado nesta pesquisa foi obtido na plataforma Kaggle, de autoria de Duque (2023). Esse banco contém informações sobre as partidas do Campeonato Brasileiro de Futebol entre os anos de 2003 e 2023. Entre os dados disponíveis, estão a rodada, a data da partida, o ano de realização, o horário, os times mandante e visitante, os respectivos placares, formações táticas e o campo de jogo.

Foi necessário incluir as partidas referentes ao ano de 2024. Para isso, as informações foram extraídas das tabelas detalhadas da Confederação Brasileira de Futebol (2024). Além disso, foram criadas tabelas auxiliares com informações que permitiram ajustes finos nos modelos estatísticos e a construção de parâmetros indiretos sobre as partidas, como, por exemplo, a identificação do mando de campo, como será melhor detalhado mais à frente.

Para a realização das análises e de todo o trabalho computacional, foram utilizados o software R Core Team (2024) e sua IDE RStudio (Posit team, 2025), além do pacote Tidyverse Wickham et al. (2019). Adicionalmente, buscou-se apoio em fóruns de fãs e torcedores disponíveis na internet, os quais auxiliaram na construção dos critérios adotados, bem como na consolidação do conhecimento e do senso esportivo por parte dos autores.

## 4.2 MÉTODOS

A seguir, estão listados e detalhados 35 modelos desenvolvidos com o objetivo de prever os resultados de partidas de futebol. É possível dividi-los em três grupos principais:

- Modelos baseados em Arruda (2000): fundamentados nos modelos propostos por Marcelo Arruda;
- Modelos baseados em (LIMA et al., 2012): derivados das propostas metodológicas apresentadas pela equipe da UFMG;
- Modelos controle: composto por modelos de metodologia simples, cuja função principal é servir como referência para comparação com os demais.

Com o objetivo de aumentar a acurácia das previsões, foram propostas modificações nos modelos originais. Essas alterações abrangem tanto aspectos metodológicos quanto o tratamento da janela temporal dos modelos. Os modelos originais, bem como suas respectivas modificações, estão descritos a seguir.

### 4.2.1 Modelo *SD 0* de Arruda (2000)

O modelo *SD 0* utiliza uma abordagem estatística para tratar os dados e realizar previsões. Em particular, concentra-se em prever  $X$  e  $Y$ , que representam, respectivamente, o número de gols do time mandante ( $time_m$ ) e do time visitante ( $time_v$ ). Segundo Arruda (2000),  $X$  e  $Y$  seguem distribuições de Poisson, mas o vetor  $(X, Y)$  segue uma distribuição de Poisson bivariada.

Dessa forma, o autor considera as seguintes esperanças marginais:  $E[X] = \lambda_x + \lambda_{xy}$  e  $E[Y] = \lambda_y + \lambda_{xy}$ . Aqui,  $\lambda_x$  e  $\lambda_y$  são parâmetros de distribuições de Poisson univariadas, que refletem, respectivamente, a frequência de gols do  $time_m$  e do  $time_v$ , assumidas como independentes entre si. Já o parâmetro  $\lambda_{xy}$  representa a covariância entre  $X$  e  $Y$ . No entanto, o modelo *SD 0* considera essa covariância nula, ou seja,  $\lambda_{xy} = 0$ .

O autor utiliza o método dos momentos para a estimação pontual dos parâmetros, valendo-se de um estimador indireto para calcular  $\lambda_x$  e  $\lambda_y$ . Esses estimadores são obtidos a partir da solução do sistema apresentado na Equação 4.1.

$$\begin{cases} E[X - Y] = \lambda_x - \lambda_y \\ E[X + Y] = \lambda_x + \lambda_y \end{cases} \quad \begin{cases} \hat{\lambda}_x = \frac{\hat{E}[X-Y] + \hat{E}[X+Y]}{2} \\ \hat{\lambda}_y = \frac{\hat{E}[X+Y] - \hat{E}[X-Y]}{2} \end{cases} \quad (4.1)$$

Por conseguinte, os valores de  $E[X + Y]$  e  $E[X - Y]$  são estimados por meio de modelos de regressão linear. Uma vez obtidos os valores de  $\lambda_X$  e  $\lambda_Y$ , sabe-se que  $X \sim \text{Poisson}(\lambda_X)$  e  $Y \sim \text{Poisson}(\lambda_Y)$ . Portanto, para obter o placar de maior probabilidade, basta cruzar as informações dessas distribuições em uma matriz de probabilidade.

#### 4.2.2 Modelo SD I de Arruda (2000)

Como variação do modelo *SD 0*, o modelo *SD I* adota os mesmos princípios metodológicos. A principal diferença entre os dois modelos reside na premissa adotada sobre a covariância das variáveis aleatórias  $X$  e  $Y$ , ou, conforme já apresentado, sobre o parâmetro  $\lambda_{xy}$ . Diferentemente do modelo anterior, neste caso admite-se  $\lambda_{xy} \neq 0$ , ou seja, assume-se a existência de dependência entre  $X$  e  $Y$ . Com essa suposição, o autor propõe o seguinte sistema:

$$\begin{cases} E[X - Y] = \lambda_x - \lambda_y \\ E[X + Y] = \lambda_x + \lambda_y + 2\lambda_{xy} \\ E[(X + Y)^2] - (E[X + Y])^2 = \lambda_x + 4\lambda_{xy} + \lambda_y \end{cases}$$

Isolando dos  $\lambda$ 's do sistema tem-se:

$$\begin{cases} \hat{\lambda}_x = \frac{\hat{E}[X - Y] + 2\hat{E}[X + Y] - \left\{ \hat{E}[(X + Y)^2] - \left( \hat{E}[X + Y] \right)^2 \right\}}{2} \\ \hat{\lambda}_y = \frac{2\hat{E}[X + Y] - \hat{E}[X - Y] - \left\{ \hat{E}[(X + Y)^2] - \left( \hat{E}[X + Y] \right)^2 \right\}}{2} \\ \hat{\lambda}_{xy} = \frac{\left\{ \hat{E}[(X + Y)^2] - \left( \hat{E}[X + Y] \right)^2 \right\} - \hat{E}[X + Y]}{2} \end{cases}$$

Assim como foi feito no modelo anterior  $E[X + Y]$ ,  $E[X - Y]$  e  $E[(X + Y)^2]$  são estimados por meio de modelos de regressão linear. Assim, da mesma forma que no modelo

*SD 0* o placar mais provável é obtido através da matriz de probabilidade das distribuições:  $X \sim \text{Poisson}(\lambda_x + \lambda_{xy})$  e  $Y \sim \text{Poisson}(\lambda_y + \lambda_{xy})$ .

### 4.2.3 Modelo *Chance I* de Arruda (2000)

O funcionamento deste modelo é muito semelhante ao que já foi estabelecido para o modelo *SD 0*, sendo que  $\lambda_x$  e  $\lambda_y$  são consideradas independentes ( $\lambda_{xy} = 0$ ) e  $X \sim \text{Poisson}(\lambda_x)$  e  $Y \sim \text{Poisson}(\lambda_y)$ , para uma dada partida do campeonato. Agora, para estimar  $\lambda_x$  e  $\lambda_y$  utiliza-se um modelo log-linear de Poisson e o método de máxima verossimilhança para estimar os parâmetros desse modelo. Assim, os logaritmos de suas funções de verossimilhança são dadas por:

$$L(\lambda_x, X) = -\lambda_x + x \log \lambda_x - \log(x!)$$

$$L(\lambda_y, Y) = -\lambda_y + y \log \lambda_y - \log(y!)$$

Segundo Arruda (2000), por meio da variável explicativa  $\lambda = e^{U\beta}$ , é possível relacionar a distribuição do número de gols marcados com as variáveis indicadoras, como mostrada a seguir.

$$L(\lambda_x, X) = -e^{U_x\beta} + xU_x\beta - \log(x!)$$

$$L(\lambda_y, Y) = -e^{U_y\beta} + yU_y\beta - \log(y!)$$

Uma vez que,  $U_x$  e  $U_y$  são vetores que indicam os times que estão participando do jogo em questão, cada um dos vetores apresenta os jogos de maneira específica visando a percepção do time que aparece em seu índice. Ainda,  $\beta$  informa parâmetros de ataque e defesa para os times do campeonato. Aplicando isso a um número  $k$  de jogos utilizados para estimar os parâmetros e unindo as informações de mandante e visitante, têm-se a seguinte função de verossimilhança:

$$L(\lambda_{x_1}, \lambda_{y_1}, \dots, \lambda_{x_k}, \lambda_{y_k}, X_1, Y_1, \dots, X_k, Y_k) = \sum_{i=1}^k (-e^{U_{x_i}\beta} - e^{U_{y_i}\beta} + x_i U_{x_i}\beta + y_i U_{y_i}\beta - \log(x_i!) - \log(y_i!)) \quad (4.2)$$

Seguindo com a resolução, é possível encontrar estimadores para os valores de  $\lambda$  utilizando as derivadas parciais da equação 4.2. Além disso, para estimar os valores de  $\beta$ , empregam-se técnicas de regressão que maximizam a função de log-verossimilhança associada à equação 4.2. Para tal, são utilizados métodos numéricos iterativos, como o algoritmo de Newton-Raphson, que são amplamente aplicados na estimação de parâmetros em modelos estatísticos (MCCULLAGH; NELDER, 1989).

Para determinar o placar mais provável e, conseqüentemente, o resultado do confronto, utilizam-se as distribuições de probabilidade de  $X$  e  $Y$ , parametrizadas pelos seus respectivos valores de  $\lambda$  obtidos, de forma análoga ao procedimento adotado no modelo *SD 0*. Com base nessas distribuições de Poisson, constrói-se uma matriz de probabilidades para os possíveis placares, permitindo identificar o resultado mais provável da partida.

#### 4.2.4 Modelo *Chance II* de Arruda (2000)

De forma análoga aos modelos *SD 0* e *SD I*, o modelo *Chance II* difere do *Chance I* na suposição de que  $\lambda_{xy} \neq 0$ , ou seja, considera-se que há uma componente de dependência entre as variáveis. Assim, temos  $X \sim \text{Poisson}(\lambda_x + \lambda_{xy})$  e  $Y \sim \text{Poisson}(\lambda_y + \lambda_{xy})$ . Logo, para estimar os valores dos parâmetros  $\lambda$ , o seguinte sistema representado em 4.3 foi encontrado, onde  $\beta_0$  é o intercepto do estimador  $\hat{\beta} = (U'U)^{-1}U'X$ . Para mais detalhes desse modelo, vide Arruda (2000, p. 24).

$$\begin{cases} E[X] = \lambda_x + \lambda_{xy} \\ E[Y] = \lambda_y + \lambda_{xy} \\ \beta_0 = \lambda_{xy} \end{cases} \quad (4.3)$$

que determina os seguintes estimadores após solucionado

$$\begin{cases} \hat{\lambda}_x = \hat{E}[X] - \hat{\beta}_0 \\ \hat{\lambda}_y = \hat{E}[Y] - \hat{\beta}_0 \\ \hat{\lambda}_{xy} = \hat{\beta}_0 \end{cases} \quad (4.4)$$

Assim como nos modelos anteriores, os valores de  $E[X]$ ,  $E[Y]$  e o intercepto  $\beta_0$  são estimados por meio do método dos mínimos quadrados. Com base nas estimativas dos parâ-

metros  $\lambda_X$  e  $\lambda_Y$ , assume-se que as variáveis aleatórias  $X$  e  $Y$  seguem distribuições de Poisson, conforme:  $X \sim \text{Poisson}(\lambda_x + \lambda_{xy})$  e  $Y \sim \text{Poisson}(\lambda_y + \lambda_{xy})$ . A partir dessas distribuições, constrói-se uma matriz de probabilidades para os possíveis placares, permitindo identificar o resultado mais provável da partida.

#### 4.2.5 Modelo *UFMG* de Lima et al. (2012)

O modelo possui um caráter matemático por usar um algoritmo baseado em constantes para ajustar as métricas dos times. No entanto, ele incorpora simulações para introduzir aleatoriedade nas previsões, operando com vetores de probabilidades conforme descrito a seguir. Inicialmente, cada time é representado por dois vetores: um para seu desempenho como mandante ( $PM = (pvm, pem, pdm)$ ) e outro como visitante ( $PV = (p dv, pev, pvv)$ ), aplicando-se a todos os times da competição a ser prevista.

Os vetores de probabilidade contêm as seguintes incógnitas: probabilidade de vitória, empate e derrota como mandante ( $pvm, pem, pdm$ ) e como visitante ( $p vv, pev, p dv$ ). A estimativa inicial do resultado de um confronto é obtida pela média entre o vetor  $PM$  do time mandante e o vetor  $PV$  do time visitante, como mostrado na equação 4.5.

$$\begin{aligned} (vm, em, vv)_{tm \times tv} &= \frac{PM_{tm} + PV_{tv}}{2} \\ &= \left( \frac{pvm_{tm} + p dv_{tv}}{2}, \frac{pem_{tm} + pev_{tv}}{2}, \frac{pdm_{tm} + pvv_{tv}}{2} \right) \end{aligned} \quad (4.5)$$

Os valores dos vetores  $PM^0$  e  $PV^0$  são escolhidos arbitrariamente pelo autor, que geralmente inicia com valores próximos de 0,33 para cada uma das componentes. Esses vetores são considerados como ponto de partida, ou seja, sem memória ou influência dos dados. Em seguida, esses vetores são atualizados conforme o algoritmo apresentado abaixo, o qual ilustra o confronto entre o time  $tm$  (mandante) e o time  $tv$  (visitante).

Caso ocorra vitória do mandante:

$$PM_{tm}^{k+1} = \frac{p \cdot PM_{tm}^k + R_{tv}(1,0,0)}{p + R_{tv}}$$

$$PV_{tv}^{k+1} = \frac{p \cdot PV_{tv}^k + (1 - R_{tm})(1,0,0)}{p + (1 - R_{tm})}$$

Caso ocorra vitória do visitante:

$$PM_{tm}^{k+1} = \frac{p \cdot PM_{tm}^k + (1 - R_{tv})(0,0,1)}{p + (1 - R_{tv})}$$

$$PV_{tv}^{k+1} = \frac{p \cdot PV_{tv}^k + R_{tm}(0,0,1)}{p + R_{tm}}$$

Onde  $p$  é um fator de peso que regula a sensibilidade do ajuste dos vetores, ou seja, quanto maior o valor de  $p$ , maior será o impacto da última atualização. Já  $R$  representa o rendimento de um determinado time, que pode ser calculado dividindo-se a soma do número de vitórias com metade do número de empates pelo número total de jogos disputados.

Agora para o caso de empate, considera-se ainda o rendimento dos times que se enfrentam, como mostrado abaixo:

Caso de empate em que o  $R_{tv} \leq 0.50$

$$PM_{tm}^{k+1} = \frac{p \cdot PM_{tm}^k + (1 - 2 \cdot R_{tv})(0, 0.5, 0.5) + 2 \cdot R_{tv}(0,1,0)}{p + 1}$$

Caso de empate em que o  $R_{tv} \geq 0.50$

$$PM_{tm}^{k+1} = \frac{p \cdot PM_{tm}^k + (2 \cdot R_{tv} - 1)(0.5, 0.5, 0) + 2 \cdot (1 - R_{tv})(0,1,0)}{p + 1}$$

O modelo aplica o mesmo raciocínio ao vetor  $PV_{tv}^{k+1}$ , considerando o desempenho do time mandante. Em partidas classificadas como clássicos, ambos os times são tratados como mandantes para atualizar os vetores de desempenho.

O próximo passo envolve incorporar aleatoriedade ao processo de previsão por meio de simulações, que consistem em sortear um número aleatório  $\gamma$  no intervalo real  $[0,1]$ . Com o valor de  $\gamma$  determinado, procede-se à definição do resultado da partida com base nesse sorteio.

- se  $0 \leq \gamma \leq vm$ , considera-se que ocorreu vitória do mandante.

- se  $vm \leq \gamma < em + vm$ , considera-se que ocorreu empate.
- se  $em + vm \leq \gamma \leq 1$ , considera-se que ocorreu vitória do visitante.

Para obter as proporções de ocorrência de cada evento, os autores reproduzem esse sorteio 10000 vezes. Obtendo agora, a probabilidade de ocorrência de cada evento.

#### 4.2.6 Modelo *UFMG* Sem sorteio

Este modelo surge como uma modificação do modelo *UFMG* original, apresentado na Seção 4.2.5, sendo uma nova proposta desenvolvida pelos autores deste trabalho. A modificação proposta parte da hipótese de que a simulação realizada no último passo do modelo *UFMG* original não cumpre seu objetivo, ou seja, não modela adequadamente a aleatoriedade presente em cada partida de futebol.

Para fundamentar essa hipótese, considerou-se que o número aleatório  $\gamma$  provém de uma distribuição uniforme no intervalo  $[0,1]$ , dado que não foi especificada nenhuma distribuição em particular e Lima et al. (2012) descreve implicitamente que cada número nesse intervalo tem a mesma chance de ser sorteado. Assim, a Lei dos Grandes Números, segundo Durrett (2010), garante que, quanto maior o número de vezes que o valor  $\gamma$  for sorteado, mais a média de ocorrência de cada evento se aproximará do tamanho do intervalo.

Portanto, o modelo *UFMG sem sorteio* exclui o passo final do sorteio presente no modelo original. As probabilidades de ocorrência dos eventos são definidas diretamente como: vitória do mandante igual a  $vm$ , empate igual a  $em$  e vitória do visitante igual a  $vv$ .

#### 4.2.7 Modelo *UFMG* Sorteio Normal

O modelo *UFMG Sorteio Normal* parte dos mesmos pressupostos do modelo *UFMG sem sorteio*. Entretanto, ao contrário de simplesmente excluir o sorteio, este modelo associa uma distribuição de probabilidade distinta da uniforme ao número aleatório. A distribuição escolhida é a normal, com média igual a 0 e desvio padrão igual a 0,0001.

#### 4.2.8 Modelo *Controle Teimoso*

O modelo *Controle Teimoso* é dividido em três variantes: *Teimoso Mandante*, *Teimoso Empate* e *Teimoso Visitante*. Esses modelos representam as abordagens mais simples utilizadas neste estudo, pois cada um prevê apenas um resultado possível para o confronto, independentemente do histórico das equipes, do local da partida ou do momento em que ela ocorre.

Por exemplo, o modelo *Teimoso Mandante* prevê, com 100% de certeza, a vitória do time mandante em todas as partidas. De forma análoga, o modelo *Teimoso Empate* prevê sempre o empate, e o modelo *Teimoso Visitante* prevê sempre a vitória do time visitante.

#### 4.2.9 Modelo *Controle Média*

O modelo *Controle Média* apresenta um funcionamento simples e direto. Ele considera o histórico de gols marcados por cada equipe e assume que a média desses valores será o número de gols que o time fará na partida a ser prevista. Por exemplo, se uma equipe marcou 12 gols nas 10 partidas anteriores, o modelo estima que ela marcará 1,20 gols na 11ª partida.

Para prever o resultado de um confronto, o modelo compara as médias de gols das duas equipes. A equipe com a maior média é considerada a vencedora. Caso as médias sejam iguais, o modelo prevê um empate. As médias são calculadas com duas casas decimais, e esse valor é utilizado diretamente na comparação para determinar o resultado previsto.

#### 4.2.10 Modelo *Controle Uniforme*

O modelo *Controle Uniforme* possui caráter aleatório, pois não utiliza informações históricas para realizar previsões. Sua aleatoriedade está associada à distribuição uniforme, sendo a previsão feita a partir do sorteio de dois números dessa distribuição, conforme detalhado a seguir:

Para determinar as probabilidades dos eventos  $vm$  (vitória do mandante),  $em$  (empate) e  $vv$  (vitória do visitante), sorteiam-se dois números independentes de uma distribuição uniforme no intervalo  $[0,1]$ . Seja  $\alpha$  o menor deles e  $\beta$  o maior. Assim, define-se:

- $P(vm) = \alpha$
- $P(em) = \beta - \alpha$
- $P(vv) = 1 - \beta$

Dessa forma, as três probabilidades são mutuamente exclusivas e somam 1, garantindo uma distribuição válida. Esse modelo serve como um controle aleatório puro, permitindo avaliar o desempenho de modelos mais sofisticados em relação a uma estratégia totalmente aleatória.

#### 4.2.11 Modelo *Controle Poisson*

O modelo *Controle Poisson* é semelhante ao modelo *Controle Uniforme*, pois também realiza previsões sem utilizar informações históricas do banco de dados. No entanto, neste caso, a aleatoriedade está associada à distribuição de Poisson, amplamente utilizada para modelar o número de gols em partidas de futebol, conforme discutido em diversos estudos (ARRUDA, 2000; JUNIOR; GAMERMAN, 2004).

Este modelo estima o placar da partida e, conseqüentemente, prediz o resultado do confronto. Para isso, sorteiam-se dois números independentes de uma distribuição de Poisson com parâmetro  $\lambda = 1$ . O primeiro número representa o número de gols do time mandante, e o segundo, do time visitante. A partir desses valores, determina-se o resultado da partida.

Este modelo serve como um controle aleatório baseado em uma distribuição estatística que reflete a natureza discreta e esparsa dos gols em partidas de futebol.

#### 4.2.12 Modelo *Controle Implícito Média*

O modelo *Implícito Média* segue a metodologia do modelo Implícito proposto por (ARRUDA, 2000). No modelo original, a previsão do placar é obtida por meio de uma matriz de probabilidades resultante do cruzamento de duas distribuições de Poisson, conforme detalhado nos modelos anteriores do mesmo autor. Originalmente, os valores de  $\lambda_X$  e  $\lambda_Y$  são determinados com base em opiniões de especialistas, sem o uso de métodos matemáticos formais, e assume-se  $\lambda_{XY} = 0$ .

A adaptação proposta neste trabalho consiste em utilizar a média do número de gols marcados por cada time como estimativa para os valores de  $\lambda_X$  e  $\lambda_Y$ , mantendo  $\lambda_{XY} = 0$ . Essa abordagem elimina a subjetividade inerente à consulta de especialistas, proporcionando uma metodologia mais sistemática e de fácil reprodução.

### 4.3 RESUMO DOS MODELOS

Cada um dos quatro modelos apresentados por Arruda (2000), juntamente ao modelo de Lima et al. (2012) e suas duas variações, foi subdividido em quatro variantes, considerando diferentes horizontes temporais de memória para as previsões. Essas subdivisões têm como objetivo identificar a janela temporal ideal para cada modelo, além de capturar possíveis mudanças estratégicas das equipes, alterações estruturais no campeonato e a dinâmica de desenvolvimento da competição ao longo do tempo.

As quatro janelas temporais consideradas são:

- **Geral:** Nesta configuração, não há perda de memória; todas as partidas anteriores àquela que será prevista são consideradas, utilizando todo o histórico disponível no banco de dados.
- **Anual:** A cada nova edição anual do campeonato, a memória é reiniciada. Assim, apenas as partidas da edição corrente, desde a primeira até a última anterior ao confronto a ser previsto, são consideradas.
- **Semestral:** A memória é reiniciada a cada semestre do campeonato, aproximando-se das janelas de transferências de jogadores. Dessa forma, apenas as partidas do semestre corrente são utilizadas nas previsões.
- **Rodada:** Esta é a configuração com o menor horizonte de memória, considerando apenas as partidas da rodada imediatamente anterior àquela que está sendo prevista.

Em decorrência das diferentes janelas temporais adotadas, os sete modelos resultam em um total de 28 variantes. A Tabela 3 resume esses modelos e apresenta uma nova nomenclatura para cada um deles, com o intuito de facilitar sua identificação e análise comparativa.

Adicionalmente, para os modelos controle, foi necessário definir o intervalo de tempo nulo, uma vez que alguns deles se baseiam em métodos aleatórios sem qualquer relação com

dados anteriores, ou seja, sem memória. Ressalta-se, portanto, esse tipo específico de janela temporal. A Tabela 3 apresenta de forma sistemática os modelos descritos anteriormente, facilitando sua compreensão. Além disso, introduz novas nomenclaturas que serão utilizadas a partir deste ponto, com o objetivo de simplificar a escrita e a representação gráfica dos modelos.

Tabela 3 – Resumo dos Modelos de Predição

Trabalho de Origem	Método Base	Janela temporal	Nome do modelo
Arruda (2000)	<i>SD 0</i>	Geral	<i>Sd0Ger</i>
		Anual	<i>Sd0Ano</i>
		Semestral	<i>Sd0Sem</i>
		Rodada	<i>Sd0Rod</i>
	<i>SD 1</i>	Geral	<i>Sd1Ger</i>
		Anual	<i>Sd1Ano</i>
		Semestral	<i>Sd1Sem</i>
		Rodada	<i>Sd1Rod</i>
	<i>Chance 1</i>	Geral	<i>Ch1Ger</i>
		Anual	<i>Ch1Ano</i>
		Semestral	<i>Ch1Sem</i>
		Rodada	<i>Ch1Rod</i>
<i>Chance 2</i>	Geral	<i>Ch2Ger</i>	
	Anual	<i>Ch2Ano</i>	
	Semestral	<i>Ch2Sem</i>	
	Rodada	<i>Ch2Rod</i>	
Lima et al. (2012)	<i>UFMG</i>	Geral	<i>UfmgGer</i>
		Anual	<i>UfmgAno</i>
		Semestral	<i>UfmgSem</i>
		Rodada	<i>UfmgRod</i>
	<i>UFMG sem Sorteio</i>	Geral	<i>UfmgSsGer</i>
		Anual	<i>UfmgSsAno</i>
		Semestral	<i>UfmgSsSem</i>
		Rodada	<i>UfmgSsRod</i>
	<i>UFMG Sorteio Normal</i>	Geral	<i>UfmgSnGer</i>
		Anual	<i>UfmgSnAno</i>
		Semestral	<i>UfmgSnSem</i>
		Rodada	<i>UfmgSnRod</i>
Modelos controle	Teimoso Mandante	<i>Nula</i>	<i>ConTeiMan</i>
	Teimoso Empate	<i>Nula</i>	<i>ConTeiEmp</i>
	Teimoso Visitante	<i>Nula</i>	<i>ConTeiVis</i>
	Média	Geral	<i>ConMed</i>
	Distribuição Uniforme	<i>Nula</i>	<i>ConUnif</i>
	Distribuição Poisson	<i>Nula</i>	<i>ConPois</i>
	Método Implícito	Geral	<i>ConImp</i>

Fonte: dos Autores

## 4.4 MÉTRICAS DE AVALIAÇÃO

Uma etapa fundamental na aplicação de modelos preditivos consiste na avaliação de seu desempenho, isto é, na verificação de sua capacidade de produzir previsões corretas e na mensuração da magnitude de seus acertos e erros. Dessa forma, neste estudo, foram utilizadas diversas métricas para a avaliação dos modelos, a saber: taxa de acerto (TA), taxa de funcionamento (TF), nível de complexidade (NC), erro preditivo médio ponderado (EPMP), medida de De Finetti (MD), medida de De Finetti detalhada (MDD), medida de verossimilhança (VERO), Precisão (MC\_Pre), Recall (MC\_Rec), F1-score (MC\_F1) e Coeficiente de Matthews Generalizado (MC\_MCC).

A seguir, são apresentadas as definições das métricas que não foram detalhadas na seção anterior.

### 4.4.1 Taxa de Acerto

Para estabelecer a TA, primeiramente é necessário definir outra métrica, o erro preditivo médio (EPM). Quando se trata de predições e estimação de parâmetros, é bastante comum utilizar o erro preditivo quadrático médio. No entanto, como essa métrica não assume valores negativos, neste estudo considerou-se apenas o erro preditivo médio.

Para observar o comportamento do modelo ao longo das rodadas, e conseqüentemente, o impacto do aumento da amostra, o EPM foi calculado em cada rodada da seguinte maneira:

$$EPM = \frac{1}{n} \sum_{i=1}^n I(\theta_i, \hat{\theta}_i)$$

Em que  $\theta_i$  é o resultado do  $i$ -ésimo jogo (utilizando a escala: "vm" para vitória do mandante, "em" para empate e "vv" para derrota do mandante);  $\hat{\theta}_i$  é o resultado previsto pelo modelo para o jogo  $i$  na mesma escala; e  $I(\cdot, \cdot)$  é uma função indicadora de erro, que assume valor 0 quando  $\theta_i = \hat{\theta}_i$  (acerto) e 1 caso contrário (erro). O EPM, calculado dessa forma, expressa se o modelo acertou ou não e faz a média do quanto o modelo errou naquela rodada.

A TA é uma métrica bastante simples, uma vez compreendido o erro preditivo médio. Como neste estudo se busca identificar o modelo que apresente os melhores resultados, faz mais

sentido classificá-los por uma medida de acerto em vez de uma medida de erro. Por esse motivo, a TA foi definida como:

$$TA = 1 - EPM$$

Além disso, todas as demais métricas mensuram erro. Assim, acredita-se que uma medida que represente diretamente o quanto o modelo acertou desperte maior interesse de quem busca identificar o “melhor” modelo.

#### 4.4.2 Erro Preditivo Médio Ponderado

Ao se avaliar a acurácia de modelos que distribuem probabilidades entre vitória do mandante, empate e vitória do visitante, predizendo que o resultado mais provável é aquele com maior probabilidade entre os três eventos, parece razoável ponderar o grau de desvio em relação ao resultado verdadeiro. Por exemplo, se o modelo 1 atribui 80% de chance de empate a uma partida, e o modelo 2 atribui 55% de chance ao mesmo evento, caso o resultado observado não seja empate, é evidente que o modelo 1 errou mais que o modelo 2. Essa métrica surge com o propósito de quantificar esse tipo de erro, indo além do determinismo do EPM. Assim, define-se:

$$EPM P_j = \frac{1}{n} \sum_{i=1}^n I(\theta_{ij}, \hat{\theta}_{ij}) p_{ij}^*$$

sendo

$$p_{ij}^* = \max(p_{ij}^v, p_{ij}^e, p_{ij}^d)$$

em que  $p_{ij}^v$ ,  $p_{ij}^e$  e  $p_{ij}^d$  representam as probabilidades previstas pelo modelo, respectivamente, de vitória, empate e derrota do mandante no jogo  $i$  da rodada  $j$ ; e  $n$  é o número de partidas da rodada. O restante da fórmula é análogo ao que foi definido para o EPM.

### 4.4.3 Taxa de Funcionamento

Essa métrica foi introduzida para lidar com um possível problema de convergência que pode ocorrer especialmente em modelos que utilizam regressão linear ou, em particular neste estudo, regressão de Poisson Arruda (2000). Situações de não convergência impedem que o modelo gere uma previsão válida, o que comprometeria a análise. Por isso, foi adotada a TF, que mede a proporção de partidas em que o modelo emite uma previsão.

A TF é calculada de forma análoga à TA, contabilizando o número total de previsões efetivamente emitidas, independente de sua exatidão, e então se calcula a média aritmética, conforme a Equação 4.6:

$$TF = \frac{1}{n} \sum_{i=1}^n \theta_i \quad (4.6)$$

Em que  $n$  representa o número de partidas analisadas, e  $\theta_i$  é uma variável indicadora que assume valor 1 quando o modelo emite uma previsão válida, ou seja, declara “vitória visitante (*vv*)”, “empate (*em*)” ou “vitória mandante (*vm*)”, e 0 caso contrário. A TF, portanto, fornece uma medida objetiva da confiabilidade operacional do modelo na prática.

### 4.4.4 Nível de Complexidade

Outra métrica relevante para a comparação entre modelos é o NC, que busca avaliar o grau de sofisticação de cada método em diferentes aspectos, como compreensão didática, conhecimento prévio exigido e esforço computacional. Por se tratar de uma métrica subjetiva, foi desenvolvido um sistema de pesos com o objetivo de quantificar essas características e garantir maior consistência e replicabilidade na avaliação dos métodos.

Para cada modelo, foram consideradas as seguintes características:

- A - **Realização de simulação:** se o modelo realiza simulações e qual o grau de complexidade dessas simulações.
- B - **Presença de distribuição de probabilidade associada à previsão:** se há uma distribuição associada e o quão complexa ela é.

- C - **Uso de regressão linear:** se o modelo utiliza regressão e qual o tipo ou complexidade dessa regressão.
- D - **Natureza algorítmica:** se o algoritmo utilizado envolve múltiplas etapas, recursividade ou se é mais direto e simples.
- E - **Tamanho da janela temporal utilizada:** se o modelo exige grande volume de partidas passadas para realizar as previsões.
- F - **Número de parâmetros:** se o modelo depende de muitos parâmetros e variáveis para suas previsões.
- G - **Quantidade de informação necessária:** se o modelo requer apenas dados básicos (como número de gols e mando de campo) ou informações adicionais (como dados sobre técnicos, cartões, etc.).

Cada uma dessas características foi classificada em aprofundamento, utilizando uma escala Likert de cinco níveis (ROBINSON, 2014), com os seguintes significados:

1. **Elementar:** método simples, que exige pouco ou nenhum conhecimento prévio, ou apenas a presença superficial da característica.
2. **Básico:** requer familiaridade com conceitos estatísticos ou computacionais elementares, sendo relativamente fácil de compreender e implementar.
3. **Mediano:** exige conhecimentos intermediários e apresenta algum grau de complexidade técnica ou computacional.
4. **Refinado:** exige conhecimento avançado da literatura e técnicas estatísticas, além de maior esforço de implementação e análise.
5. **Sofisticado:** apresenta elevado grau de complexidade, exigindo domínio teórico aprofundado, múltiplos parâmetros e alto esforço computacional.

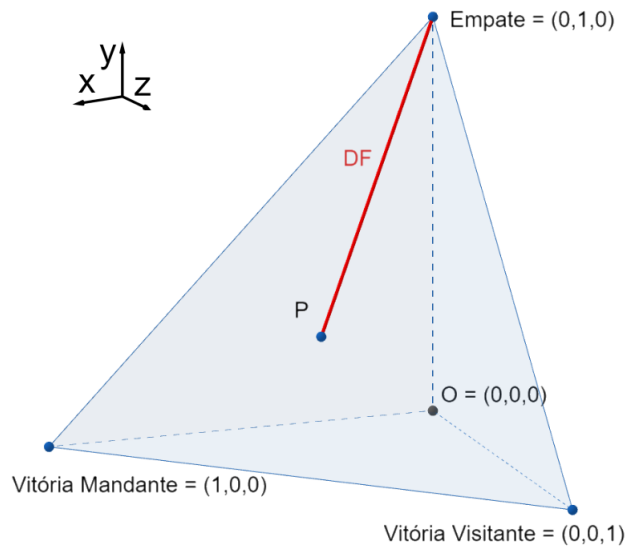
Para determinar o NC de cada modelo, atribuiu-se um valor de aprofundamento entre 1 e 5 a cada uma das características listadas de A a G, seguindo a escala Likert descrita anteriormente. Assim, para cada modelo avaliado, calculou-se a média aritmética simples dos valores atribuídos às sete características. Essa média representa o escore final de complexidade do modelo, permitindo uma comparação objetiva entre diferentes métodos, mesmo diante da natureza

subjetiva dessa métrica. Quanto maior a média obtida, maior o NC associado ao modelo, tanto do ponto de vista teórico quanto prático.

#### 4.4.5 Medida de De Finetti

Outra métrica de acurácia amplamente utilizada para avaliar a qualidade das previsões de modelos como os empregados neste trabalho é a MD. Essa métrica foi originalmente proposta por De Finetti (1972) como uma alternativa para eventos tricotômicos. Ela consiste na distância euclidiana quadrática entre dois pontos em um espaço  $\mathbb{R}^3$ , cujos eixos ortogonais variam de 0 a 1, tendo origem no 0. A distância é calculada do ponto  $P$ , com coordenadas  $(VM, EM, VV)$  (as probabilidades previstas de vitória do mandante, empate e vitória do visitante), até o vértice  $V$  que representa o evento observado, conforme ilustrado na Figura 3.

Figura 3 – Representação gráfica da medida de De Finetti (MD)



Fonte: Dos autores.

O cálculo dessa medida ocorre de três formas, dependendo do evento observado, segundo as seguintes equações:

Se ocorreu vitória do mandante,  $V = (1,0,0)$ :

$$MD = (VM - 1)^2 + (EM - 0)^2 + (VV - 0)^2$$

Se ocorreu empate,  $V = (0,1,0)$ :

$$MD = (VM - 0)^2 + (EM - 1)^2 + (VV - 0)^2$$

Se ocorreu vitória do visitante,  $V = (0,0,1)$ :

$$MD = (VM - 0)^2 + (EM - 0)^2 + (VV - 1)^2$$

Para exemplificar esse cálculo, imagine-se a partida hipotética Flamengo  $\times$  Atlético Mineiro, na qual o modelo atribuiu probabilidades de (0.34, 0.28, 0.38) à vitória do mandante, empate e vitória do visitante, respectivamente. Entretanto, o resultado foi empate, e portanto o valor de MD será igual a 0.7784.

#### 4.4.6 Medida de De Finetti Detalhada

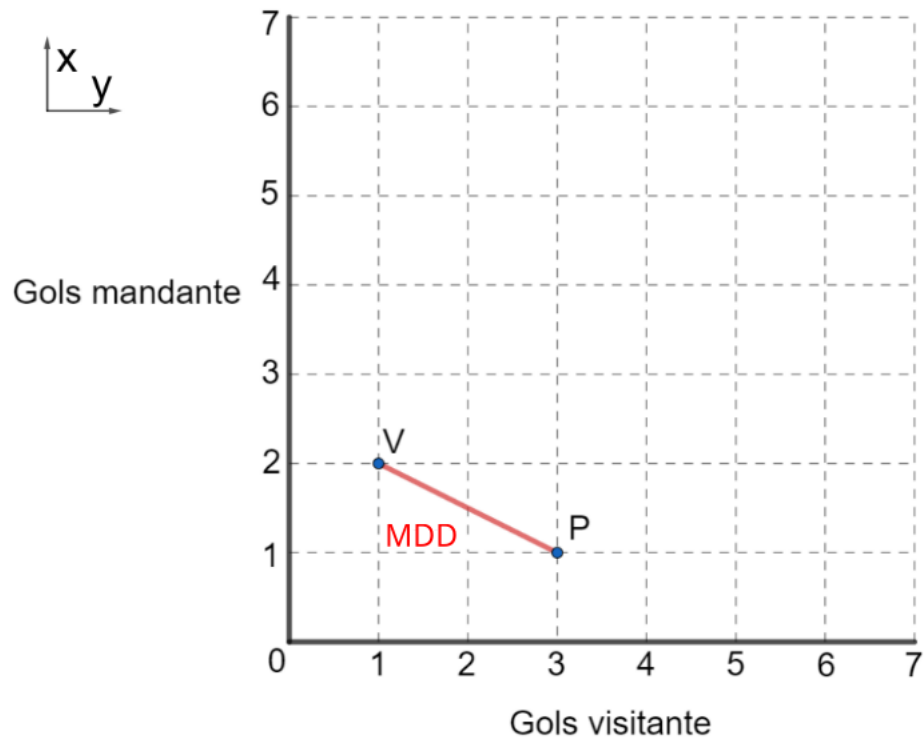
Este trabalho propõe a seguinte métrica, inspirada na medida de De Finetti e nos pressupostos do erro preditivo médio ponderado. Nesse sentido, considere a medida de De Finetti detalhada como a distância euclidiana entre o ponto  $P$  e o ponto  $V$  em um plano cartesiano, sendo que  $P$  tem coordenadas correspondentes ao placar mais provável previsto pelo modelo, e  $V$ , ao placar efetivamente observado, como ilustrado na Figura 4.

O cálculo dessa medida ocorre da seguinte forma:

$$MDD = \sqrt{(gmp - gmo)^2 + (gvp - gvo)^2} \quad (4.7)$$

Em que  $gmp$  e  $gmo$  representam, respectivamente, o número de gols previstos e observados do mandante, enquanto  $gvp$  e  $gvo$  correspondem aos gols previstos e observados do visitante. Diferentemente do exemplo anterior, essa métrica é sensível ao quanto a predição se aproximou do resultado real. Assim, considerando o confronto hipotético Atlético Mineiro  $\times$  Internacional, se o modelo previu o placar de  $1 \times 3$ , mas o placar verdadeiro foi  $2 \times 1$ , a MDD será de 2,236.

Figura 4 – Representação gráfica da medida de De Finetti Detalhada (MDD).



Fonte: Dos autores.

#### 4.4.7 VEROSSIMILHANÇA DO RESULTADO

A VERO é uma métrica estatística que quantifica o quão compatíveis são os resultados observados com os parâmetros assumidos pelo modelo. No contexto dos modelos que utilizam distribuições de probabilidade, considera-se que os placares do mandante e do visitante  $(X, Y)$  seguem distribuições de Poisson com médias  $\lambda_X$  e  $\lambda_Y$ , respectivamente. Assim, a VERO de um resultado observado  $(x, y)$  é dada por:

$$\mathcal{L}(x, y | \lambda_1, \lambda_2) = P(X = x | \lambda_1) P(Y = y | \lambda_2),$$

sendo calculada como o produto de duas probabilidades de Poisson. A função implementada retorna, para um conjunto de partidas, a verossimilhança média dos resultados observados sob os parâmetros previstos pelo modelo.

A interpretação dessa métrica baseia-se na ideia de compatibilidade entre modelo e realidade. Valores altos de VERO indicam que o modelo atribui probabilidades relativamente grandes aos resultados observados, sugerindo boa aderência entre previsão e realidade. Já valo-

res muito baixos sinalizam que os parâmetros do modelo geram probabilidades pequenas para os placares observados.

#### 4.4.8 Matriz de Confusão e Métricas de Desempenho

A matriz de confusão é uma ferramenta muito útil para a avaliação de modelos de classificação, como os expostos neste estudo, que determinam um dos três resultados possíveis para cada partida. Ela compara as classes previstas pelo modelo com as classes reais observadas, permitindo verificar onde o modelo acerta e onde comete erros. Em sua forma mais reduzida, a matriz é utilizada para problemas binários; entretanto, ela pode ser expandida de forma natural para problemas com eventos tricotômicos.

Considere um problema de classificação com três classes distintas:  $C_1$ ,  $C_2$  e  $C_3$ . A matriz de confusão tricotômica pode ser representada como:

Real\Prevista	$C_1$	$C_2$	$C_3$
$C_1$	$n_{11}$	$n_{12}$	$n_{13}$
$C_2$	$n_{21}$	$n_{22}$	$n_{23}$
$C_3$	$n_{31}$	$n_{32}$	$n_{33}$

Cada elemento  $n_{ij}$  representa o número de observações cuja classe real é  $C_i$  e que foram classificadas pelo modelo como pertencentes à classe  $C_j$ . Dessa forma, os elementos da diagonal principal correspondem às classificações corretas, enquanto os elementos fora da diagonal representam erros de classificação. Por exemplo, um valor elevado de  $n_{12}$  indica que o modelo apresenta dificuldade em distinguir a classe  $C_1$  da classe  $C_2$ , confundindo-as com frequência.

Além de sua utilidade para análise visual do desempenho do classificador, a matriz de confusão constitui a base para o cálculo de diversas métricas quantitativas. Para cada classe, podem ser obtidas medidas como Precisão, Recall, F1-score e o Coeficiente de Matthews Generalizado. Ressalta-se que, inicialmente, essas métricas são apresentadas de forma individual, por classe. Entretanto, quando retomadas na seção de resultados, elas se referem às métricas globais do modelo, calculadas a partir da média das métricas obtidas para todas as classes.

#### 4.4.8.1 Precisão

A Precisão da classe  $C_i$  é definida por:

$$\text{Precisão}(C_i) = \frac{n_{ii}}{\sum_{j=1}^3 n_{ji}}$$

Ela indica a proporção de previsões da classe  $C_i$  que estavam corretas, ou seja, é a razão entre o número de previsões corretas de uma classe e o total de previsões feitas.

#### 4.4.8.2 Recall ou Sensibilidade

O recall da classe  $C_i$  é definido por:

$$\text{Recall}(C_i) = \frac{n_{ii}}{\sum_{j=1}^3 n_{ij}}$$

O recall, mede a capacidade do modelo de identificar corretamente as observações que realmente pertencem a uma classe  $C_i$ . Assim, o recall indica a proporção de casos reais de  $C_i$  que o modelo conseguiu captar. Valores altos significam que o modelo comete poucos falsos negativos, reconhecendo a maior parte dos elementos da classe, enquanto valores baixos indicam que o modelo frequentemente deixa de identificar observações que pertencem a  $C_i$ .

#### 4.4.8.3 F1-Score

Para cada classe é dado pela média harmônica entre precisão e recall:

$$F1(C_i) = 2 \cdot \frac{\text{Precisão}(C_i) \cdot \text{Recall}(C_i)}{\text{Precisão}(C_i) + \text{Recall}(C_i)}$$

O F1-score combina precisão e recall em uma única medida, utilizando a média harmônica entre elas para equilibrar os dois aspectos do desempenho do modelo. Essa métrica é especialmente útil quando há desequilíbrio entre as classes, pois penaliza situações em que uma das duas medidas é muito baixa: para que o F1-score seja alto, o modelo precisa simultaneamente identificar corretamente a maior parte dos elementos reais da classe (alto recall) e evitar atribuições incorretas a essa classe (alta precisão). Dessa forma, o F1-score fornece uma avaliação mais robusta do desempenho por classe do que usar apenas precisão ou recall isoladamente.

#### 4.4.8.4 Coeficiente de Matthews Generalizado

O MC\_MCC, originalmente definido para classificação binária, pode ser estendido para múltiplas classes. Para três classes, ele pode ser calculado como:

$$\text{MC\_MCC} = \frac{\sum_{k=1}^3 \sum_{l=1}^3 \sum_{m=1}^3 n_{kk}n_{lm} - n_{kl}n_{mk}}{\sqrt{\left(\sum_{k=1}^3 \left(\sum_{l=1}^3 n_{kl}\right) \left(\sum_{m \neq k} \sum_{l=1}^3 n_{ml}\right)\right) \left(\sum_{k=1}^3 \left(\sum_{l=1}^3 n_{lk}\right) \left(\sum_{m \neq k} \sum_{l=1}^3 n_{lm}\right)\right)}}$$

Esse coeficiente varia de  $-1$  a  $+1$ , sendo:

- $+1$  desempenho perfeito;
- $0$  equivalente a uma classificação aleatória;
- $-1$  desempenho totalmente inverso.

A principal vantagem do MC\_MCC em relação a outras métricas é sua robustez, especialmente quando há desequilíbrio entre as classes — uma situação comum em dados reais. Enquanto métricas tradicionais podem inflar a percepção de desempenho quando a classe majoritária domina, o MC\_MCC penaliza proporcionalmente erros distribuídos de forma desigual entre classes. Assim, ele oferece uma medida equilibrada e confiável da capacidade discriminativa do modelo em todo o espaço de classes, tornando-se uma das métricas mais indicadas para avaliar classificadores multiclasse de maneira justa e independente das proporções amostrais.

## 4.5 ESTUDO DE SIMULAÇÃO

Como este trabalho propõe diferentes tipos de janelas temporais para cada modelo, surge naturalmente o questionamento sobre qual seria o intervalo de tempo ideal para cada um deles. Em termos práticos, a memória ideal corresponde ao número de rodadas anteriores que proporciona o melhor desempenho preditivo do modelo, equilibrando a quantidade de informação utilizada com a capacidade de generalização.

Foram analisados os seguintes modelos: *Sd0Rod*, *Sd1Rod*, *Ch1Rod*, *Ch2Rod*, *UfmgRod*, *UfmgSsRod* e *UfmgSnRod*.

Cada um desses modelos, em sua forma original, utiliza apenas as partidas da rodada imediatamente anterior àquela que se deseja prever. Para o estudo de simulação, os algoritmos foram modificados a fim de considerar  $n$  rodadas anteriores como base para a previsão. De forma arbitrária, o valor de  $n$  foi limitado ao intervalo  $(1, 114)$ , levando em conta o custo computacional envolvido na simulação. Esse intervalo permite ao modelo considerar desde apenas uma rodada prévia até três anos de competições, visto que cada campeonato contém 38 rodadas ( $38 \times 3 = 114$ ). Como conjunto de dados, foram utilizados os campeonatos de 2021, 2022, 2023 e 2024.

Para avaliar o desempenho, consideraram-se exclusivamente as previsões referentes ao campeonato de 2024. Essa escolha garante que todas as previsões sejam realizadas com a mesma quantidade de informações disponíveis, permitindo uma comparação justa entre as diferentes quantias de rodadas utilizadas. Para comparar o desempenho dos modelos, foram selecionadas cinco métricas amplamente utilizadas na literatura: Taxa de Acerto, Medida De Finetti, Precisão, Recall, F1-Score e Coeficiente de Correlação de Matthews. A escolha dessas métricas deve-se à sua popularidade em estudos anteriores, o que possibilita comparações com outros trabalhos e contribui para uma discussão mais aprofundada sobre os resultados.

## 5 RESULTADOS

Nesta seção, apresentam-se os resultados obtidos a partir da aplicação dos modelos e da análise do banco de dados. Inicialmente, são discutidos aspectos descritivos dos campeonatos observados e das equipes participantes, de modo a fornecer um panorama geral dos dados. Essa etapa introdutória tem como objetivo contextualizar o cenário competitivo, além de fomentar a rivalidade entre torcedores e leitores, servindo também como base informativa para futuros estudos.

Em seguida, analisam-se as métricas de desempenho alcançadas por cada modelo, com destaque para suas principais contribuições, limitações e implicações práticas. Por fim, são apresentados os resultados de um estudo voltado à determinação do valor de memória ideal para alguns modelos selecionados, buscando otimizar tanto a precisão quanto a eficiência das previsões.

### 5.1 ANÁLISE EXPLORATÓRIO DO CAMPEONATO BRASILEIRO

Durante os 22 anos de competição e 8785 partidas analisadas, o número padrão de times participantes por edição mudou duas vezes. Essa alteração impactou diretamente o número de rodadas, jogos e gols, conforme pode ser observado na Tabela 4, onde também estão destacados os times que foram campeões de cada edição. Além disso, soma-se um total de 45 times que participaram de pelo menos uma dessas edições. A seguir estão destacados esses times e suas aparições nesse intervalo de competições; estes e outros resultados estão presentes na Tabela 6 apresentada no Apêndice A.

Iniciando pelos times com menor número de participações, destacam-se Barueri, Grêmio Prudente, América-RN, CSA, Ipatinga, Joinville, Santo André e Brasiense, todos com uma participação registrada. Com menos de cinco participações, encontram-se Santa Cruz, Guarani, Paysandu, Portuguesa, Cuiabá e São Caetano. Esses clubes tiveram presença limitada nas competições, sendo mais influentes em outras divisões do campeonato.

Em seguida, os times com menos de dez participações são Bragantino, Criciúma, Náutico, América-MG, Paraná, Avaí, Ceará, Chapecoense, Atlético-GO, Juventude, Fortaleza e Ponte Preta. Esses times demonstram presença intermitente, figurando nas competições ao

longo do tempo, embora em número reduzido.

Clubes como Figueirense, Sport, Vitória, Bahia, Coritiba, Goiás, Vasco, Botafogo-RJ e Cruzeiro marcam uma fase intermediária, com participações entre 10 e 20 vezes, refletindo equipes que competiram regularmente e apresentaram boas campanhas, mas sem manter uma presença contínua ao longo dos anos.

Tabela 4 – Número de jogos, times, rodadas, gols e time campeão de cada ano analisado.

<b>Ano</b>	<b>Nº de jogos</b>	<b>Nº de times</b>	<b>Nº de rodadas</b>	<b>Nº de gols</b>	<b>Campeão</b>
2003	552	24	46	1592	Cruzeiro
2004	552	24	46	1534	Santos
2005	462	22	42	1451	Corinthians
2006	380	20	38	1030	São Paulo
2007	380	20	38	1047	São Paulo
2008	380	20	38	1035	São Paulo
2009	380	20	38	1094	Flamengo
2010	380	20	38	978	Fluminense
2011	380	20	38	1017	Corinthians
2012	380	20	38	940	Fluminense
2013	380	20	38	936	Cruzeiro
2014	380	20	38	860	Cruzeiro
2015	380	20	38	897	Corinthians
2016	379	20	38	912	Palmeiras
2017	380	20	38	923	Corinthians
2018	380	20	38	827	Palmeiras
2019	380	20	38	876	Flamengo
2020	380	20	38	944	Flamengo
2021	380	20	38	842	Atlético MG
2022	380	20	38	905	Palmeiras
2023	380	20	38	946	Palmeiras
2024	380	22	38	929	Botafogo

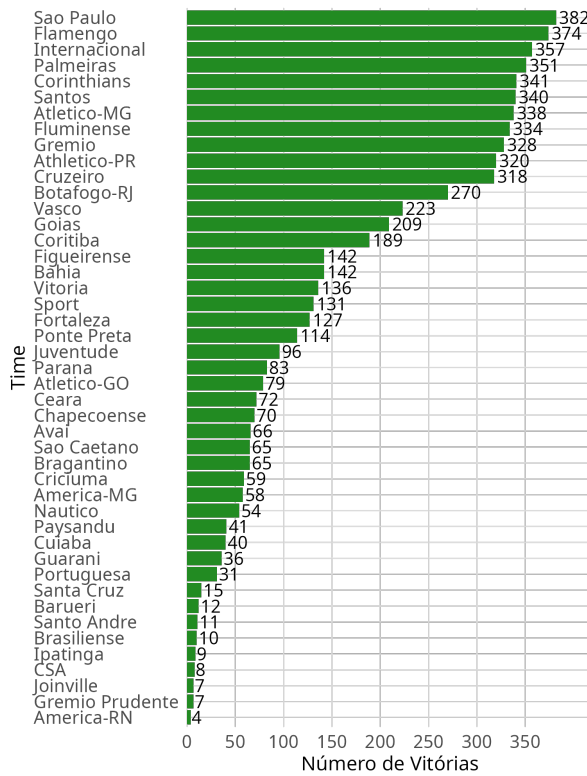
Fonte: Dos Autores

Por fim, os times com as maiores participações são Grêmio e Palmeiras, com 20 participações cada, seguidos por Athletico Paranaense, Atlético-MG, Corinthians, Internacional e Santos, todos com um rebaixamento a menos. Já São Paulo, Fluminense e Flamengo, com 22 participações cada, são os clubes com maior presença nas competições, destacando-se nacionalmente e mantendo-se na elite do futebol brasileiro desde 2003, sem rebaixamentos nesse período. Esses clubes possuem um legado histórico e uma frequência altíssima de participação em diversas edições do campeonato, solidificando suas posições no cenário do futebol nacional e em competições internacionais como consequência.

Ainda analisando valores absolutos, com o intuito de obter uma representação histó-

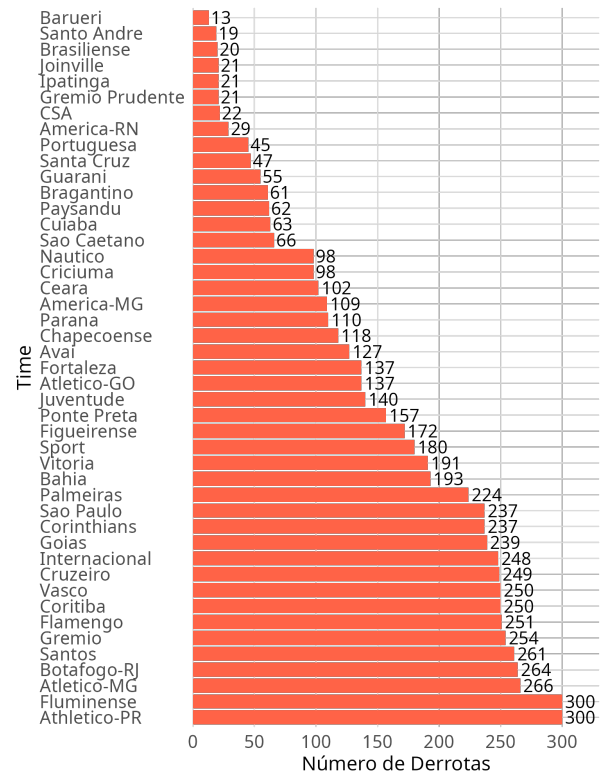
rica do desempenho dos times nesta competição, observam-se o número de vitórias, empates e derrotas de cada equipe, conforme apresentado nas Figuras 5, 6 e 7.

Figura 5 – Contagem de vitórias por time.



Fonte: Dos Autores

Figura 6 – Contagem de derrotas por time.



Fonte: Dos Autores

A análise do desempenho histórico dos clubes no Campeonato Brasileiro pode ser aprofundada a partir da observação conjunta das Figuras 5, 6 e 7, que apresentam, respectivamente, a contagem total de vitórias, derrotas e empates por time ao longo da história da competição. Esses dados fornecem uma visão ampla e comparativa sobre a participação e o rendimento dos clubes, permitindo identificar padrões de regularidade, consistência e competitividade.

A Figura 5, que mostra a contagem de vitórias, evidencia o domínio de clubes tradicionais com forte presença na elite do futebol nacional. O São Paulo lidera a lista com 382 vitórias, seguido por Flamengo (374), Internacional (357), Palmeiras (351) e Corinthians (341). Esses números refletem a regularidade dessas equipes ao longo dos anos e sua capacidade de manter desempenho competitivo em diversas edições. Logo atrás, encontram-se clubes como Atlético-MG (338), Fluminense (330) e Grêmio (324), também com histórico sólido no torneio.

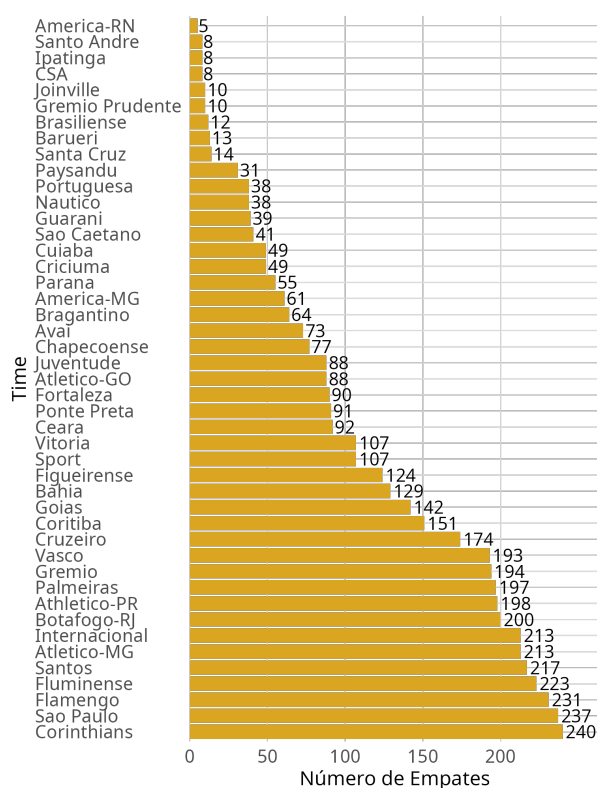
Na faixa intermediária, times como Botafogo-RJ, Vasco, Goiás, Coritiba, Figueirense e Vitória apresentam desempenho relevante, mesmo que com menor número de vitórias. Por fim,

na parte inferior do gráfico, identificam-se equipes com participações pontuais, como Grêmio Prudente e América-RN, ambas com apenas 7 vitórias, o que sugere campanhas limitadas em poucas edições do campeonato.

A Figura 6 , por sua vez, apresenta a contagem de derrotas e oferece uma perspectiva complementar à análise anterior. Aqui, o Athletico-PR aparece com o maior número de derrotas (266), seguido de Fluminense (264), Atlético-MG (260), Botafogo-RJ (256) e Grêmio (251). Embora esses números possam inicialmente sugerir baixo desempenho, é importante considerar que muitas dessas equipes possuem extensa trajetória na competição, o que implica, naturalmente, em maior número absoluto de partidas e, portanto, de derrotas.

Na faixa intermediária, clubes como Sport, Vitória, Figueirense, Ponte Preta e Juventude acumulam entre 150 e 200 derrotas, o que revela certa instabilidade ou desempenho irregular ao longo dos anos. Já na base do gráfico, encontram-se times com campanhas breves, como Barueri (13 derrotas), Santo André (19) e Brasiense (20), o que se alinha a participações muito pontuais, muitas vezes limitadas a apenas uma ou duas temporadas.

Figura 7 – Contagem de empates por time.



Fonte: Dos Autores

A Figura 7, que aborda o número de empates, reforça a influência do tempo de perma-

nência na competição sobre as estatísticas acumuladas. O Corinthians lidera com 240 empates, seguido por São Paulo (237), Fluminense (231), Atlético-MG (233) e Internacional (223). Esses valores reforçam a tese de que a longevidade na elite tende a elevar os totais em todas as categorias (vitórias, derrotas e empates), sendo necessária cautela na análise comparativa entre clubes com trajetórias tão distintas.

Na parte intermediária do gráfico, observa-se uma ampla concentração de clubes com números entre 80 e 150 empates, o que inclui equipes como Figueirense, Vitória, Goiás, Cruzeiro, Grêmio e Palmeiras. Esse padrão sugere certa estabilidade competitiva, com campanhas que mesclaram vitórias, derrotas e resultados equilibrados. Na base do gráfico, novamente se destacam clubes com presenças fugazes no torneio, como América-RN (5 empates), Santo André (8) e Ipatinga (8), cujos baixos totais resultam da baixa quantidade de jogos disputados.

Ao cruzar os dados das três figuras, torna-se possível traçar perfis distintos dos clubes. Alguns se destacam pelo alto número de vitórias e relativamente poucas derrotas, como é o caso do São Paulo e do Palmeiras, sugerindo campanhas historicamente eficientes. Outros apresentam números elevados em todas as categorias, como Fluminense, Atlético-MG e Grêmio, refletindo longa permanência no campeonato, ainda que nem sempre com desempenho de elite. Já clubes com poucos jogos somam números baixos em todas as métricas, o que limita inferências mais profundas.

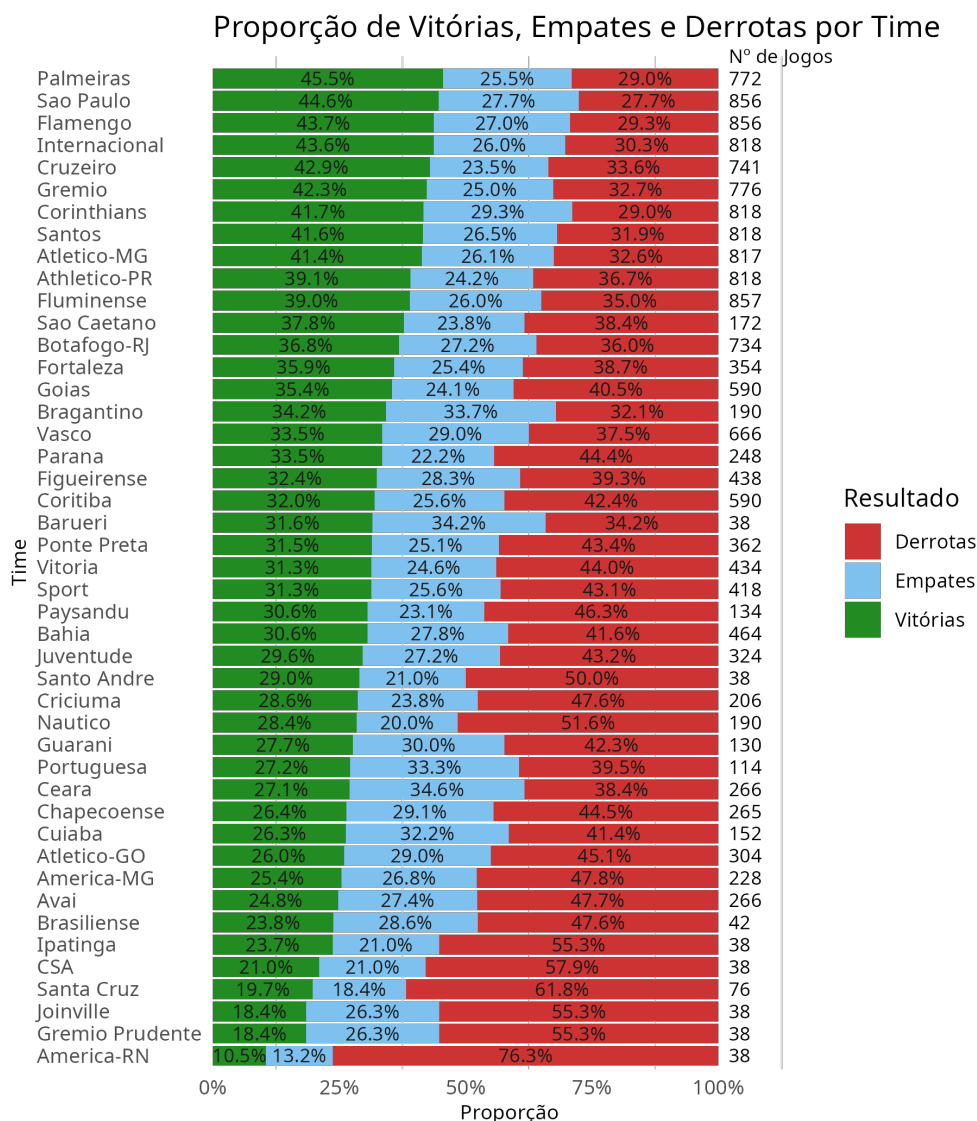
Portanto, a análise integrada das vitórias, derrotas e empates oferece um panorama robusto da trajetória dos clubes no Campeonato Brasileiro, destacando tanto a quantidade de participações quanto o nível de desempenho esportivo. Essa leitura é essencial para compreender a evolução histórica da competição e as dinâmicas que marcam o futebol brasileiro ao longo das décadas.

Após a análise absoluta do desempenho dos clubes em termos de vitórias, derrotas e empates, é fundamental observar os dados sob uma ótica proporcional, ou seja, ponderando os resultados pelo número total de jogos que cada time disputou no Campeonato Brasileiro. Essa abordagem, ilustrada na Figura 8, proporciona uma comparação mais justa e equilibrada, já que corrige as distorções causadas por clubes que participaram de mais edições ou jogaram mais partidas. Dessa forma, é possível avaliar a eficiência histórica de cada equipe, independentemente do tempo de permanência na competição.

A Figura 8 apresenta um gráfico composicional das proporções de vitórias (em verde), empates (amarelo) e derrotas (vermelho) para cada time. A ordenação das equipes foi feita com

base na proporção de vitórias, permitindo identificar de forma clara quais clubes apresentam melhor aproveitamento histórico.

Figura 8 – Gráfico composicional das vitórias, derrotas e empates ponderados pelo número de jogos que cada time participou.



Fonte: Dos Autores

No topo do gráfico, observa-se que os clubes com os melhores desempenhos proporcionais são Palmeiras (41,6% de vitórias), São Paulo (41,4%) e Flamengo (41,3%), seguidos de perto por Internacional, Cruzeiro e Grêmio, todos com mais de 40% de vitórias em relação ao total de jogos disputados. Esses dados reafirmam a regularidade e a competitividade histórica desses clubes, que mantêm altos índices de sucesso, mesmo quando descontadas as diferenças no número de participações.

Outros clubes tradicionais, como Corinthians, Santos, Atlético-MG e Athletico-PR, também aparecem com bons desempenhos proporcionais, com índices de vitória na casa dos 39% a 41%, mantendo-se competitivos historicamente. Vale destacar também a performance relativamente elevada de clubes como São Caetano, que apesar de poucas participações, apresenta 23,8% de vitórias, índice compatível com clubes de maior tradição.

Na parte inferior do gráfico, aparecem equipes com desempenho historicamente mais modesto, como Santa Cruz (18,4%), CSA (21%), Ipatinga (21%), América-RN (13,2%) e Grêmio Prudente (26,3%). Esses baixos índices sugerem campanhas de baixo aproveitamento nas poucas participações que tiveram, o que também se reflete no alto percentual de derrotas.

Chama atenção o fato de que alguns clubes, mesmo com presença constante, apresentam alta proporção de empates, como Portuguesa (33,3%), Ceará (34,4%) e Barueri (34,2%). Já outros, como América-RN, têm grande concentração de derrotas (mais de 70%), revelando grande dificuldade competitiva nas edições em que participaram.

De forma geral, a análise proporcional oferece uma visão mais refinada e coerente do desempenho dos clubes, eliminando o viés do tempo de participação e destacando a eficiência esportiva de cada equipe. Enquanto os gráficos anteriores evidenciaram a quantidade bruta de resultados, a Figura 8 permite comparar qualitativamente os desempenhos, revelando, por exemplo, que alguns clubes com menos vitórias absolutas possuem, proporcionalmente, aproveitamento superior a outros com mais participações.

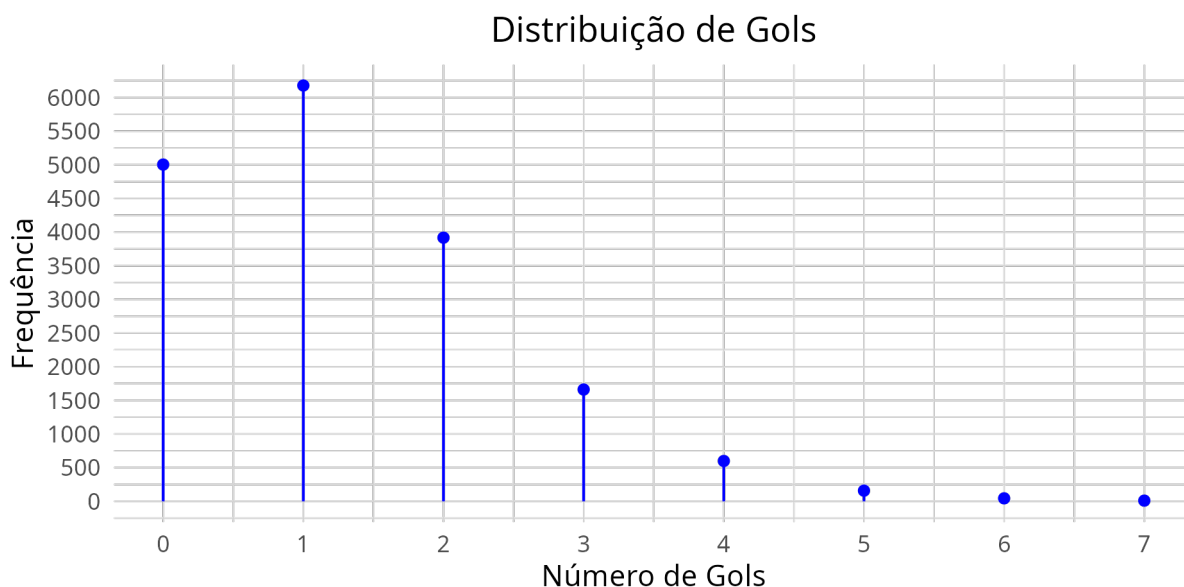
Quanto ao número de gols nesses 22 anos de campeonato analisados, foram marcados 22.515 gols, com uma média geral de 1,28 gols por partida e um desvio padrão de 1,16, indicando certa variabilidade nos placares. Quando analisados separadamente, os times mandantes apresentaram média de 1,54 gols por jogo, com desvio padrão de 1,22, enquanto os visitantes marcaram em média 1,03 gols, com desvio padrão de 1,03. Esses dados reforçam a existência da vantagem do mando de campo, comum em competições desse porte.

A distribuição dos gols por partida revela que o placar mais frequente foi de 1 gol, com 6.178 ocorrências, seguido por jogos sem gols (0), que ocorreram 5.003 vezes. Partidas com 2 gols foram registradas em 3.917 oportunidades, e a frequência cai progressivamente à medida que o número de gols aumenta. Placar com 4 ou mais gols é relativamente raro, com apenas 10 partidas registrando 7 gols.

Esse padrão evidencia uma tendência de resultados mais equilibrados e com placares baixos, compatível com o perfil competitivo do campeonato ao longo dos anos. A distribuição

completa dos gols pode ser visualizada de forma clara por meio de um gráfico de agulhas, que ilustra essa concentração em resultados de 0 a 2 gols.

Figura 9 – Gráfico da frequência de gols observados.



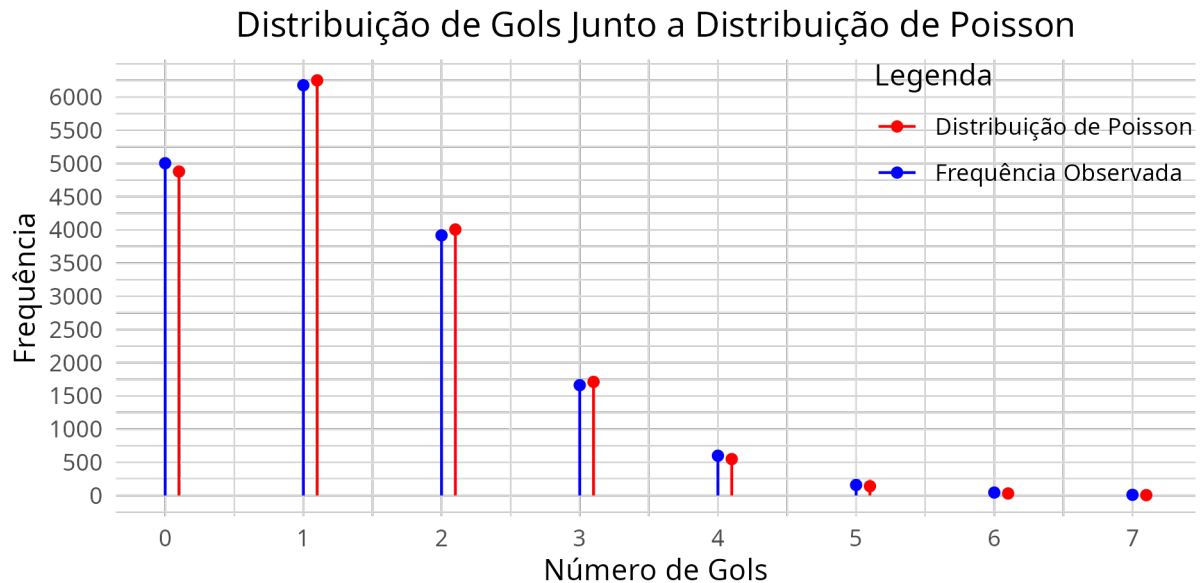
Fonte: Dos autores.

Em relação à frequência de gols apresentada na Figura 9, um aspecto importante que chama a atenção, tanto de pesquisadores quanto de desenvolvedores de modelos, é o formato da distribuição observada. Essa visualização suscita questões relevantes, como: é possível modelar esses dados e descrever essa distribuição de forma analítica?

Diversos trabalhos na literatura utilizam a distribuição de Poisson com esse objetivo. Nesse contexto, a Figura 10 apresenta, lado a lado, a distribuição empírica de gols observados e a distribuição de Poisson simulada por computador, com parâmetro  $\lambda = 1,28$ , correspondente à média dos gols registrados.

Com o objetivo de verificar se a distribuição do número de gols por partida poderia ser modelada por uma distribuição de Poisson, foi aplicado o teste do qui-quadrado de aderência. As frequências observadas foram comparadas com as frequências teóricas esperadas de uma distribuição de Poisson com parâmetro  $\lambda$  estimado pela média amostral dos dados. Para garantir a validade do teste, as classes com baixa frequência esperada (gols igual ou superior a 5) foram agrupadas em uma única categoria, conforme recomendação de que todas as frequências esperadas devem ser, no mínimo, iguais a 5.

Figura 10 – Gráfico da frequência de gols observados comparada à distribuição de Poisson com  $\lambda = 1,28$ .



Fonte: Dos autores.

O valor do estatístico do teste foi  $\chi^2 = 4,03$ , com 4 graus de liberdade (devido à estimação de um parâmetro) e valor-p igual a  $p = 0,5457$ . Como o valor-p é superior ao nível de significância usual ( $\alpha = 0,05$ ), não há evidências estatísticas para rejeitar a hipótese nula de que os dados seguem uma distribuição de Poisson. Portanto, os resultados indicam que a distribuição do número de gols por partida pode ser adequadamente modelada por uma distribuição de Poisson.

## 5.2 MODELOS E SUAS MÉTRICAS

Nesta seção, são apresentadas as análises provenientes da aplicação dos modelos detalhados anteriormente na Seção 4.2, voltadas à previsão das partidas constantes no banco de dados. Inicialmente, é válido observar as estatísticas de cada uma das métricas de avaliação, a fim de construir uma base comparativa para os modelos que serão analisados a seguir. Para isso, a Tabela 5 apresenta um resumo estatístico das métricas utilizadas na avaliação dos modelos: Taxa de Acerto (TA), Erro Preditivo Médio Ponderado (EPMP), Medida de De Finetti (MD), Medida de De Finetti Detalhada (MDD), Taxa de Funcionamento (TF), Nível de Complexidade (NC), Precisão de acordo com a Matriz de Confusão (MC\_Pre), Recall (MC\_Rec), F1-score

(MC\_F1), Coeficiente de Correlação de Matthews (MC\_MCC) e Verossimilhança do Resultado (VERO). (Para uma versão expandida da Tabela 5 vide a Tabela 8 presente no Apêndice B.)

Tabela 5 – Estatísticas das métricas dos modelos

Métrica	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desv. Pad.	NA's
TA	0,1692	0,3719	0,4046	0,3934	0,4425	0,4963	0,0814	0
EPMP	0,2017	0,2484	0,2849	0,3173	0,3309	0,7605	0,1273	2
MD	0,6349	0,6467	0,6924	0,7726	0,7514	1,5210	0,2221	2
MDD	1,439	1,551	1,712	1,836	1,883	2,806	0,4093	16
TF	0,5375	1,0000	1,0000	0,9711	1,0000	1,0000	0,1066	0
NC	1,000	2,929	3,429	3,143	3,714	4,143	0,9154	0
MC_Pre	0,2606	0,3361	0,3520	0,3510	0,3596	0,4379	0,0299	3
MC_Rec	0,1815	0,3348	0,3407	0,3339	0,3540	0,3657	0,0409	3
MC_F1	0,2119	0,2955	0,3182	0,3068	0,3252	0,3651	0,0335	3
MC_MCC	-0,0099	0,0065	0,0167	0,0213	0,0434	0,0559	0,0193	3
Vero	0,0401	0,0677	0,0690	0,0640	0,0710	0,0742	0,0124	18

Fonte: Dos Autores

A TA apresentou valores entre 0,1692 e 0,4963, com média de 0,3934. A mediana foi de 0,4046, o que sugere que metade dos modelos superou esse valor. O primeiro e o terceiro quartis situaram-se em 0,3719 e 0,4425, respectivamente. Dessa forma, conclui-se que os modelos mantiveram, em geral, um desempenho em torno de 40% de acerto, com baixa variabilidade, dado que o desvio-padrão não alcançou 10%.

O EPMP variou de 0,2017 a 0,7605, com média de 0,3173, evidenciando uma variação mais ampla na precisão preditiva entre os modelos — sendo que valores menores são desejáveis. Apenas dois modelos, *ConPois* e *ConMed*, apresentaram dados ausentes (NA) para essa métrica, o que se deve à natureza dessas metodologias, que não fornecem probabilidades de ocorrência dos resultados, mas sim previsões diretas de placares.

A MD, associada à calibração das previsões, oscilou entre 0,6349 e 1,5210, com média de 0,7726. A mediana de 0,6924 indica que a maior parte dos modelos apresentou calibração inferior a esse valor. Da mesma forma que no EPMP, não foi possível calcular essa métrica para os modelos baseados apenas em previsão de placares, por não disponibilizarem as probabilidades necessárias ao cálculo.

A MDD, calculada apenas para um subconjunto de modelos, variou de 1,439 a 2,806,

com média de 1,836 e mediana de 1,712, indicando leve assimetria à direita. Essa métrica não pôde ser estimada em 16 modelos que não apresentam previsões de placares, mas apenas probabilidades dos resultados da partida (vitória, empate ou derrota). Esses modelos incluem as versões baseadas no modelo *UFMG*, bem como os controles Teimosos e Uniforme.

A TF foi de 100% em todos os modelos, com exceção daqueles baseados nos métodos Chance I e Chance II. Nesses casos, destaca-se que esses modelos são baseados em um processo de regressão linear; dessa forma, é possível que apresentem problemas de convergência, impossibilitando a realização das previsões.

Por fim, o NC variou entre 1,000 e 4,143, com média de 3,143. A distribuição dessa métrica, com os quartis posicionados em 2,929 (Q1) e 3,714 (Q3), revela uma concentração de modelos entre os níveis 3 e 4 da escala definida. Apesar de essa métrica adotar critérios arbitrários, ela cumpre seu papel de associar um parâmetro de parcimônia aos modelos, de forma a se aproximar da ideia de uma normal. As demais métricas serão melhor detalhadas de forma isolada a seguir.

### 5.2.1 Análise das Métricas Isoladas

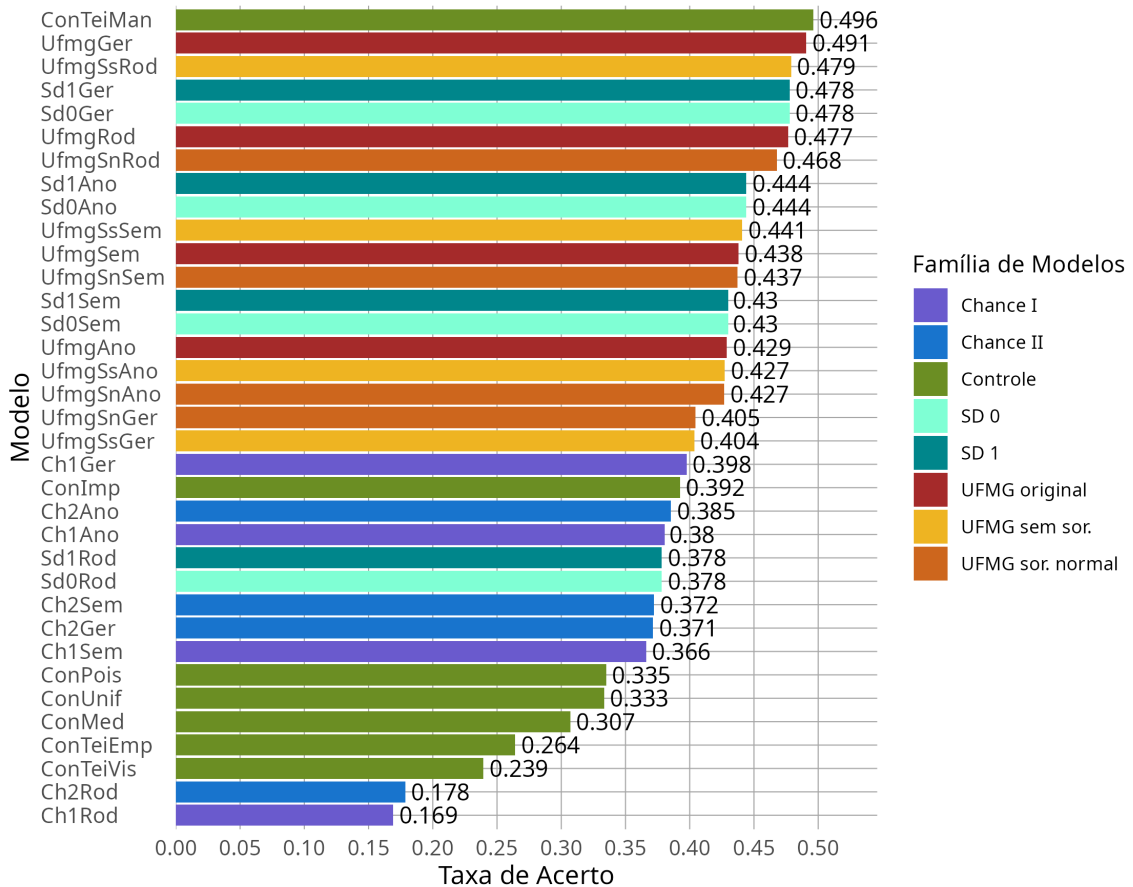
De forma categórica, analisou-se o desempenho de cada modelo individualmente, considerando suas métricas de forma isolada. A Figura 11 apresenta a TA obtida por cada um dos modelos aplicados à previsão dos resultados das partidas. Observa-se uma ampla variação de desempenho, com valores entre 0,169 e 0,496. O modelo *ConTeiMan* apresentou a melhor TA, alcançando 0,496, o que significa que quase metade de suas previsões foi correta. Em seguida, destacam-se os modelos *UfmgGer* (0,491), *UfmgSsRod* (0,479), *Sd1Ger* (0,478) e *Sd0Ger* (0,478), todos com desempenho superior a 47% de acertos e, portanto, entre os mais eficazes na tarefa.

Outros estudos também registram desempenhos relevantes com base na acurácia, permitindo contextualizar os resultados obtidos. Por exemplo, Wong et al. (2025) alcançou 62,3% de acurácia utilizando o modelo XGBoost aplicado ao campeonato *Premier League* de 2022, enquanto Ramos, Fernandes e Batista (2021) registrou 52% de acurácia com seu modelo aplicado ao “Brasileirão Série A” de 2018.

Um segundo grupo de modelos apresentou desempenho intermediário, com TA’s situ-

adas entre 0,40 e 0,47. Nesse intervalo encontram-se, por exemplo, os modelos *UfmgSnRod* (0,466), *Sd1Ano* e *Sd0Ano* (ambos com 0,444), além dos modelos *UfmgSsSem* e *UfmgSem*, ambos próximos de 0,441. Observa-se que diversas variantes das famílias *UFMG* e *SD* mantiveram desempenho consistente nessa faixa.

Figura 11 – Taxa de acerto dos modelos.

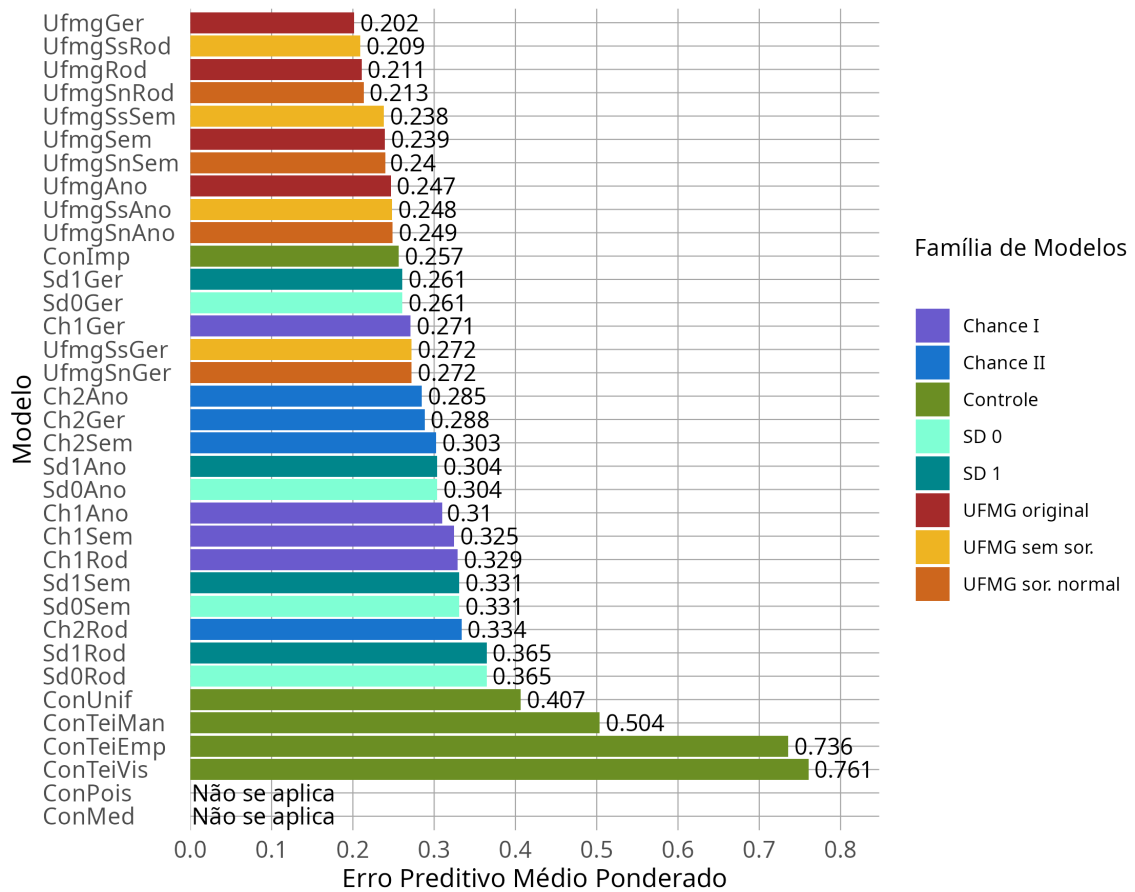


Fonte: Dos Autores

Por outro lado, alguns modelos apresentaram taxas de acerto inferiores à média geral (aproximadamente 0,393). Entre eles destacam-se *Ch1Ger* (0,398), *ConImp* (0,392), *Ch2Ano* (0,385), *Ch1Ano* (0,380), *ConTeiVis* (0,239), *Ch2Rod* (0,178) e *Ch1Rod* (0,169), todos com TA próximas de 30%.

Para o Erro Preditivo Médio Ponderado, a Figura 12 apresenta os valores obtidos por cada modelo avaliado. Essa métrica quantifica o desvio médio entre as previsões realizadas e os resultados observados. Assim, quanto menor o valor do EPMP, maior a precisão preditiva do modelo, o que torna essa métrica valiosa para avaliar a confiabilidade das previsões.

Figura 12 – Erro preditivo médio ponderado dos modelos.



Fonte: Dos Autores

O modelo *UfmgGer* apresentou o menor valor dessa métrica (0,202), sendo seguido de perto pelos modelos *UfmgSsRod* (0,209), *UfmgRod* (0,211) e *UfmgSnRod* (0,213). Esses resultados reforçam que os modelos da família *UFMG* não apenas se destacam pela TA, mas também pela precisão probabilística de suas estimativas. A maioria desses modelos apresenta EPMP inferior a 0,25, posicionando-se entre os mais consistentes e confiáveis para aplicações práticas. Contudo, isso indica que seus palpites, ou melhor, suas probabilidades, são distribuídas de forma mais uniforme; ou seja, eles se caracterizam como modelos indecisos, ou pouco convictos de sua previsão.

Também foram observados resultados satisfatórios no modelo *ConImp*, que obteve EPMP de 0,257, reforçando seu desempenho equilibrado entre TA, baixa complexidade e precisão probabilística. Os modelos *Sd1Ger* e *Sd0Ger*, ambos com EPMP de 0,261, também figuram entre os destaques positivos, embora apresentem variações ligeiramente superiores.

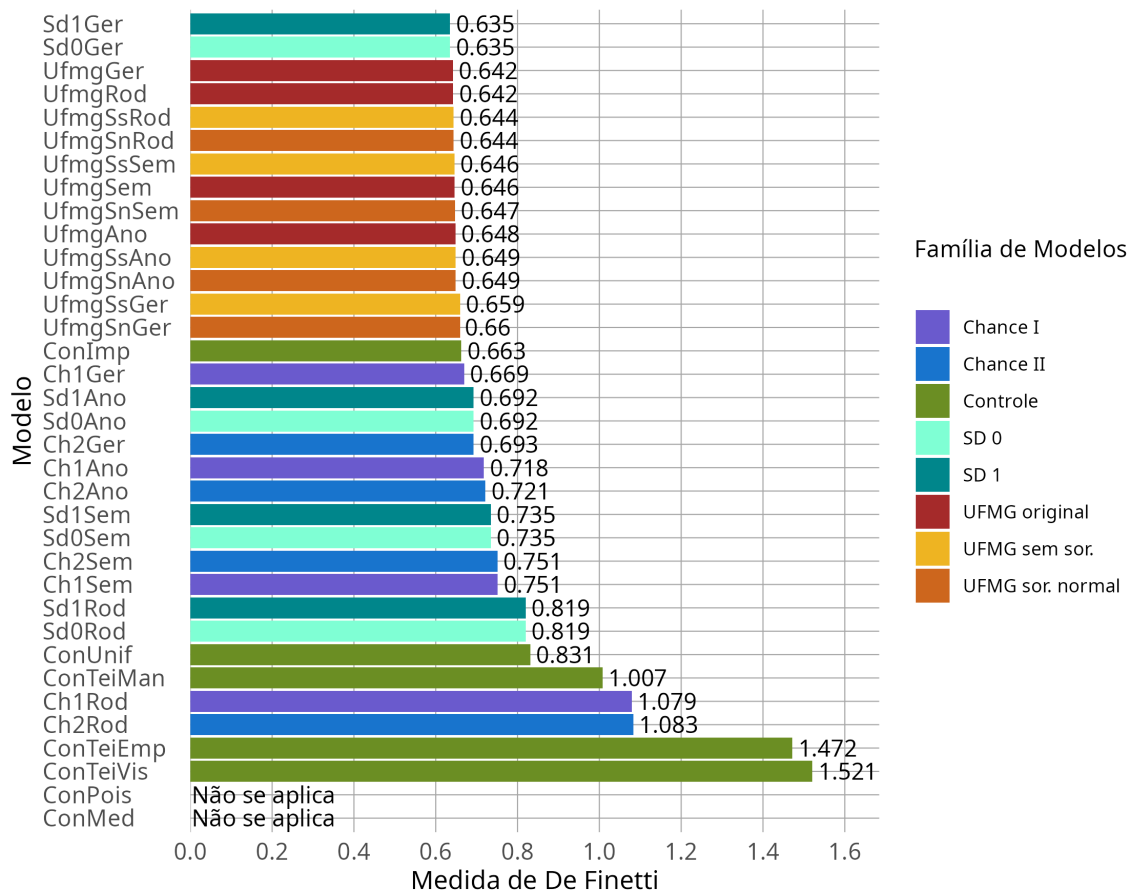
Na extremidade inferior da avaliação encontram-se os modelos com os maiores valo-

res de EPMP. Os principais destaques negativos são *ConTeiMan* (0,504), *ConTeiEmp* (0,736) e *ConTeiVis* (0,761), todos apresentando desempenho substancialmente inferior aos demais. Isso ocorre devido ao caráter convicto do resultado; assim, esses modelos distribuem sua probabilidade inteiramente no resultado escolhido.

Por fim, é importante observar que os modelos *ConPois* e *ConMed* não apresentam valor aplicável para essa métrica, pois não fornecem probabilidades de ocorrência dos resultados, sendo, portanto, excluídos da análise. Outra observação importante nessa métrica é que a metodologia tem um peso significativo, uma vez que as famílias de modelos estão quase sempre agrupadas.

Continuando a análise dos modelos pela ótica das métricas de desempenho, a Medida de De Finetti, apresentada na Figura 13, representa o erro cometido pelo modelo; assim, quanto menor o valor da medida, melhor o desempenho do modelo.

Figura 13 – Medida de De Finetti dos modelos.



Fonte: Dos Autores

Os modelos que apresentaram os melhores resultados foram *Sd1Ger* e *Sd0Ger*, ambos

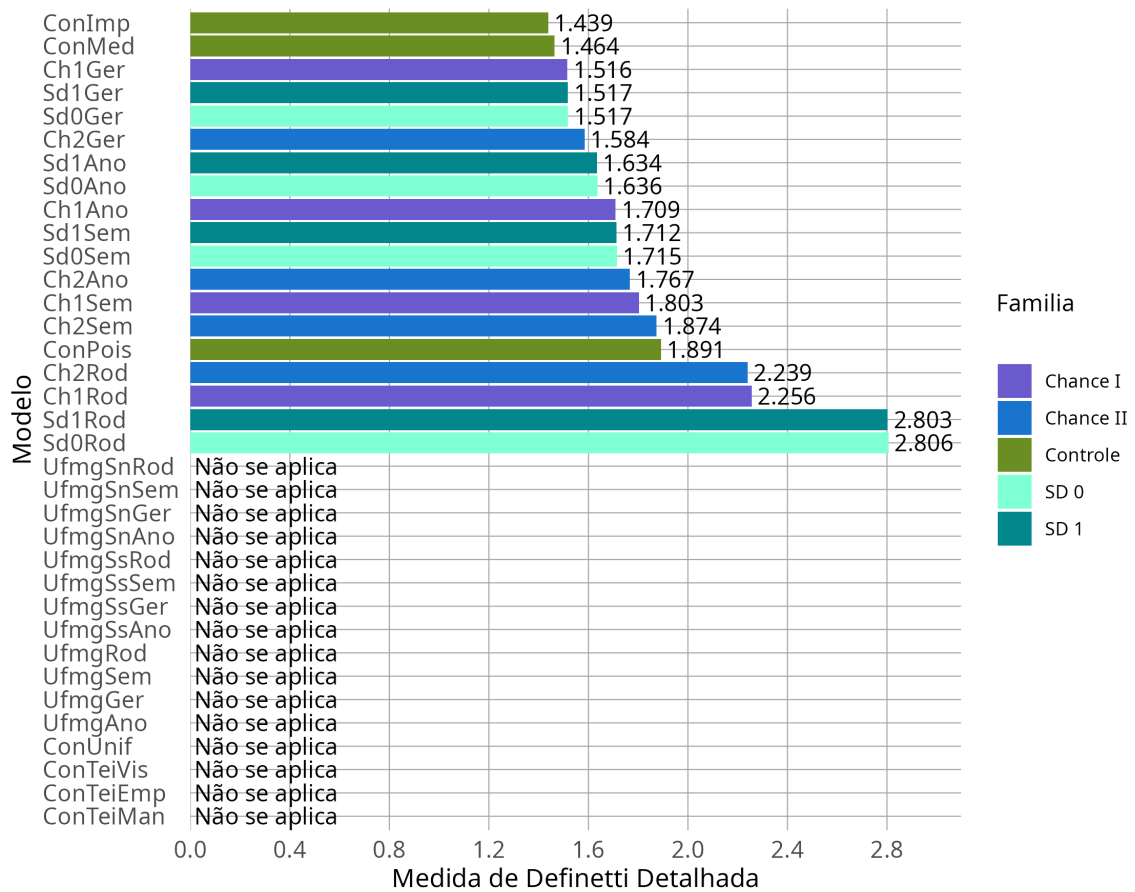
com o menor valor observado, igual a 0,635, seguidos por *UfmgGer* (0,642), *UfmgRod* (0,644), *UfmgSnRod* (0,644) e outros modelos pertencentes à família *UFMG*. Entretanto, considerando-se os 15 modelos que obtiveram as melhores posições nessa métrica, observa-se que a amplitude de valores é igual a 0,028, o que evidencia um desempenho muito próximo entre eles. Valores dessa métrica próximos a 0,660 são bastante comuns na literatura, como em (ARAÚJO et al., 2015; RAMOS; FERNANDES; BATISTA, 2021), que obtiveram valores iguais ou ligeiramente inferiores.

No extremo inferior do gráfico, destacam-se os modelos com as maiores medidas de De Finetti, como *Sd1Rod* e *Sd0Rod* (ambos com 0,819), *ConUnif* (0,822), *ConTeiMan* (1,007), *Ch1Rod* (1,079), *Ch2Rod* (1,083), *ConTeiEmp* (1,472) e *ConTeiVis* (1,521). Por fim, vale destacar que os modelos *ConPois* e *ConMed* não apresentaram valores aplicáveis para essa métrica, uma vez que não fornecem probabilidades para os eventos, apenas enunciam os placares ou resultados sem probabilidade associada.

A Figura 14 apresenta os valores da Medida de De Finetti Detalhada, que avalia a capacidade dos modelos de prever os placares das partidas. O modelo com a melhor calibração segundo essa métrica foi o *ConImp*, com MDD igual a 1,439, seguido por *ConMed* (1,464), *Ch1Ger* (1,516), *Sd1Ger* (1,517) e *Sd0Ger* (1,517). Ainda dentro de um intervalo considerado favorável, encontram-se os modelos *Ch2Ger* (1,584), *Sd1Ano* (1,634) e *Sd0Ano* (1,636).

Entre os modelos com desempenho mais fraco destacam-se *ConPois* (1,891), *Ch2Rod* (2,239), *Ch1Rod* (2,256), *Sd1Rod* (2,803) e *Sd0Rod* (2,806). Já os modelos da família *UFMG*, além de *ConTeiEmp*, *ConTeiMan*, *ConTeiVis* e *ConUnif*, não possuem valores aplicáveis para essa métrica, pois não apresentam previsões para o placar. Assim, a análise da Medida de De Finetti Detalhada evidencia a superioridade de modelos como *ConImp*, *ConMed* que são dois modelos que utilizam diretamente em suas previsões o número médio de gols de cada time.

Figura 14 – Medida de De Finetti detalhada dos modelos.



Fonte: Dos Autores

A Figura 15 apresenta um gráfico de dispersão que relaciona a TA com o Nível de Complexidade dos modelos avaliados. Essa visualização permite examinar a eficiência relativa de cada abordagem, isto é, identificar quais modelos conseguem maximizar o desempenho preditivo (alta taxa de acerto) ao mesmo tempo em que mantêm baixos níveis de complexidade, característica desejável tanto do ponto de vista prático quanto computacional. Para mais informações sobre o nível de complexidade, o Apêndice D apresenta a distribuição de valores determinada pelos autores neste trabalho.

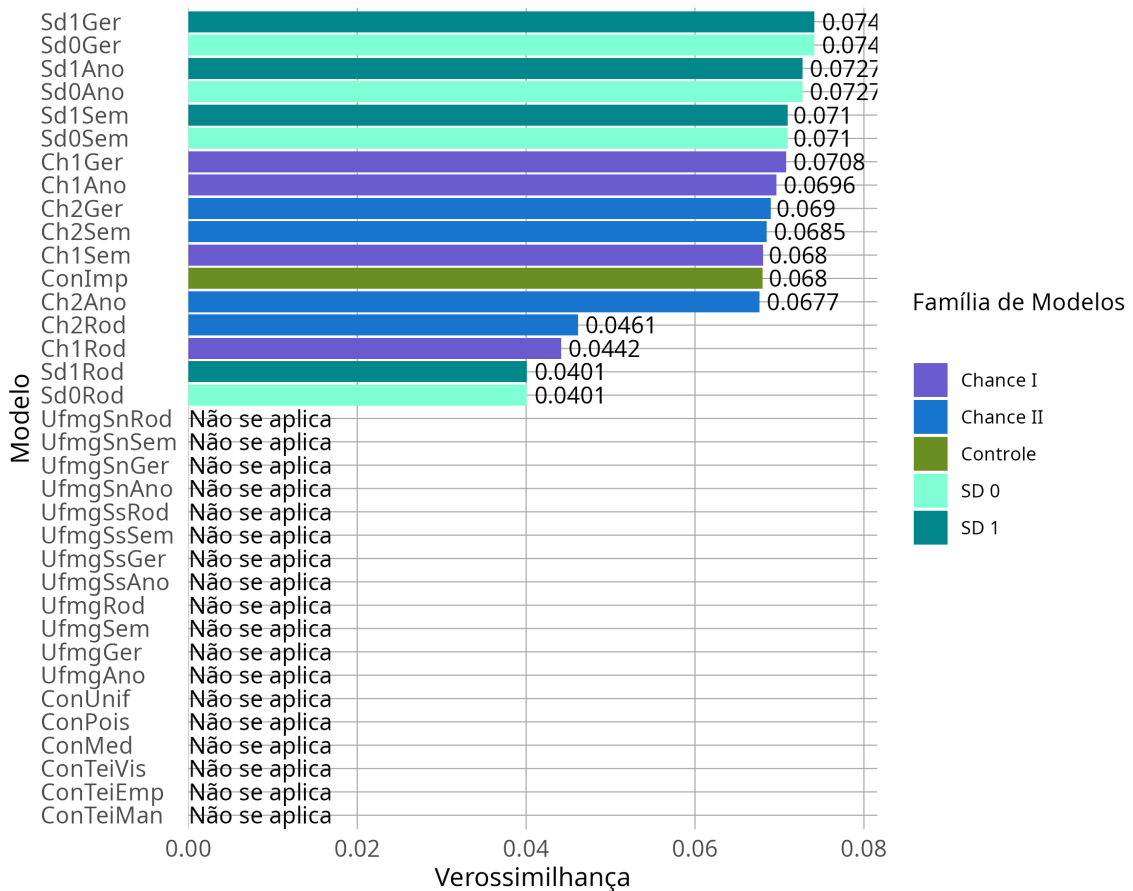
O modelo que mais se destaca na figura é o *ConTeiMan*, posicionado no canto superior esquerdo do gráfico. Ele apresenta a maior TA entre todos os modelos (aproximadamente 0,50), combinada com o menor nível de complexidade registrado (1). Essa configuração torna o modelo especialmente atrativo em cenários nos quais se busca maximizar o desempenho mantendo uma estrutura extremamente simples.

Um segundo grupo de destaque é composto pelos modelos da família *UFMG*, que se



mílias *SD 0* e *SD 1*. Entre eles, destacam-se *Sd1Ger* e *Sd0Ger*, ambos com verossimilhança igual a 0,074, seguidos por *Sd1Ano* (0,0725), *Sd0Ano* (0,0727) e *Sd1Sem* e *Sd0Sem* (aproximadamente 0,071). Modelos das famílias *Chance I* e *Chance II* também aparecem na faixa superior, com valores entre 0,068 e 0,070. Mesmo considerando os valores mais elevados, nenhum deles se aproxima efetivamente de 1, que pode ser tomado como um valor de referência para uma boa previsão.

Figura 16 – Valores de verossimilhança do resultado dos modelos.



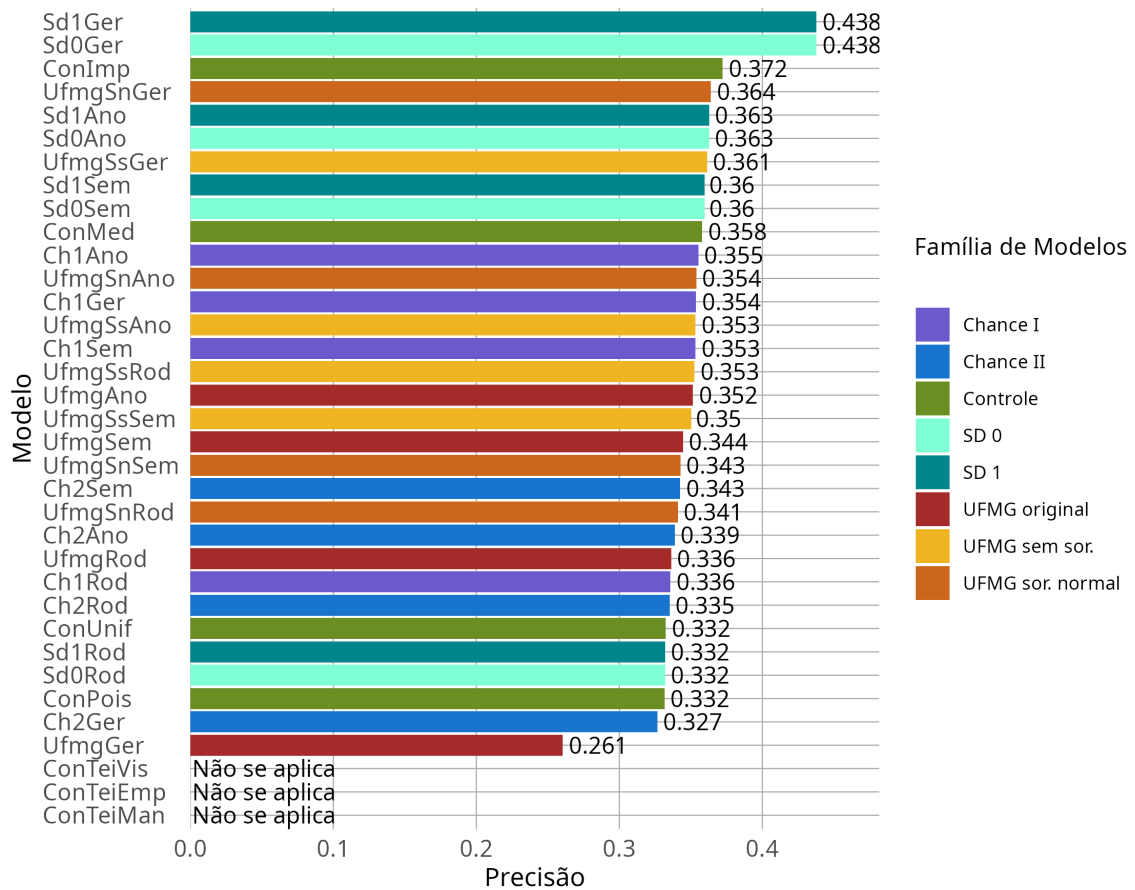
Fonte: Dos Autores

Vale destacar que grande parte dos modelos das famílias *UFMG* e *Controle* não apresenta valores aplicáveis para essa métrica, pois não possuem distribuição de probabilidade associada para a comparação dos parâmetros.

Agora, começando a analisar as métricas originadas da matriz de confusão temos a Figura 17 com os valores de Precisão. Essa métrica indica a proporção de previsões positivas corretamente identificadas, O melhor desempenho destacam-se *Sd1Ger* e *Sd0Ger*, ambos com precisão igual a 0,438, constituindo os maiores valores observados na análise. Em seguida,

aparece o modelo *ConImp* (0,372), com uma diferença de 0.056% em relação aos modelos anteriores.

Figura 17 – Precisão dos modelos.



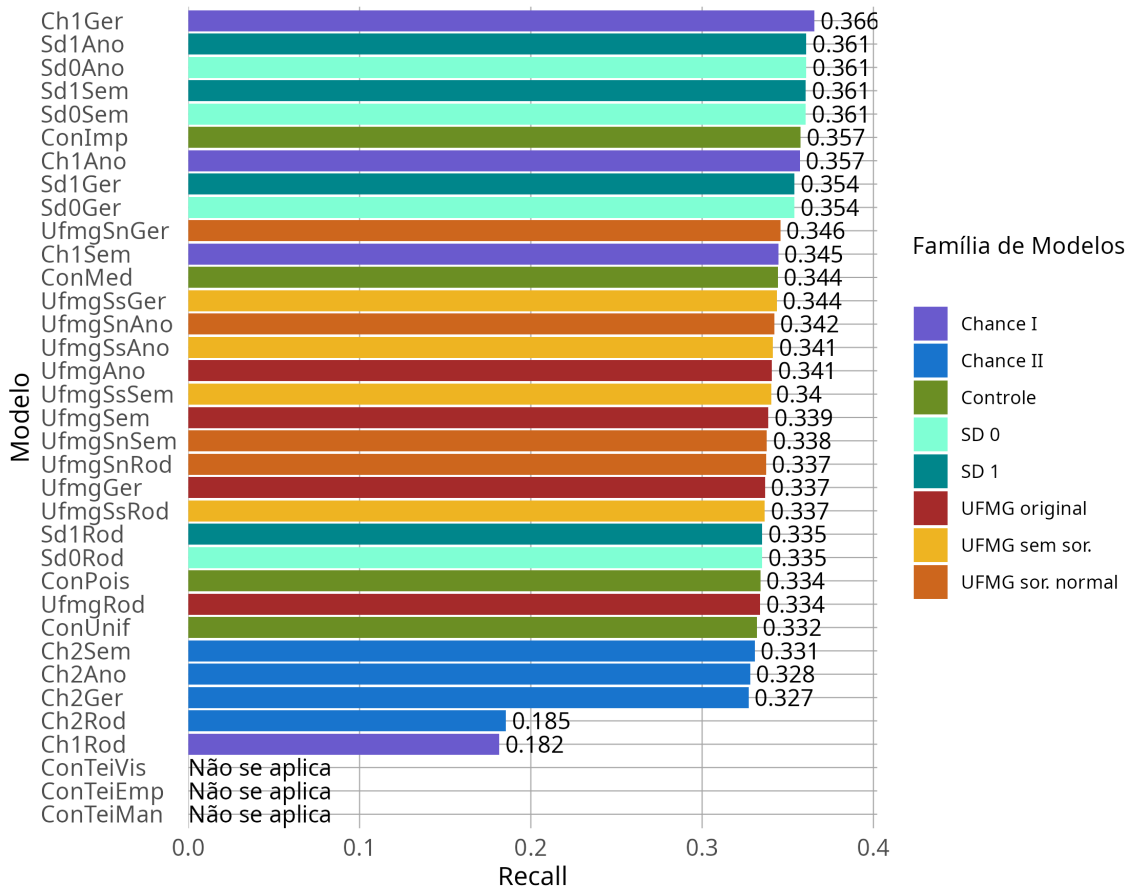
Fonte: Dos Autores

Destaca-se que, para essa métrica, as famílias apresentam-se mais segmentadas. Embora metodologias próximas obtenham resultados semelhantes, elas não formam um bloco único entre si. Os modelos mais penalizados por essa métrica foram *UfmgGer* (0,261), *Ch2Ger* (0,327), *ConPois* (0,332) e *Sd0Rod* (0,332), indicando menor precisão e maior tendência a erros de classificação. Os modelos *ConTeiVis*, *ConTeiEmp* e *ConTeiMan* foram excluídos da comparação, visto que suas previsões apresentam apenas uma classe, o que poderia gerar comparações injustas com os demais métodos.

A Figura 18 apresenta os valores de *Recall* (MC\_Rec), também derivados da matriz de confusão. Essa métrica quantifica a capacidade do modelo de recuperar corretamente os elementos de uma determinada classe. Contudo, seu poder comparativo foi limitado: a variação entre os 30 modelos não ultrapassou 0,04, com valores médios em torno de 34%, desempenho

consideravelmente inferior ao observado em Wong et al. (2025) que observou 53,7%.

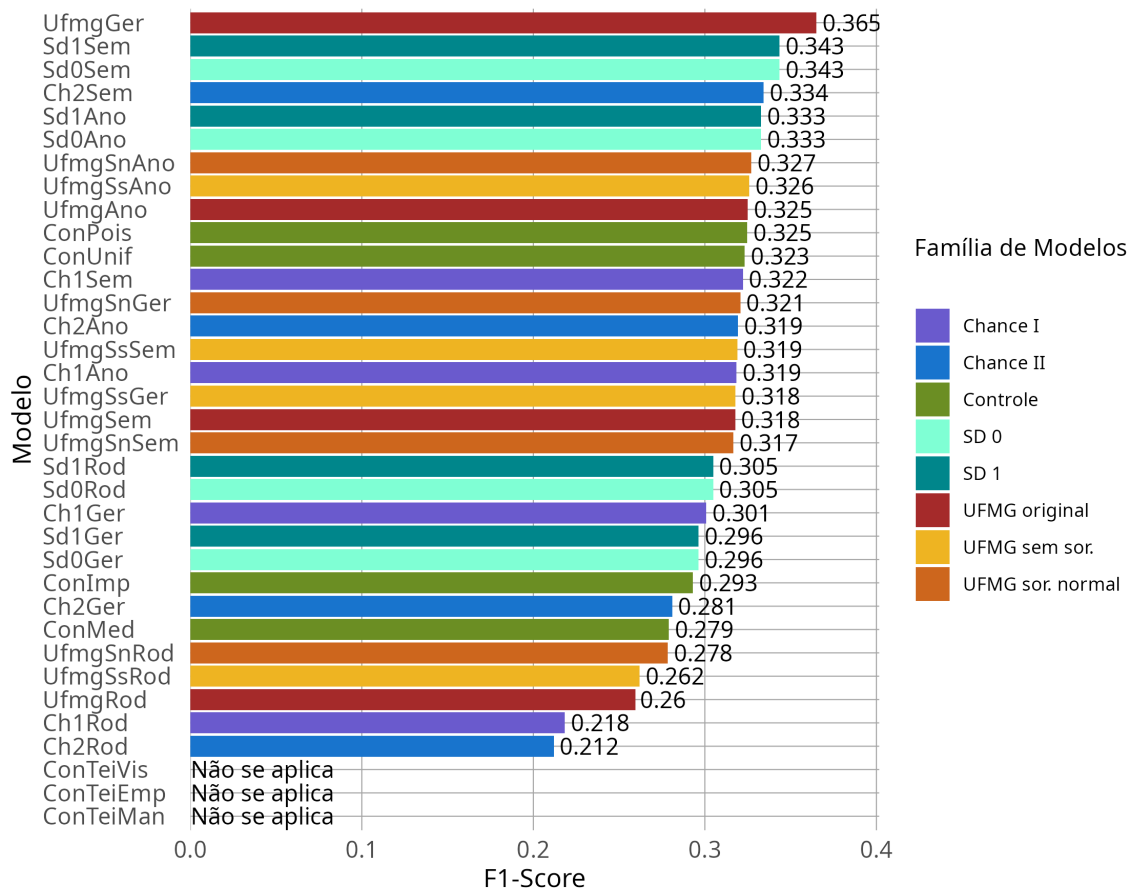
Figura 18 – Recall dos modelos.



Fonte: Dos Autores

*F1-Score*, é uma métrica muito utilizada para lidar com previsões, principalmente se essas previsões envolverem classes ou categorias. Sabendo disso, podemos observar a Figura 19. Os melhores desempenhos foram obtidos pelos modelos *UfmgGer* (0,365), *Sd1Sem* (0,343), *Sd0Sem* (0,343) e *Ch2Sem* (0,334), que lideram a análise. Na parte inferior da distribuição destacam-se modelos como *UfmgRod* (0,260), *Ch1Rod* (0,218) e *Ch2Rod* (0,212), que exibem F1-Scores sensivelmente mais baixos, indicando maior desequilíbrio entre precisão e recall. Tais modelos tendem a falhar simultaneamente em capturar eventos verdadeiros e em evitar falsos positivos.

Figura 19 – Valores de F1-Score dos modelos.

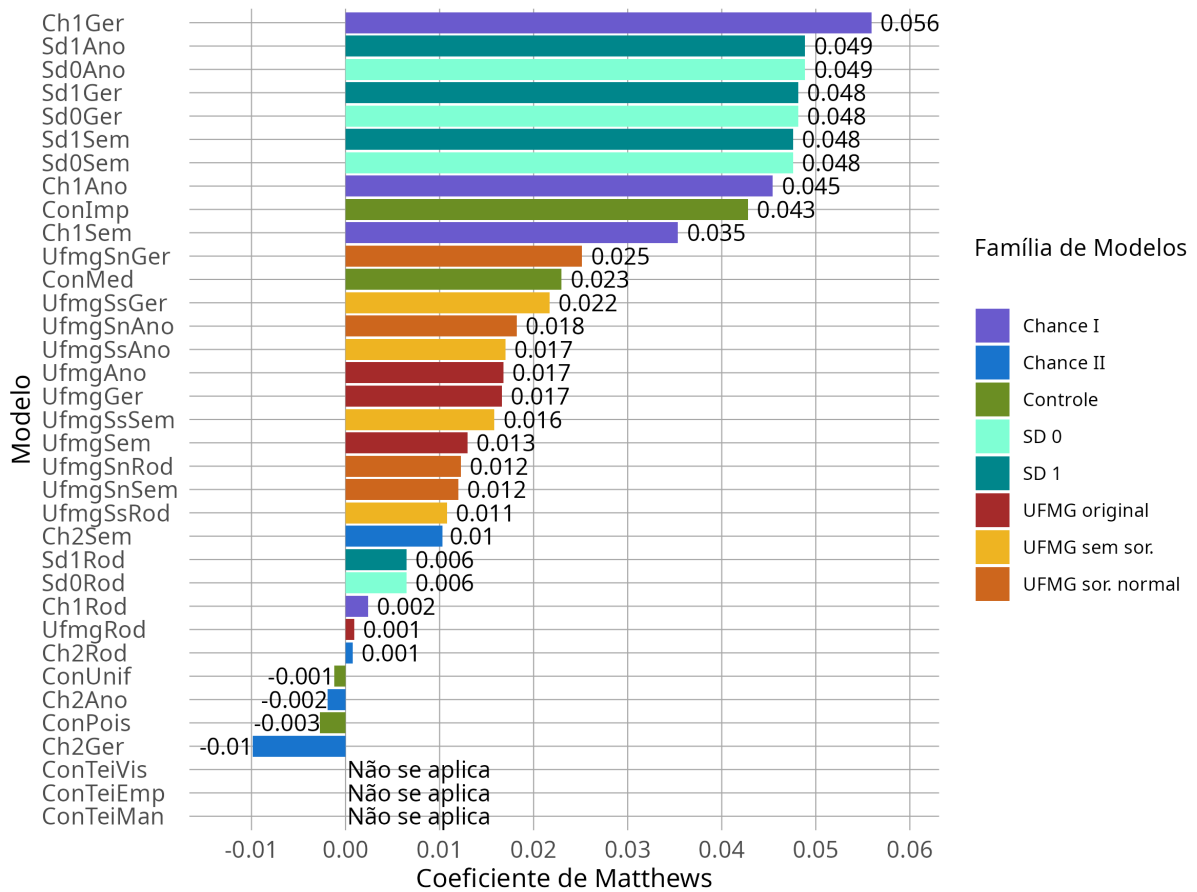


Fonte: Dos Autores

A Figura 20 apresenta os valores do *Coefficiente de Correlação de Matthews* (MCC), uma métrica robusta para avaliar modelos de classificação, variando de  $-1$  (desempenho ruim) a  $+1$  (classificação perfeita). Os melhores resultados foram obtidos pelos modelos *Ch1Ger* (0,056), *Sd1Ano* (0,049), *Sd0Ano* (0,049), *Sd1Ger* (0,048) e *Sd0Ger* (0,048). Embora baixos, esses valores indicam desempenho superior ao acaso e coerente com outras métricas, reforçando a maior estabilidade das famílias *SD* e *Chance I*. Modelos como *ConImp* (0,043) e *Ch1Sem* (0,035) também apresentaram correlações moderadas entre previsões e resultados.

No extremo inferior, modelos como *Ch2Ger* ( $-0,003$ ), *Ch2Ano* ( $-0,002$ ) e *ConUnif* ( $-0,001$ ) exibiram *MC\_MCC* negativo, que indicaria desempenho inferior ao acaso, mas o que prevalece para esses modelos é a proximidade do valor nulo. Assim, como *UfmgRod*, *Sd0Rod* e *Sd1Rod*, apresentaram valores próximos de zero, refletindo baixa capacidade discriminativa.

Figura 20 – Valores do coeficiente de correlação de matthews dos modelos.



Fonte: Dos Autores

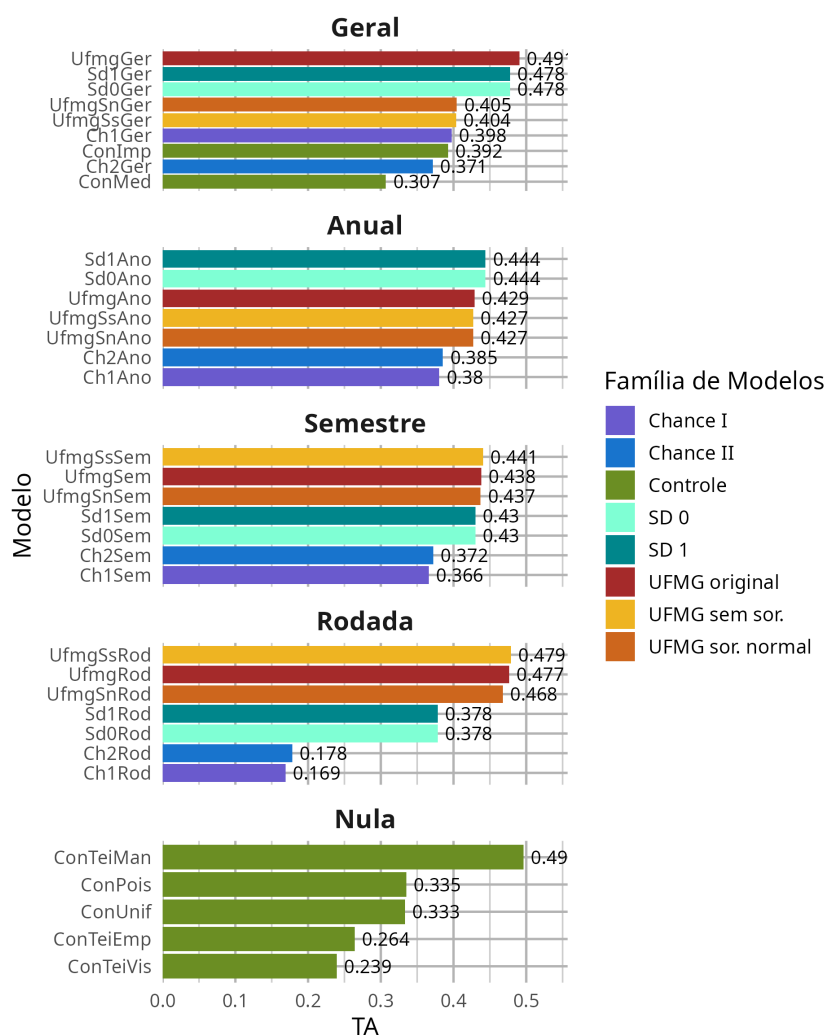
## 5.2.2 Comparação dos Resultados Agrupados por Janelas de Tempo

Nesta subseção são apresentados gráficos de três métricas: TA, MD e F1. Calculadas para os modelos agrupados por Janelas de Tempo, a fim de avaliar o impacto de diferentes quantidades de rodadas nas previsões. Essas métricas foram escolhidas de forma arbitrária, buscando contemplar aspectos relevantes do desempenho dos modelos e dialogar com a literatura existente, que tradicionalmente faz uso dessas medidas em estudos semelhantes.

A análise da Figura 21, que apresenta a TA dos modelos agrupados por intervalos de tempo, revela padrões importantes sobre a influência da quantidade de informação histórica na capacidade preditiva dos modelos. O grupo Geral, que utiliza toda a série histórica como referência, apresenta os maiores valores registrados, sendo liderado pelo modelo *UfmgGer* (0,491), seguido por *Sd1Ger* e *Sd0Ger* (ambos com 0,478).

Apesar disso, as demais janelas de tempo não se encontram distantes desses resultados, evidenciando a importância do banco de dados. Nos grupos Anual e Semestral, embora os valores sejam menores, observa-se certa consistência; esses dois grupos, inclusive, apresentam desempenhos muito próximos, indicando baixo impacto dessa divisão temporal. Já no grupo de memória Nula, observa-se grande variabilidade, uma vez que esses modelos são essencialmente dependentes de sua metodologia e não utilizam informações prévias, o que resulta em maior instabilidade nas previsões.

Figura 21 – Taxa de acerto dos modelos agrupada por janelas de tempo.

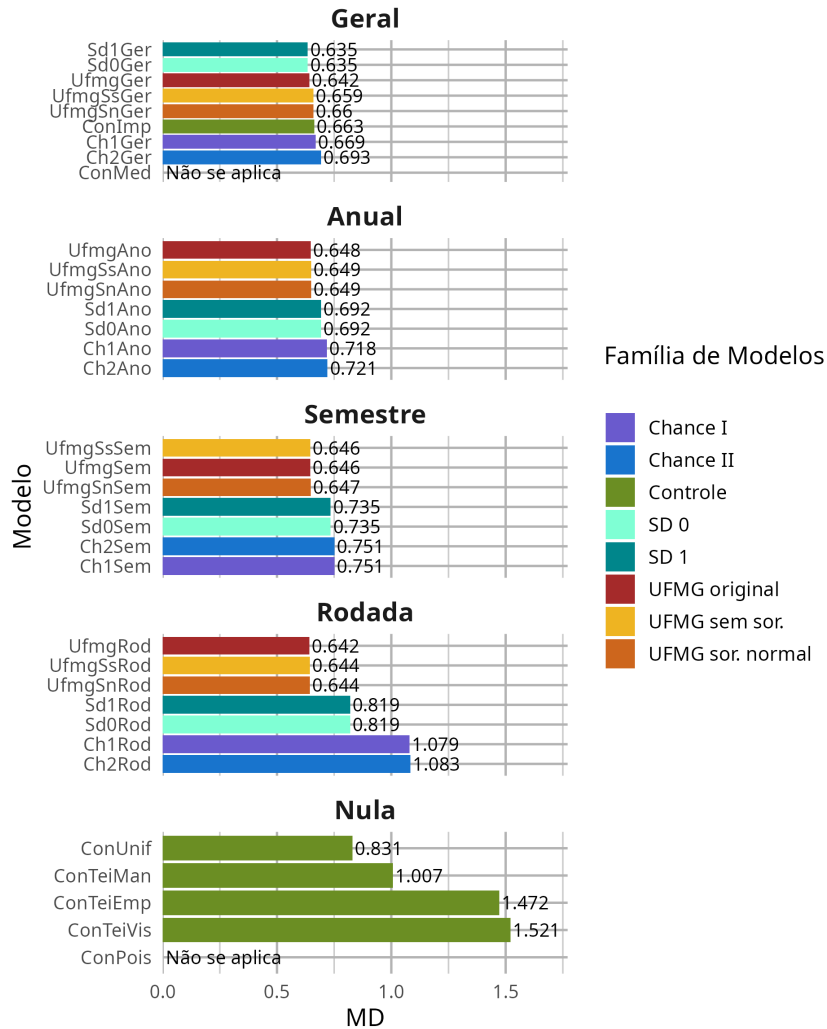


Fonte: Dos Autores

Em relação à MD, representada na Figura 22, os mesmos comportamentos destacados anteriormente com a TA foram observados, evidenciando que tanto medidas de precisão quanto de acurácia não sofrem alterações significativas com diferentes amostragens do banco de dados para os modelos avaliados. Ainda assim, pode-se destacar que os modelos da família *UFMG*

tiveram resultados ligeiramente melhores para previsões anuais, semestrais e de rodadas, enquanto os modelos de cada família apresentaram desempenhos próximos entre si.

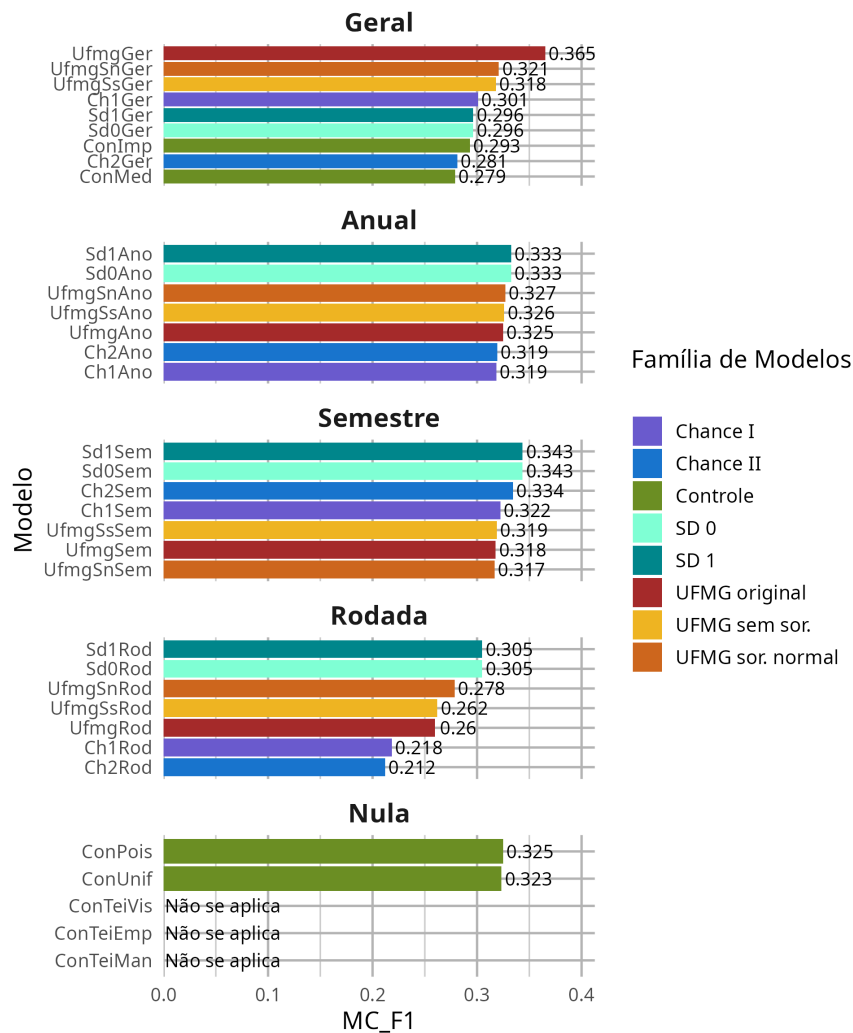
Figura 22 – Medida de De Finetti dos modelos agrupada por janelas de tempo.



Fonte: Dos Autores

A Figura 23, referente ao *F1-Score*, reforça as tendências observadas nas métricas anteriores, uma vez que os melhores resultados se concentram nas previsões que utilizam a maior parcela do banco de dados. Ainda assim, os valores permanecem muito próximos entre si. Contudo, há uma mudança em relação à análise da métrica MD: embora as famílias permaneçam próximas, os modelos *Sd0* e *Sd1* apresentaram desempenho superior nas janelas de tempo anual, semestral e de rodada, seguidos pelos modelos da família *UFMG*.

Figura 23 – F1-score dos modelos agrupados por janelas de tempo.



Fonte: Dos Autores

De forma integrada, a comparação dos três gráficos evidencia que o janela selecionada influencia a qualidade das previsões, independentemente da métrica utilizada. Modelos que utilizam janelas amplas (*Geral*, *Anual* e *Semestral*) apresentam desempenho mais consistente e confiável. Em contraste, modelos dependentes de janelas muito curtas (*Rodada*) apresentam desempenho inferior em todas as métricas, sugerindo que informações de curto prazo não são suficientes para capturar a dinâmica das partidas. Porém, isso levanta uma questão: Qual seria a quantidade ideal de rodadas que um modelo deve considerar para obter precisões mais consistente? A sessão 5.3 busca responder esta pergunta. Já modelos sem memória (*Nula*), como esperado, apresentam resultados aleatórios ou “impessoais”, podendo ou não acertar. Esses resultados reforçam que a escolha da janela temporal é um fator crítico no desempenho dos modelos de previsão.

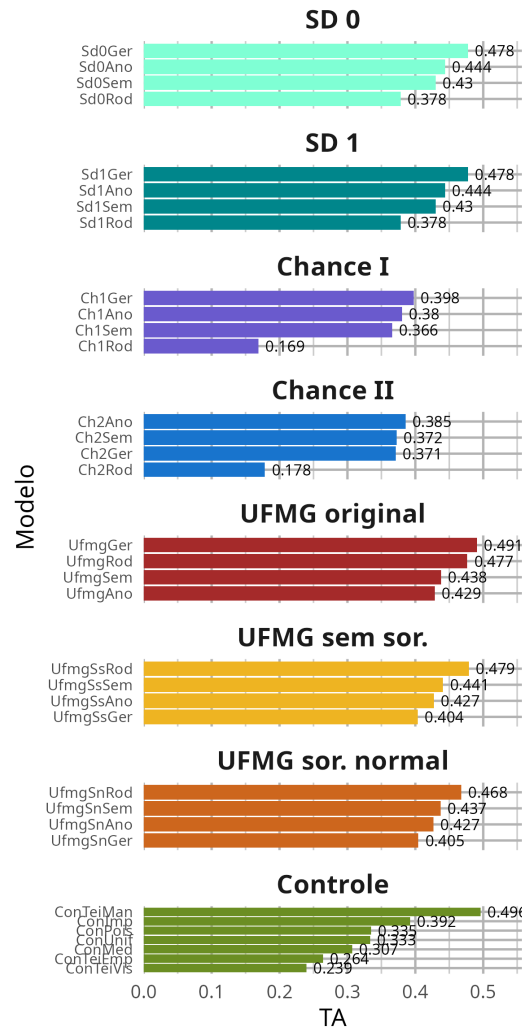
Para trazer evidências estatísticas das informações trazidas acima, foi feito um teste de Kruskal–Wallis e Dunn post-hoc com correção de Holm que apontaram que para a TA, não houve diferença estatística significativa entre as classes de Memória ( $H = 3.205$ ,  $p = 0.5242$ ). Já MD, observou-se diferença estatisticamente significativa entre as classes de Memória ( $H = 11.867$ ,  $p = 0.0184$ ); identificar pares específicos que diferem são: Geral vs Nula. E MC\_F1, observou-se diferença estatisticamente significativa entre as classes de Memória ( $H = 18.256$ ,  $p = 0.0011$ ); para os pares específicos Rodada vs Semestre. Para mais detalhes sobre o teste feito vide no Apêndice E e F as tabelas 10, 11, 12 e 13.

### 5.2.3 Comparação de resultados agrupados por família

Da mesma forma que na subseção anterior aqui será discutido os valores das métricas: TA, MD e F1, dos modelos agrupados em relação às famílias, ou melhor, em relação à metodologia central dos modelos. Sendo assim, a Figura 24 apresenta a TA dos modelos agrupados. Com já visto na Figura 11 as TA de grande parte dos modelos se concentram entre 30% à 40%, devido a essa baixa amplitude as comparações aqui propostas são sutis e baseadas em centésimos ou milésimos. Portanto, as famílias *SD 0* e *SD 1* apresentam desempenho sólido e bastante homogêneo, com valores que variam de 0,378 a 0,478.

Entre as famílias probabilísticas baseadas em aleatoriedade controlada, observa-se contraste significativo. A família *Chance I* apresenta valores intermediários, com destaque para *Ch1Ger* (0,398) e quedas mais acentuadas em *Ch1Rod* (0,169). Já a família *Chance II* mostra desempenho um pouco inferior, com valores variando entre 0,371 e 0,385. Isso reflete sua metodologia, mas também essa métrica foi a mais afetada pela TF, uma vez que diversas vezes o modelo não conseguiu realizar a regressão e isso resultou na ausência de previsão para essas partidas.

Figura 24 – Taxa de acerto dos modelos agrupada por família.



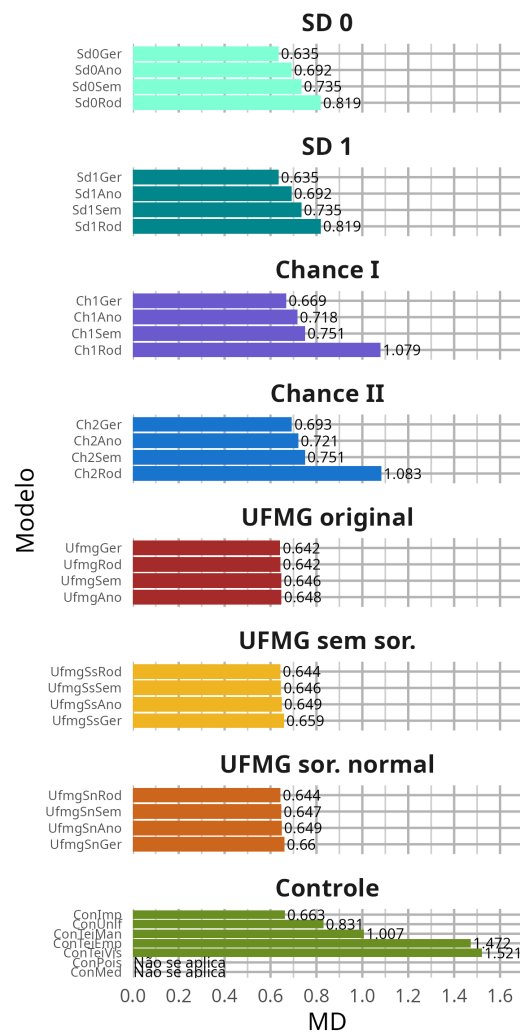
Fonte: Dos Autores

As famílias derivadas do modelo da *UFMG* exibem desempenho globalmente superior. A versão *UFMG original* apresenta os maiores valores entre todas as famílias (com destaque para *UfmgGer*, 0,491), enquanto as variantes *UFMG sem sorteio* e *UFMG sorteio normal* também mantêm ótimo desempenho, com valores que variam entre 0,404 e 0,479. Esses resultados mostram que as diferentes adaptações do modelo original resultaram em pouco efeito sobre o desempenho geral.

Os modelos *Controle* apresenta comportamento heterogêneo, indo do melhor resultado individual observado no estudo (*ConTeiMan* com 0,496) a desempenhos significativamente inferiores, como *ConTeiVis* (0,239). Ainda assim, essa família surpreende por alcançar resultados próximos aos das demais, mesmo sem utilizar metodologias robustas ou grandes volumes de dados.

A Figura 25 apresenta a MD dos modelos agrupados por família. Assim, iniciando com as famílias *SD 0* e *SD 1* observa-se um comportamento praticamente idêntico, com valores variando entre 0,635 e 0,819. Em ambas, os modelos baseados em janelas temporais amplas (geral, anual e semestral) alcançam valores baixos de MD. Contudo, quando a janela é reduzida ao nível de rodada, a métrica aumenta significativamente para 0,819, refletindo pior qualidade probabilística.

Figura 25 – Medida de De Finetti dos modelos agrupada por família.



Fonte: Dos Autores

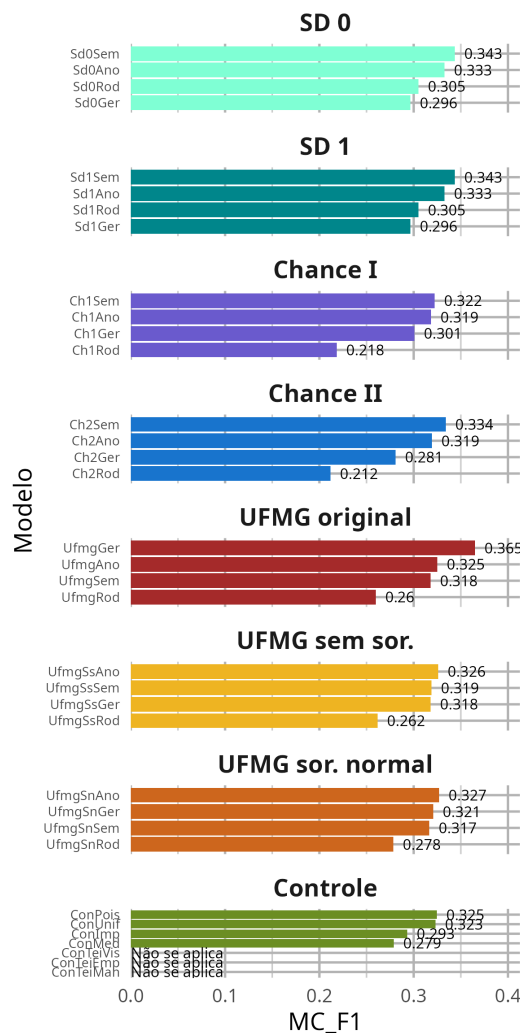
As famílias *Chance I* e *Chance II* apresentam resultados intermediários e elevada sensibilidade à redução da memória. Nas versões gerais, os valores situam-se entre 0,669 e 0,693, aumentando gradualmente nas versões anuais e semestrais, que atingem até 0,751. A pior calibração ocorre nas versões de rodada, com *Ch1Rod* alcançando 1,079 e *Ch2Rod* chegando a 1,083, pois, com pouca informação disponível, essas famílias tendem a ser mais convictas em

seus “palpites”.

Os modelos da família *UFMG* apresentam medidas muito próximas entre si, o que se explica pelo fato de não serem modelos convictos no resultado, isto é, distribuírem as probabilidades entre os três eventos possíveis. As três variantes derivadas do modelo *UFMG* exibem os melhores desempenhos gerais, com valores estáveis entre 0,642 e 0,660, independentemente da janela de tempo.

Como já observado, os modelos da família *Controle* são os que mais se distanciam entre si. Enquanto *ConImp* (0,663) e *ConUnif* (0,831) apresentam valores aceitáveis de MD, outros modelos, como *ConTeiMan* (1,007), *ConTeiEmp* (1,472)

Figura 26 – F1-score dos modelos agrupada por família.



Fonte: Dos Autores

A Figura 26 apresenta os valores de *F1-Score* dos modelos agrupados por família, permitindo avaliar o equilíbrio entre precisão e recall dentro de cada abordagem. Como o F1 repre-

senta a média harmônica dessas duas métricas, valores mais altos indicam tanto boa capacidade de identificar corretamente as instâncias positivas quanto baixa taxa de falsos positivos.

As famílias *SD 0* e *SD 1* demonstram comportamento quase idêntico, com valores variando entre 0,296 e 0,343. Em ambos os grupos, os melhores desempenhos aparecem nas versões semestrais (0,343), enquanto as versões gerais apresentam os menores valores (0,296). Esse padrão sugere que essas famílias se beneficiam de janelas de tempo moderadas, capazes de capturar tendências recentes sem acumular ruído excessivo. A similaridade entre *SD 0* e *SD 1* reforça a robustez estrutural dos modelos estocásticos descontínuos.

Já modelos, *Chance I* e *Chance II* apresentam desempenhos intermediários, porém menos consistentes. Na família *Chance I*, os valores variam entre 0,218 e 0,322, com queda pronunciada no modelo baseado em rodada (0,218). O mesmo ocorre na família *Chance II*, onde os valores vão de 0,212 a 0,334. A perda significativa de desempenho nas versões de memória curta indica que modelos com maior componente aleatório não conseguem equilibrar precisão e recall quando há pouca informação histórica disponível.

As três famílias derivadas do modelo *UFMG* apresentam excelente desempenho relativo e maior estabilidade interna. A versão *UFMG original* destaca-se com o maior F1 do gráfico (0,365), enquanto suas variações *sem sorteio* e *com sorteio normal* mantêm valores próximos entre si, variando de 0,278 a 0,327. Esse padrão mostra que o método da *UFMG* preserva seu bom equilíbrio entre precisão e recall independentemente da forma de sorteio utilizada, o que evidencia a solidez estatística da estrutura do modelo.

Por fim, a família *Controle* apresenta o conjunto mais irregular. Modelos como *ConPoís* (0,325) e *ConUnif* (0,323) exibem valores competitivos, mas outros modelos, *ConTeiVis*, *ConTeiEmp* e *ConTeiMan*, aparecem como “Não se aplica”, indicando que suas estruturas não permitem o cálculo adequado do F1. Essa heterogeneidade mostra que a família de controle não mantém uma lógica uniforme de classificação, apresentando forte dependência das características individuais de cada variante.

Também foram considerados os resultados de um teste de Kruskal–Wallis e Dunn post-hoc com correção de Holm, que observou que para a TA há diferença estatisticamente significativa entre as classes de Família ( $H = 18.070$ ,  $p = 0.0118$ ); para os pares (ex.: Controle vs *SD 0* e Controle vs *SD 1*). Contudo, para a MD e MC\_F1, não houve diferença estatística significativa entre as classes de Família, para mais detalhes vide as Tabelas: 14, 15 e 16 17 presentes no Apêndices G, H e I.

#### 5.2.4 Tabela Geral de Resultados

O quadro apresentado na Figura 27 apresenta de forma consolidada todas as métricas calculadas para os modelos avaliados ao longo deste estudo. Essa visão integrada permite examinar simultaneamente desempenho, consistência probabilística, robustez classificatória e a parcimoniedade.

A análise detalhada da tabela revela que os modelos da família *UFMG* especialmente *UFMGGer*, ocupam as primeiras posições em diversas métricas, destacando-se pela combinação de alta TA (0,4907), excelente calibração probabilística (MD = 0,6420), bom F1-Score (0,3651) e um dos maiores valores de MC\_MCC (0,0166). As famílias *SD 0* e *SD 1* apresentam resultados consistentes e homogêneos, com destaque para *Sd1Ger* e *Sd0Ger*, que alcançam TA de 0,4777 e valores baixos de MD (0,6349). Esses modelos também figuram entre os melhores em precisão e recall, mantendo valores equilibrados de F1-Score e MCC.

De modo geral, a tabela sintetiza e confirma os achados descritos nas análises anteriores: as famílias *UFMG*, *SD 0* e *SD 1* apresentam os resultados mais sólidos e consistentes em múltiplas métricas, enquanto *Chance I*, *Chance II* e *Controle* revelam maior variabilidade e fragilidade estrutural. A integração entre acurácia, calibração probabilística, equilíbrio classificatório e simplicidade torna os modelos *UFMG* particularmente competitivos e os destaca como candidatos mais adequados para aplicações práticas de previsão.

Figura 27 – Quadro de resultados das métricas.

Modelo	TA	EPMP	MD	MDD	NC	TF	Vero	MC_Pre	MC_Rec	MC_F1	MC_MCC
UfmgSnRod	0.4678	0.2133	0.6437	NA	3.286	1.0000	NA	0.3409	0.3373	0.2785	0.0123
UfmgSnSem	0.4372	0.2401	0.6467	NA	3.429	1.0000	NA	0.3427	0.3378	0.3166	0.0120
UfmgSnGer	0.4046	0.2724	0.6598	NA	3.714	1.0000	NA	0.3641	0.3457	0.3207	0.0251
UfmgSnAno	0.4269	0.2490	0.6488	NA	3.571	1.0000	NA	0.3538	0.3422	0.3271	0.0182
UfmgSsRod	0.4789	0.2092	0.6436	NA	2.714	1.0000	NA	0.3525	0.3366	0.2619	0.0108
UfmgSsSem	0.4409	0.2382	0.6459	NA	2.857	1.0000	NA	0.3503	0.3405	0.3189	0.0158
UfmgSsGer	0.4036	0.2722	0.6593	NA	3.143	1.0000	NA	0.3613	0.3436	0.3179	0.0217
UfmgSsAno	0.4272	0.2484	0.6487	NA	3.000	1.0000	NA	0.3533	0.3414	0.3259	0.0170
UfmgRod	0.4766	0.2107	0.6424	NA	3.143	1.0000	NA	0.3362	0.3338	0.2597	0.0009
UfmgSem	0.4380	0.2394	0.6461	NA	3.286	1.0000	NA	0.3445	0.3387	0.3177	0.0130
UfmgGer	0.4907	0.2017	0.6420	NA	3.571	1.0000	NA	0.2606	0.3367	0.3651	0.0166
UfmgAno	0.4289	0.2468	0.6479	NA	3.429	1.0000	NA	0.3515	0.3408	0.3250	0.0168
Sd1Rod	0.3783	0.3647	0.8195	2.8025	3.571	1.0000	0.0401	0.3319	0.3350	0.3049	0.0065
Sd1Sem	0.4299	0.3309	0.7348	1.7123	3.714	1.0000	0.0710	0.3596	0.3605	0.3435	0.0476
Sd1Ger	0.4777	0.2609	0.6349	1.5172	4.000	1.0000	0.0742	0.4379	0.3540	0.2963	0.0481
Sd1Ano	0.4441	0.3039	0.6924	1.6341	3.857	1.0000	0.0727	0.3627	0.3609	0.3327	0.0489
Sd0Rod	0.3783	0.3647	0.8195	2.8060	3.143	1.0000	0.0401	0.3319	0.3350	0.3049	0.0065
Sd0Sem	0.4299	0.3309	0.7348	1.7147	3.286	1.0000	0.0710	0.3596	0.3605	0.3435	0.0476
Sd0Ger	0.4777	0.2609	0.6349	1.5174	3.571	1.0000	0.0742	0.4379	0.3540	0.2963	0.0481
Sd0Ano	0.4441	0.3039	0.6924	1.6363	3.429	1.0000	0.0727	0.3627	0.3609	0.3327	0.0489
Ch2Rod	0.1785	0.3339	1.0831	2.2392	3.714	0.5540	0.0461	0.3352	0.1854	0.2119	0.0008
Ch2Sem	0.3723	0.3025	0.7508	1.8743	3.857	0.9673	0.0685	0.3426	0.3308	0.3343	0.0103
Ch2Ger	0.3714	0.2883	0.6929	1.5844	4.143	0.9995	0.0690	0.3266	0.3272	0.2810	-0.0099
Ch2Ano	0.3854	0.2849	0.7212	1.7667	4.000	0.9822	0.0677	0.3388	0.3281	0.3194	-0.0019
Ch1Rod	0.1692	0.3290	1.0790	2.2556	3.714	0.5375	0.0442	0.3358	0.1815	0.2184	0.0024
Ch1Sem	0.3660	0.3246	0.7514	1.8028	3.857	0.9628	0.0680	0.3533	0.3445	0.3223	0.0353
Ch1Ger	0.3976	0.2711	0.6693	1.5157	4.143	0.9994	0.0708	0.3537	0.3657	0.3009	0.0559
Ch1Ano	0.3804	0.3100	0.7179	1.7094	4.000	0.9839	0.0696	0.3553	0.3571	0.3185	0.0454
ConUnif	0.3333	0.4066	0.8308	NA	1.571	1.0000	NA	0.3325	0.3320	0.3233	-0.0012
ConPois	0.3349	NA	NA	1.8910	1.714	1.0000	NA	0.3317	0.3340	0.3248	-0.0027
ConMed	0.3070	NA	NA	1.4644	1.857	1.0000	NA	0.3579	0.3442	0.2790	0.0230
ConTeiVis	0.2395	0.7605	1.5210	NA	1.000	1.0000	NA	NA	NA	NA	NA
ConTeiEmp	0.2642	0.7358	1.4716	NA	1.000	1.0000	NA	NA	NA	NA	NA
ConTeiMan	0.4963	0.5037	1.0074	NA	1.000	1.0000	NA	NA	NA	NA	NA
ConImp	0.3924	0.2566	0.6626	1.4387	2.714	1.0000	0.0680	0.3723	0.3575	0.2931	0.0428

Fonte: Dos Autores

### 5.3 RESULTADOS DA SIMULAÇÃO

De forma geral, não foram identificados picos ou vales marcantes que indicassem a existência de um intervalo ideal de rodadas para o desempenho dos modelos. Apenas previsões feitas antes da 10ª rodada mostram efeitos negativos mais evidentes, especialmente em modelos

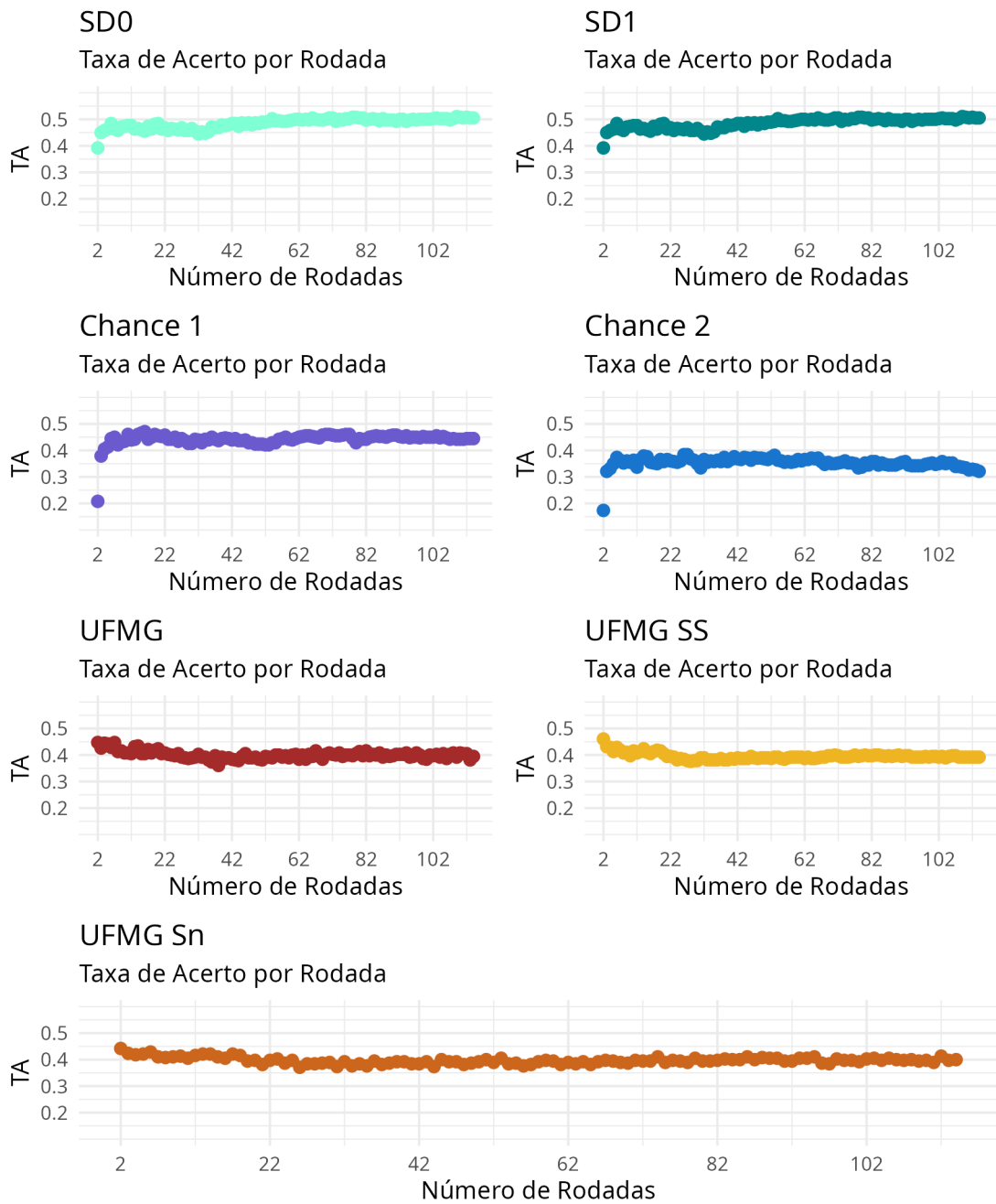
mais dependentes da quantidade de dados disponíveis.

Os gráficos por rodada revelam um comportamento consistente entre as métricas analisadas. A TA cresce rapidamente nas primeiras rodadas e, em seguida, estabiliza-se: as famílias *SD0* e *SDI* atingem valores médios entre 0,45 e 0,50, as variantes da *UFMG* situam-se em torno de 0,38 a 0,42 e os modelos *Chance* oscilam entre 0,33 e 0,39. A MD também apresenta forte redução de erro nas rodadas iniciais, passando depois a se estabilizar. A família *UFMG* mantém as menores e mais estáveis MDs, enquanto *Chance* e alguns modelos de Controle exibem maior variabilidade e piores calibrações.

As métricas derivadas da matriz de confusão reforçam esse padrão. O *F1-Score* melhora gradualmente para os modelos *SD* e *UFMG*, estabilizando-se após as primeiras rodadas, enquanto os modelos *Chance* apresentam quedas acentuadas, indicando desequilíbrio entre falsos positivos e falsos negativos quando a janela temporal é curta. O MCC segue a mesma tendência: *SD* e *UFMG* mantêm valores positivos e crescentes, ao passo que os modelos *Chance* permanecem próximos de zero ou levemente negativos.

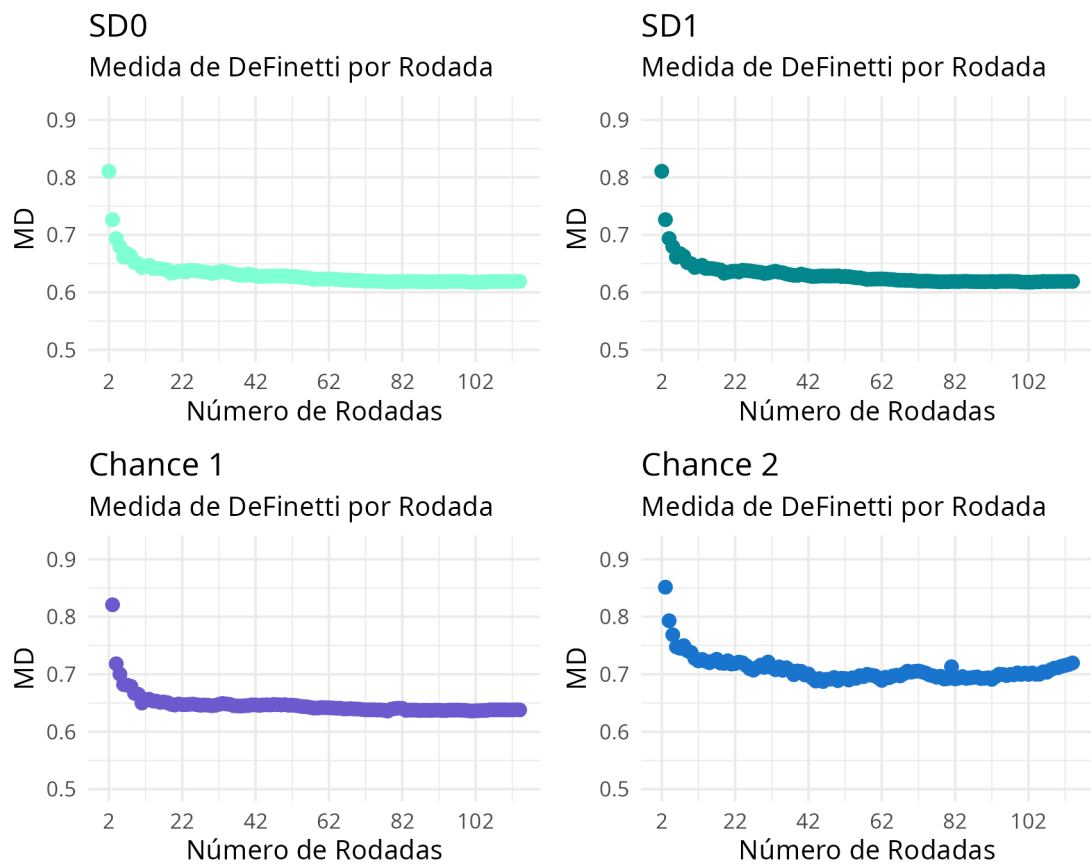
Esses resultados apontam duas conclusões principais. Primeiro, janelas temporais muito curtas, como a granularidade por rodada, produzem modelos instáveis, mal calibrados e pouco confiáveis, especialmente nas métricas probabilísticas. Segundo, modelos que incorporam uma quantidade maior de histórico, como as famílias *UFMG* e *SD*, convergem rapidamente para desempenhos estáveis, apresentando simultaneamente boa acurácia (TA), melhor calibração (MD baixa) e equilíbrio entre precisão e recall (F1 e MCC positivos). Assim, para aplicações que exigem consistência probabilística e robustez classificatória, as variantes *UFMG* e *SD* com janelas anuais, semestrais ou com uso da série histórica completa são as mais recomendadas. Os resultados da simulação podem ser observados nas Figuras 28, 29, 30, 31, 32 e 33.

Figura 28 – Resultados comparativos da taxa de acerto para o estudo de simulação nos modelos selecionados.



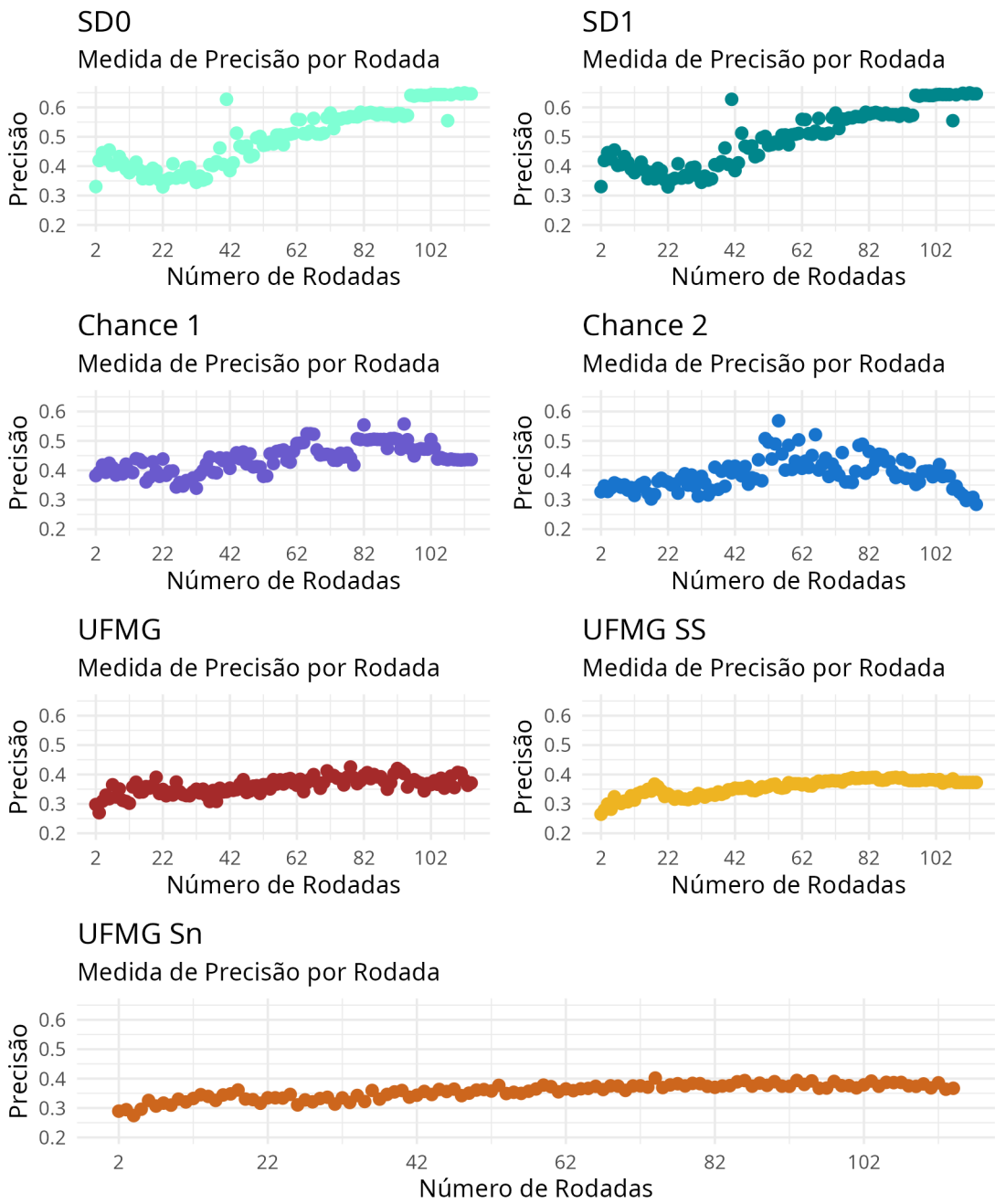
Fonte: Dos Autores

Figura 29 – Resultados comparativos da métrica MD para o estudo de simulação nos modelos selecionados.



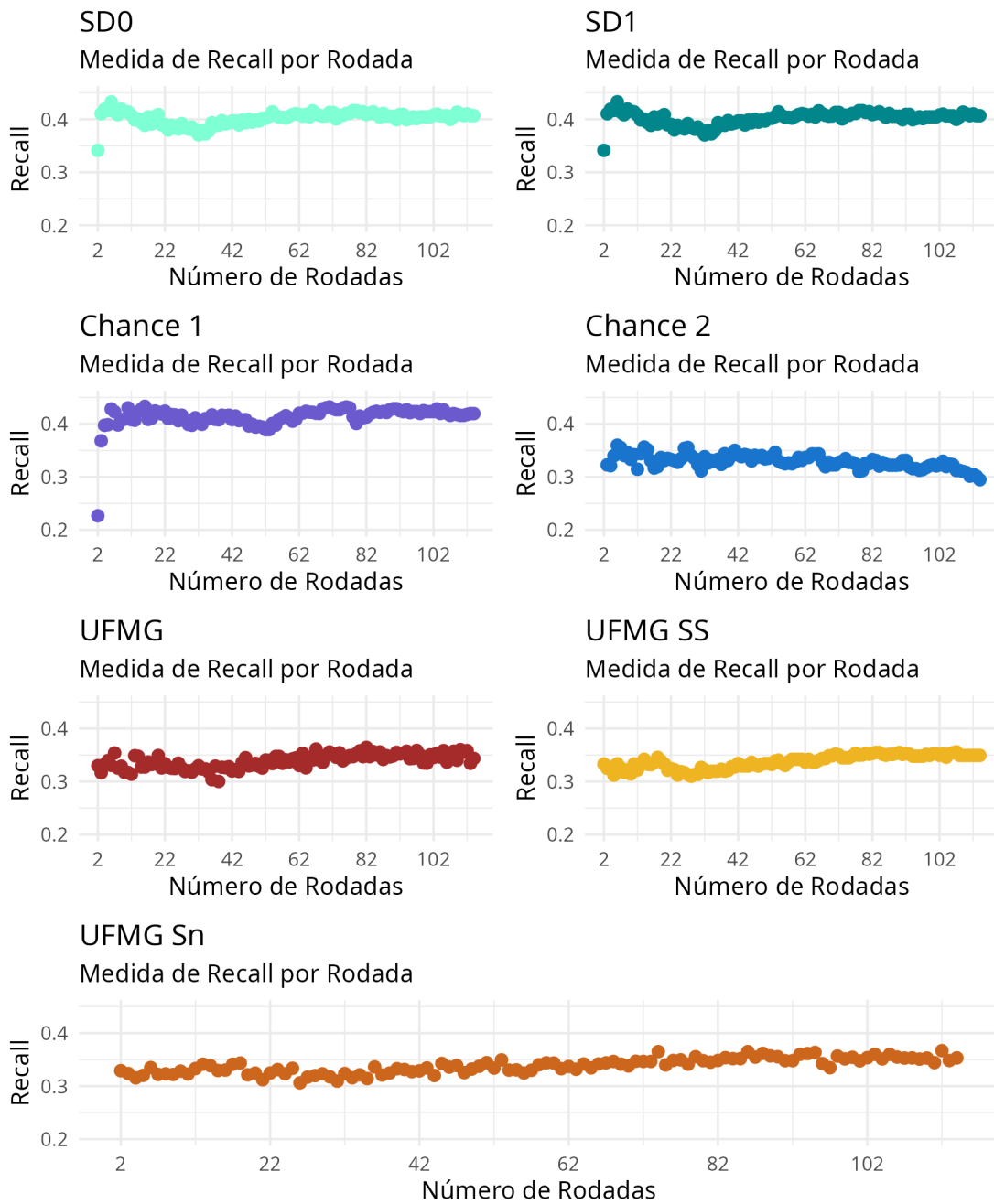
Fonte: Dos Autores

Figura 30 – Resultados comparativos da precisão para o estudo de simulação nos modelos selecionados.



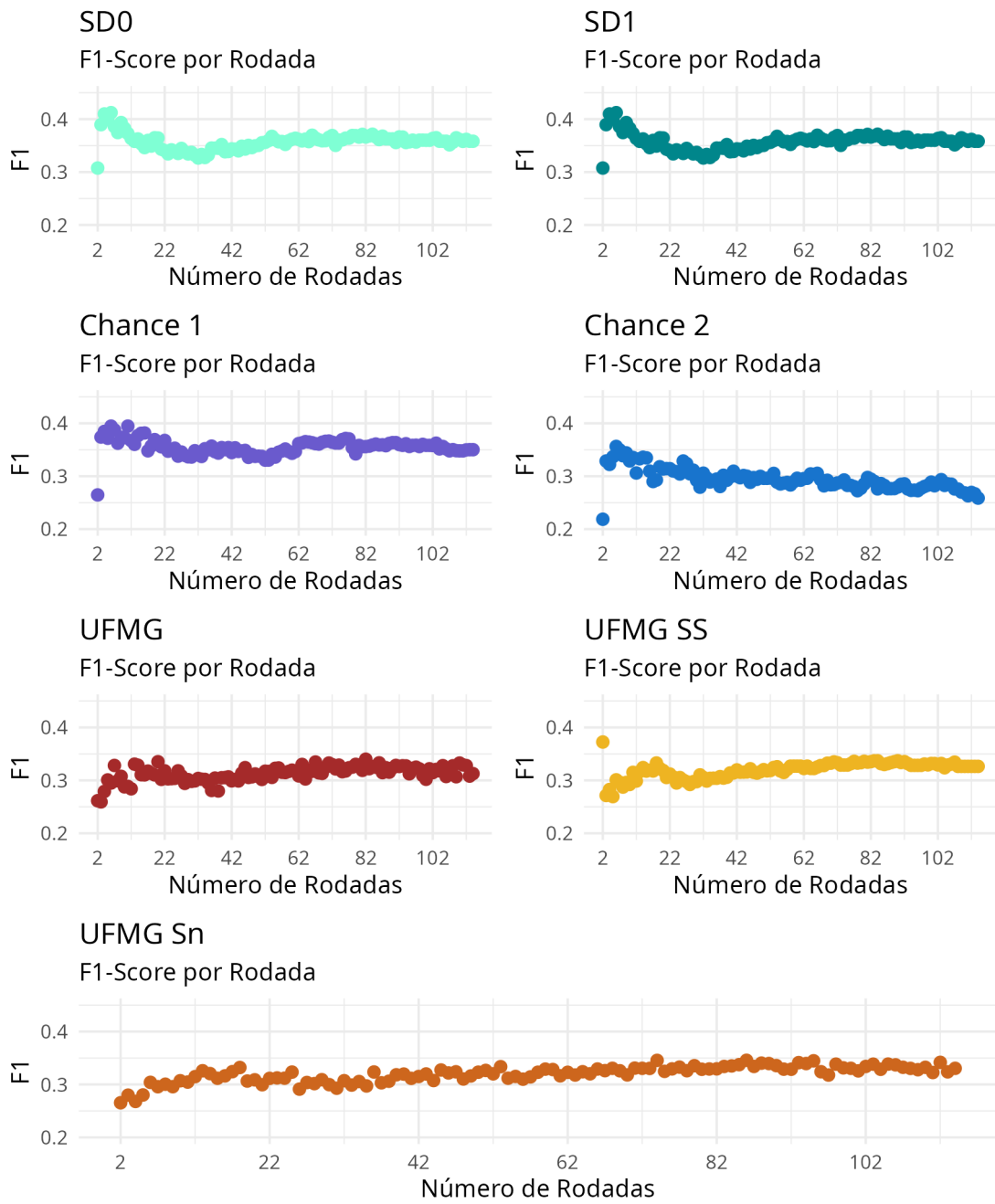
Fonte: Dos Autores

Figura 31 – Resultados comparativos do recall para o estudo de simulação nos modelos selecionados.



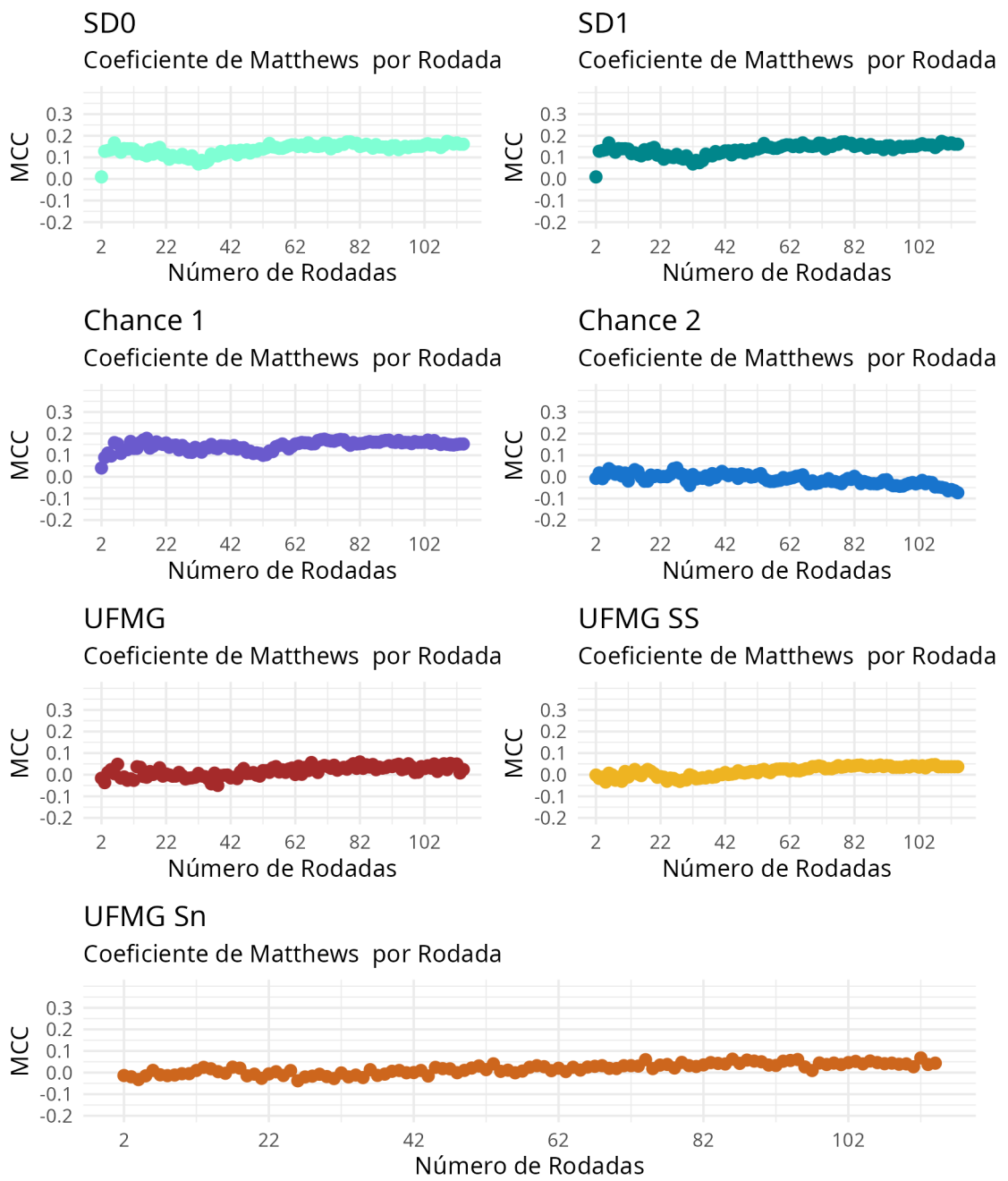
Fonte: Dos Autores

Figura 32 – Resultados comparativos do F1-score para o estudo de simulação nos modelos selecionados.



Fonte: Dos Autores

Figura 33 – Resultados comparativos do coeficiente de correlação de Matthews (MCC) para o estudo de simulação nos modelos selecionados.



Fonte: Dos Autores

## 6 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo comparar o desempenho de diferentes modelos preditivos aplicados ao Campeonato Brasileiro de Futebol, avaliando-os sob múltiplas métricas e distintas estruturas de janela temporal. A análise contemplou desde modelos consagrados na literatura, como o proposto pela Lima et al. (2012) e os métodos de Arruda (2000), até adaptações metodológicas e modelos de controle, totalizando 35 variações avaliadas.

Os resultados revelaram um cenário bastante monótono entre os modelos analisados. Tanto os métodos selecionados quanto as variações propostas não apresentaram efeitos significativos ou amplitudes expressivas nos resultados, permanecendo os valores concentrados em uma faixa pouco satisfatória em todas as métricas avaliadas. Esse comportamento sugere a existência de uma barreira metodológica, ou possivelmente inerente ao próprio contexto das previsões de resultados de futebol, que os modelos não foram capazes de ultrapassar, mesmo após adaptações. Além disso, tais adaptações não apontaram direções claras ou promissoras para superar essas limitações.

Outro fenômeno observado, e que pode causar certo incômodo a pesquisadores e leitores, foi o desempenho do modelo *ConTeiMan*. Apesar de sua extrema simplicidade, uma vez que sempre prevê vitória do mandante, esse modelo registrou a maior taxa de acerto entre todos os métodos avaliados. Tal resultado contraria a expectativa intuitiva de que agregar informações, estruturar modelos mais elaborados e fundamentados teoricamente seria sempre mais vantajoso do que manter um único palpite fixo ao longo de todas as partidas.

Apesar disso, foi possível identificar um grupo de modelos com desempenho superior nas diversas frentes de avaliação. O modelo *UfmgGer* destacou-se como a abordagem mais consistente: obteve TA igual a 0,491, foi o primeiro colocado no EPMP, terceiro melhor em MDD e registrou o maior F1-Score. Seu desempenho entre as diversas métricas o coloca como o melhor modelo do estudo.

Em segundo lugar, o modelo *SdIGer* apresentou resultados igualmente expressivos. Conforme sintetizado na Figura 27, foi o melhor modelo em verossimilhança, precisão e Medida de DeFinetti. Já o modelo *ConTeiMan* ocupou a terceira posição, pelos motivos já citados nesta seção. Mesmo que suas métricas de precisão não se destaquem, o equilíbrio entre metodologia e acurácia pesa a seu favor.

Vale ressaltar que, mesmo com a classificação e a composição desse pódio, os modelos

listados são muito semelhantes e não obtiveram grandes diferenças em seus resultados. Foram considerados critérios subjetivos, que levam em conta a metodologia e o histórico dos autores, e não apenas o desempenho nas métricas de avaliação. Isso não se restringe aos três modelos listados, pois a maioria dos modelos avaliados neste estudo apresentou resultados muito próximos e sujeitos à variabilidade.

Outro aspecto fundamental observado foi que as diferentes janelas de tempo, geral, anual, semestral e rodada, exerceram impacto reduzido sobre o desempenho. Apenas a janela por rodada apresentou efeitos negativos significativos, devido à baixa quantidade de informação disponível. As demais divisões temporais geraram resultados praticamente equivalentes, sugerindo que os modelos tendem a utilizar de maneira eficiente a informação histórica, mesmo quando agregada de forma mais sintética. No mesmo sentido, as variações propostas nos métodos — como as adaptações do modelo UFMG com e sem sorteio — não resultaram em melhorias consideráveis, reforçando a adequação da formulação original e favorecendo, por parcimônia, a versão sem sorteio.

O estudo de simulação também mostrou que o aumento progressivo do número de rodadas consideradas nas previsões tem impacto limitado. Após aproximadamente dez rodadas, as métricas tornam-se estáveis, indicando que a quantidade de informação adicional pouco altera o comportamento dos modelos. Esse achado tem implicações práticas relevantes, pois sugere que o desempenho preditivo converge rapidamente e que modelos estatísticos desse tipo não demandam longas séries temporais para atingirem seu potencial máximo.

Em síntese, os resultados apontam para a eficácia relativa dos modelos estudados, ao mesmo tempo em que evidenciam as limitações existentes. Embora nenhum dos métodos tenha alcançado a meta de ultrapassar 60% de taxa de acerto, foi possível identificar estratégias estatísticas mais coerentes, com destaque para as famílias UFMG e SD. A partir destes achados, futuras pesquisas poderão aprofundar o uso de abordagens bayesianas, modelos híbridos ou técnicas de aprendizado de máquina, além de expandir a modelagem probabilística incorporando características contextuais das partidas e dependências entre rodadas.

Os resultados deste trabalho contribuem para a literatura ao oferecer uma avaliação ampla e sistemática das principais metodologias de previsão aplicadas ao futebol brasileiro, fornecendo diretrizes sólidas para aprimoramentos futuros e para o desenvolvimento de modelos mais robustos, flexíveis e interpretáveis.

## 6.1 TRABALHOS FUTUROS

Nesta subseção, são destacados os ganhos e as questões levantadas ao longo do desenvolvimento deste trabalho que os autores consideram relevantes para investigações futuras. Esses pontos não foram explorados neste estudo por extrapolarem os objetivos propostos, mas podem ser aprofundados em pesquisas subsequentes.

Um desdobramento pertinente a este estudo, mas que não foi implementado a tempo da submissão deste documento, é a realização de simulações com os dados analisados, bem como a construção de intervalos de confiança para a análise de variância e a realização de testes estatísticos que possibilitem a diferenciação estatística dos resultados obtidos em cada métrica.

Destaca-se ainda a possibilidade de analisar com maior rigor as métricas de avaliação propostas neste trabalho, bem como outras medidas existentes na literatura, com o objetivo de avaliar o desempenho dessas métricas e selecionar um conjunto que ofereça parâmetros mais adequados para a comparação de modelos preditivos de resultados de partidas de futebol.

Outro aspecto que merece investigação diz respeito aos fatores que influenciam o resultado de uma partida. Assim como discutido em Bilek e Ulas (2019), é possível examinar o impacto do primeiro gol marcado em jogos do Campeonato Brasileiro de Futebol. Da mesma forma, pode-se analisar o peso do fator campo na definição de uma partida, observando esse efeito pela ótica de métricas como as apresentadas em Paludo, Figueiredo e Ferreira (2023).

Também se pode considerar a aplicação dos modelos aqui desenvolvidos ou utilizados em outras ligas de futebol, como a Champions League ou a La Liga, ou ainda em outros esportes. Além disso, os dados estatísticos provenientes da análise exploratória deste trabalho podem ser utilizados para gerar mapas e relacioná-los a indicadores sociais, permitindo discutir o papel sociocultural e econômico do futebol no Brasil.

Outro caminho possível consiste na comparação dos modelos estudados com métodos de aprendizado de máquina ou outras metodologias voltadas à área computacional, ampliando as abordagens utilizadas para previsão de resultados esportivos.

## REFERÊNCIAS

- Agência Estado. Cbf vê pedido de unificação dos títulos como "factive". **Gazeta do Povo**, nov. 2010. Acesso em 4 de março de 2025. Disponível em: <<https://www.gazetadopovo.com.br/esportes/cbf-ve-pedido-de-unificacao-dos-titulos-como-factive-09tcl0qd9gsp55am91jzwqn2m/>>.
- ARAÚJO, C. T. P. de et al. Modelagem estatística para a previsão de jogos de futebol: Uma aplicação no campeonato brasileiro de futebol 2014. **Revista da Estatística da Universidade Federal de Ouro Preto**, Ouro Preto, 2015.
- ARRUDA, M. L. d. **Poisson, Bayes, Futebol e DeFinetti**. 2000. Tese (Doutorado) — Universidade de São Paulo, São Paulo.
- ASTIVIA, O. L. O. Issues, problems and potential solutions when simulating continuous, non-normal data in the social sciences. **Meta-Psychology**, v. 4, 2020.
- BETING, M. Taça brasil é brasileiro? robertão é brasileiro? entenda. ou não. **Blog do Mauro Beting - UOL**, maio 2016. Acesso em 4 de março de 2025. Disponível em: <<https://maurobetting.blogosfera.uol.com.br/2016/05/31/taca-brasil-e-brasileirao-robertao-e-brasileirao-entenda-ou-nao/>>.
- BILEK, G.; ULAS, E. Predicting match outcome according to the quality of opponent in the english premier league using situational variables and team performance indicators. **International Journal of Performance Analysis in Sport**, Taylor & Francis, v. 19, n. 6, p. 930–941, 2019.
- BUNKER, R.; YEUNG, C.; FUJII, K. Machine learning for soccer match result prediction. In: **Artificial Intelligence, Optimization, and Data Sciences in Sports**. [S.l.]: Springer, 2024. p. 7–49.
- CAMPBELL, J. The poisson correlation function. **Proceedings of the Edinburgh Mathematical Society**, Cambridge University Press, v. 4, n. 1, p. 18–26, 1934.
- Confederação Brasileira de Futebol. **Campeonato Brasileiro Série A - Tabela Detalhada / Edição 2024**. 2024. Acesso em: 2 abr. 2025. Disponível em: <<https://objectstorage.sa-saopaulo-1.oraclecloud.com/n/grsa9ybqykir/b/portalcbf/o/Tabela%20Brasileir%C3%A3o%20S%C3%A9rie%20A%202024%20-%2006.12.24.pdf>>.
- Correio da Manhã. Inicia-se hoje o torneio dos campeões. **Correio da Manhã**, jan. 1937. Acesso em 4 de março de 2025.
- DE FINETTI, B. **Probability, Induction and Statistics: The Art of Guessing**. New York: John Wiley, 1972.

DUQUE, A. **Campeonato Brasileiro de Futebol**. 2023. Acesso em: 2 abr. 2025. Disponível em: <<https://www.kaggle.com/datasets/adaoduque/campeonato-brasileiro-de-futebol>>.

DURRETT, R. **Probability: Theory and Examples**. [S.l.]: Cambridge University Press, 2010.

ESPORTIVO, M. **Globo obtém média de 23 pontos de audiência em SP e no RJ com transmissão do Brasileirão**. 2023. <<https://maquinadoesporte.com.br/midia/globo-obtem-media-de-23-pontos-de-audiencia-em-sao-paulo-e-no-rio-com-transmissao-do-brasileirao/>>. Acesso em 2025-07-31.

ESTADÃO. Brasileirão: confira lista atualizada com todos os campeões. **Estadão**, dez. 2023. Acesso em 4 de março de 2025. Disponível em: <<https://www.estadao.com.br/esportes/futebol/brasileirao-confira-lista-atualizada-com-todos-os-campeoes/>>.

FURTADO, B. Cbf recebe dossiê dos 'campeões de 59 a 70' e clubes aguardam reconhecimento. **Superesportes**, nov. 2010. Acesso em 4 de março de 2025. Disponível em: <<https://www.uol.com.br/esporte/futebol/ultimas-noticias/2010/11/09/cruzeiro-vai-a-cbf-tentar-reconhecer-taca-brasil-como-titulo-nacional.jhtm>>.

GALVÃO, L. R.; BUENO FILHO, J. S. S. Estimação de rating no futebol: campeonato brasileiro de 2017. **Brazilian Journal of Biometrics**, v. 38, n. 1, p. 1–17, 2020.

GLOBO, E. E. . **Média de público do Brasileirão 2023 é a maior dos últimos 12 anos; Flamengo lidera rankings**. 2023. <<https://ge.globo.com/espiao-estatistico/noticia/2023/07/07/media-de-publico-do-brasileirao-2023-e-a-maior-dos-ultimos-12-anos-flamengo-lidera-rankings.ghtml>>. Acesso em 2025-07-31.

HAMDAN, M.; AL-BAYYATI, H. A note on the bivariate poisson distribution. **The American Statistician**, Taylor & Francis, v. 23, n. 4, p. 32–33, 1969.

HIRATA, E. Clube dos 13: ícone inacabado da modernização do futebol brasileiro (1980-2012). **EFDeportes.com, Revista Digital**, n. 181, jun 2013. Acesso em 4 de março de 2025. Disponível em: <<http://www.efdeportes.com/efd181/clubedos13.htm>>.

HO, L. L. **Análise de Contagens Multivariadas**. 1995. Tese (Doutorado) — Escola Politécnica, Universidade de São Paulo, São Paulo, Brasil, orientador: Prof. Dr. Pedro Luiz O. Costa Neto; Coorientador: Prof. Dr. Julio da Motta Singer.

HOLGATE, P. Estimation for the bivariate poisson distribution. **Biometrika**, Oxford University Press, v. 51, n. 1-2, p. 241–287, 1964.

JUNIOR, O. G. de S.; GAMERMAN, D. Previsão de partidas de futebol usando modelos dinâmicos. In: **XXXVI Simpósio Brasileiro de Pesquisa Operacional**. São João del-Rei: [s.n.], 2004. p. 650 – 659.

LIMA, B. N. B. d. et al. Probabilidades no esporte. **TRIM: revista de investigación multidisciplinar**, n. 5, p. 39–53, 2012.

MAGALHÃES, M. N. **Probabilidade e variáveis aleatórias**. São Paulo, SP: Edusp, 2006.

MARSHALL, A. W.; OLKIN, I. A family of bivariate distributions generated by the bivariate bernoulli distribution. **Journal of the American Statistical Association**, Taylor & Francis, v. 80, n. 390, p. 332–338, 1985.

MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. 2nd. ed. London: Chapman and Hall/CRC, 1989. (Monographs on Statistics and Applied Probability). ISBN 978-0412317606. Disponível em: <<https://www.taylorfrancis.com/books/mono/10.1201/9780203753736/generalized-linear-models-mccullagh>>.

MENDES-NEVES, T. et al. Forecasting soccer matches through distributions. **arXiv preprint arXiv:2501.05873**, 2025.

NASCIMENTO, J.; BRASIL, C. **Brasileirão tem segunda maior média de público da história**. 2024. <<https://www.cnnbrasil.com.br/esportes/futebol/brasileirao-tem-segunda-maior-media-de-publico-da-historia/>>. Segundo a CBF, mais de 9,6 milhões de torcedores nos estádios em 2024. Acesso em 2025-07-31.

NOGUEIRA, C. Os passos do futebol do país até chegar ao campeonato brasileiro. **Globo**, maio 2017. Acesso em 4 de março de 2025. Disponível em: <<https://ge.globo.com/blogs/especial-blog/memoria-ec/post/os-passos-do-futebol-brasileiro-ate-chegar-ao-campeonato-brasileiro.html>>.

PALUDO, G. F.; FIGUEIREDO, N. N. de; FERREIRA, E. B. Proposta de uma métrica para o efeito de casa baseada em pontos ganhos. **Ciência e Natura**, v. 45, 2023.

Posit team. **RStudio: Integrated Development Environment for R**. Boston, MA, 2025. Disponível em: <<http://www.posit.co/>>.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2024. Disponível em: <<https://www.R-project.org/>>.

RAMOS, L. F. P.; FERNANDES, H. C.; BATISTA, B. D. de O. Modelagem matemática para previsão de jogos de futebol. **ReviSeM**, v. 6, n. 1, p. 46–64, 2021.

RIBEIRO, F. É tri! cbf reconhece atlético-mg como campeão brasileiro de 1937. **Globo Esporte**, ago. 2023. Acesso em 4 de março de 2025. Disponível em: <<https://ge.globo.com/futebol/times/atlético-mg/noticia/2023/08/25/e-tri-cbf-reconhece-atletico-mg-como-campeao-brasileiro-de-1937.ghtml>>.

RICO-GONZÁLEZ, M. et al. Machine learning application in soccer: a systematic review. **Biology of sport**, Termedia, v. 40, n. 1, p. 249–263, 2023.

ROBINSON, J. Likert scale. In: \_\_\_\_\_. **Encyclopedia of Quality of Life and Well-Being Research**. Dordrecht: Springer Netherlands, 2014. p. 3620–3621. ISBN 978-94-007-0753-5. Disponível em: <[https://doi.org/10.1007/978-94-007-0753-5\\_1654](https://doi.org/10.1007/978-94-007-0753-5_1654)>.

ROSS, S. **A first course in probability**. Boston, MA: Pearson, 1976. v. 9.

SANTANA, H. et al. Modelagem estatística e de aprendizado de máquina: previsão do campeonato brasileiro série a 2017. **Matemática e Estatística em Foco**, v. 7, n. 1, p. 42–66, 2020.

STAFYLIDIS, A. et al. Decoding success: Predictive analysis of uefa euro 2024 to uncover key factors influencing soccer match outcomes. **Applied Sciences**, MDPI, v. 14, n. 17, p. 7740, 2024.

TEICHER, H. On the multivariate poisson distribution. **Scandinavian Actuarial Journal**, Taylor & Francis, v. 1954, n. 1, p. 1–9, 1954.

WICKHAM, H. et al. Welcome to the tidyverse. **Journal of Open Source Software**, v. 4, n. 43, p. 1686, 2019.

WONG, A. et al. A predictive analytics framework for forecasting soccer match outcomes using machine learning models. **Decision Analytics Journal**, Elsevier, v. 14, p. 100537, 2025.

## APÊNDICES

**APÊNDICE A** - Tabela de resultados quantitativos do Campeonato Brasileiro de Futebol Série A, dos anos de 2003 à 2024.

Tabela 6 – Desempenho absoluto dos times no Campeonato Brasileiro

Time	Nº jogos	Part	Vit	Emp	Der	Gols Feitos	Gols Sofridos
America-MG	228	6	58	61	109	227	332
America-RN	38	1	4	5	29	24	80
Athletico-PR	818	21	320	198	300	1072	1011
Athletico-GO	304	8	79	88	137	317	421
Athletico-MG	817	21	338	213	266	1178	1032
Avai	266	7	66	73	127	274	415
Bahia	464	12	142	129	193	529	614
Barueri	38	1	12	13	13	59	52
Botafogo-RJ	734	19	270	200	264	935	910
Bragantino	190	5	65	64	61	247	228
Brasiliense	42	1	10	12	20	49	68
Ceara	266	7	72	92	102	277	317
Chapecoense	265	7	70	77	118	261	359
Corinthians	818	21	341	240	237	1053	877
Coritiba	590	15	189	151	250	693	787
Criciuma	206	5	59	49	98	236	325
Cruzeiro	741	19	318	174	249	1055	890
CSA	38	1	8	8	22	24	58
Cuiaba	152	4	40	49	63	134	167
Figueirense	438	11	142	124	172	530	622
Flamengo	856	22	374	231	251	1230	1004
Fluminense	857	22	334	223	300	1131	1073
Fortaleza	354	9	127	90	137	427	460
Goiás	590	15	209	142	239	789	839
Gremio	776	20	328	194	254	1063	884
Gremio Prudente	38	1	7	10	21	39	64
Guarani	130	3	36	39	55	140	180
Internacional	818	21	357	213	248	1094	888
Ipatinga	38	1	9	8	21	37	67
Joinville	38	1	7	10	21	26	48
Juventude	324	8	96	88	140	381	499
Nautico	190	5	54	38	98	224	318
Palmeiras	772	20	351	197	224	1137	873
Parana	248	6	83	55	110	312	369
Paysandu	134	3	41	31	62	193	245
Ponte Preta	362	9	114	91	157	414	534
Portuguesa	114	3	31	38	45	137	157
Santa Cruz	76	2	15	14	47	86	145
Santo Andre	38	1	11	8	19	46	61
Santos	818	21	340	217	261	1178	982
Sao Caetano	172	4	65	41	66	209	199
Sao Paulo	856	22	382	237	237	1216	922
Sport	418	11	131	107	180	463	568
Vasco	666	17	223	193	250	832	938
Vitoria	434	11	136	107	191	537	633

Fonte: dos Autores.

**APÊNDICE B-** Tabela de resultados quantitativos do Campeonato Brasileiro de Futebol Série A, ponderada pelo número de jogos .

Tabela 7 – Desempenho dos times no Campeonato Brasileiro ponderado por jogo.

<b>Time</b>	<b>Nº jogos</b>	<b>Part</b>	<b>Vit_Med</b>	<b>Emp_Med</b>	<b>Der_Med</b>	<b>Gols_F</b>	<b>Gols_S</b>
América-MG	228	6	0,2544	0,2675	0,4781	0,9956	1,4561
América-RN	38	1	0,1053	0,1316	0,7632	0,6316	2,1053
Athletico-PR	818	21	0,3912	0,2421	0,3667	1,3105	1,2359
Atlético-GO	304	8	0,2599	0,2895	0,4507	1,0428	1,3849
Atlético-MG	817	21	0,4137	0,2607	0,3256	1,4419	1,2632
Avaí	266	7	0,2481	0,2744	0,4774	1,0301	1,5602
Bahia	464	12	0,3060	0,2780	0,4159	1,1401	1,3233
Barueri	38	1	0,3158	0,3421	0,3421	1,5526	1,3684
Botafogo-RJ	734	19	0,3678	0,2725	0,3597	1,2738	1,2398
Bragantino	190	5	0,3421	0,3368	0,3211	1,3000	1,2000
Brasiliense	42	1	0,2381	0,2857	0,4762	1,1667	1,6190
Ceará	266	7	0,2707	0,3459	0,3835	1,0414	1,1917
Chapecoense	265	7	0,2642	0,2906	0,4453	0,9849	1,3547
Corinthians	818	21	0,4169	0,2934	0,2897	1,2873	1,0721
Coritiba	590	15	0,3203	0,2559	0,4237	1,1746	1,3339
Criciúma	206	5	0,2864	0,2379	0,4757	1,1456	1,5777
Cruzeiro	741	19	0,4291	0,2348	0,3360	1,4238	1,2011
CSA	38	1	0,2105	0,2105	0,5789	0,6316	1,5263
Cuiabá	152	4	0,2632	0,3224	0,4145	0,8816	1,0987
Figueirense	438	11	0,3242	0,2831	0,3927	1,2100	1,4201
Flamengo	856	22	0,4369	0,2699	0,2932	1,4369	1,1729
Fluminense	857	22	0,3897	0,2602	0,3501	1,3197	1,2520
Fortaleza	354	9	0,3588	0,2542	0,3870	1,2062	1,2994
Goiás	590	15	0,3542	0,2407	0,4051	1,3373	1,4220
Grêmio	776	20	0,4227	0,2500	0,3273	1,3698	1,1392
Grêmio Prudente	38	1	0,1842	0,2632	0,5526	1,0263	1,6842
Guarani	130	3	0,2769	0,3000	0,4231	1,0769	1,3846
Internacional	818	21	0,4364	0,2604	0,3032	1,3374	1,0856
Ipatinga	38	1	0,2368	0,2105	0,5526	0,9737	1,7632
Joinville	38	1	0,1842	0,2632	0,5526	0,6842	1,2632
Juventude	324	8	0,2963	0,2716	0,4321	1,1759	1,5401
Náutico	190	5	0,2842	0,2000	0,5158	1,1789	1,6737
Palmeiras	772	20	0,4547	0,2552	0,2902	1,4728	1,1308
Palmeiras	772	20	0,4547	0,2552	0,2902	1,4728	1,1308
Paraná	248	6	0,3347	0,2218	0,4435	1,2581	1,4879
Paysandu	134	3	0,3060	0,2313	0,4627	1,4403	1,8284
Ponte Preta	362	9	0,3149	0,2514	0,4337	1,1436	1,4751
Portuguesa	114	3	0,2719	0,3333	0,3947	1,2018	1,3772
Santa Cruz	76	2	0,1974	0,1842	0,6184	1,1316	1,9079
Santo André	38	1	0,2895	0,2105	0,5000	1,2105	1,6053
Santos	818	21	0,4156	0,2653	0,3191	1,4401	1,2005
São Caetano	172	4	0,3779	0,2384	0,3837	1,2151	1,1570
São Paulo	856	22	0,4463	0,2769	0,2769	1,4206	1,0771
Sport	418	11	0,3134	0,2560	0,4306	1,1077	1,3589
Vasco	666	17	0,3348	0,2898	0,3754	1,2492	1,4084
Vitória	434	11	0,3134	0,2465	0,4401	1,2373	1,4585

Fonte: dos Autores.

**APÊNDICE C-** Tabela de resultados das métricas calculadas para cada um dos modelos de predição.

Tabela 8 – Resultados Geral das Métricas de avaliação dos modelos

Modelo	TA	EPMP	MD	MDD	NC	TF	Vero	MC_Pre	MC_Rec	MC_F1	MC_MCC
UfmgSnRod	0.4678	0.2133	0.6437	NA	3.286	1	NA	0.3409	0.3373	0.2785	0.0123
UfmgSnSem	0.4372	0.2401	0.6467	NA	3.429	1	NA	0.3427	0.3378	0.3166	0.012
UfmgSnGer	0.4046	0.2724	0.6598	NA	3.714	1	NA	0.3641	0.3457	0.3207	0.0251
UfmgSnAno	0.4269	0.249	0.6488	NA	3.571	1	NA	0.3538	0.3422	0.3271	0.0182
UfmgSsRod	0.4789	0.2092	0.6436	NA	2.714	1	NA	0.3525	0.3366	0.2619	0.0108
UfmgSsSem	0.4409	0.2382	0.6459	NA	2.857	1	NA	0.3503	0.3405	0.3189	0.0158
UfmgSsGer	0.4036	0.2722	0.6593	NA	3.143	1	NA	0.3613	0.3436	0.3179	0.0217
UfmgSsAno	0.4272	0.2484	0.6487	NA	3	1	NA	0.3533	0.3414	0.3259	0.017
UfmgRod	0.4766	0.2107	0.6424	NA	3.143	1	NA	0.3362	0.3338	0.2597	9e-04
UfmgSem	0.438	0.2394	0.6461	NA	3.286	1	NA	0.3445	0.3387	0.3177	0.013
UfmgGer	0.4907	0.2017	0.642	NA	3.571	1	NA	0.2606	0.3367	0.3651	0.0166
UfmgAno	0.4289	0.2468	0.6479	NA	3.429	1	NA	0.3515	0.3408	0.325	0.0168
Sd1Rod	0.3783	0.3647	0.8195	2.8025	3.571	1	0.0401	0.3319	0.335	0.3049	0.0065
Sd1Sem	0.4299	0.3309	0.7348	1.7123	3.714	1	0.071	0.3596	0.3605	0.3435	0.0476
Sd1Ger	0.4777	0.2609	0.6349	1.5172	4	1	0.0742	0.4379	0.354	0.2963	0.0481
Sd1Ano	0.4441	0.3039	0.6924	1.6341	3.857	1	0.0727	0.3627	0.3609	0.3327	0.0489
Sd0Rod	0.3783	0.3647	0.8195	2.806	3.143	1	0.0401	0.3319	0.335	0.3049	0.0065
Sd0Sem	0.4299	0.3309	0.7348	1.7147	3.286	1	0.071	0.3596	0.3605	0.3435	0.0476
Sd0Ger	0.4777	0.2609	0.6349	1.5174	3.571	1	0.0742	0.4379	0.354	0.2963	0.0481
Sd0Ano	0.4441	0.3039	0.6924	1.6363	3.429	1	0.0727	0.3627	0.3609	0.3327	0.0489
Ch2Rod	0.1785	0.3339	1.0831	2.2392	3.714	0.554	0.0461	0.3352	0.1854	0.2119	8e-04
Ch2Sem	0.3723	0.3025	0.7508	1.8743	3.857	0.9673	0.0685	0.3426	0.3308	0.3343	0.0103
Ch2Ger	0.3714	0.2883	0.6929	1.5844	4.143	0.9995	0.069	0.3266	0.3272	0.281	-0.0099
Ch2Ano	0.3854	0.2849	0.7212	1.7667	4	0.9822	0.0677	0.3388	0.3281	0.3194	-0.0019
Ch1Rod	0.1692	0.329	1.079	2.2556	3.714	0.5375	0.0442	0.3358	0.1815	0.2184	0.0024
Ch1Sem	0.366	0.3246	0.7514	1.8028	3.857	0.9628	0.068	0.3533	0.3445	0.3223	0.0353
Ch1Ger	0.3976	0.2711	0.6693	1.5157	4.143	0.9994	0.0708	0.3537	0.3657	0.3009	0.0559
Ch1Ano	0.3804	0.31	0.7179	1.7094	4	0.9839	0.0696	0.3553	0.3571	0.3185	0.0454
ConUnif	0.3333	0.4066	0.8308	NA	1.571	1	NA	0.3325	0.332	0.3233	-0.0012
ConPois	0.3349	NA	NA	1.891	1.714	1	NA	0.3317	0.334	0.3248	-0.0027
ConMed	0.307	NA	NA	1.4644	1.857	1	NA	0.3579	0.3442	0.279	0.023
ConTeiVis	0.2395	0.7605	1.521	NA	1	1	NA	NA	NA	NA	NA
ConTeiEmp	0.2642	0.7358	1.4716	NA	1	1	NA	NA	NA	NA	NA
ConTeiMan	0.4963	0.5037	1.0074	NA	1	1	NA	NA	NA	NA	NA
ConImp	0.3924	0.2566	0.6626	1.4387	2.714	1	0.068	0.3723	0.3575	0.2931	0.0428

**APÊNDICE D-** Tabela dos valores atribuídos para cada critério do Nível de Complexidade

Tabela 9 – valores dos critérios de avaliação do Nível de Complexidade por modelo

<b>Modelo</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>	<b>NC</b>
ConTeiMan	1	1	1	1	1	1	1	1,0000
ConTeiEmp	1	1	1	1	1	1	1	1,0000
ConTeiVis	1	1	1	1	1	1	1	1,0000
ConUnif	2	2	1	2	1	2	1	1,5714
ConPois	2	2	1	2	2	2	2	1,7143
ConMed	1	2	5	1	2	2	2	1,8571
ConImp	2	1	2	5	1	2	2	2,7143
UfmgSsRod	3	1	1	5	1	1	3	2,7143
UfmgSsSem	3	1	1	5	1	1	3	2,7143
UfmgSsAno	3	1	1	5	1	1	4	3,0000
Sd0Rod	3	4	2	4	3	3	3	3,1429
UfmgRod	3	2	1	5	1	2	3	3,1429
UfmgSsGer	1	2	1	5	2	2	4	3,1429
Sd0Sem	3	4	2	4	3	3	3	3,2857
UfmgSem	3	2	1	5	1	2	4	3,2857
UfmgSnRod	3	2	1	5	1	2	4	3,2857
Sd0Ano	3	3	2	4	3	3	3	3,4286
UfmgAno	3	2	1	5	2	2	4	3,4286
UfmgSnSem	3	2	1	5	2	2	4	3,4286
Sd0Ger	3	4	2	4	3	3	3	3,5714
Sd1Rod	3	5	3	4	4	3	4	3,5714
UfmgGer	3	2	1	5	2	2	5	3,5714
UfmgSnAno	3	3	3	4	3	3	3	3,5714
Ch1Rod	3	3	4	4	4	3	4	3,7143
Sd1Sem	3	5	3	4	3	3	4	3,7143
UfmgSnGer	3	3	3	4	3	3	4	3,7143
Ch1Sem	3	4	5	4	4	3	4	3,8571
Ch2Sem	3	5	4	3	4	3	4	3,8571
Sd1Ano	3	5	3	4	4	3	4	3,8571
Ch1Ano	3	4	3	5	4	4	4	4,0000
Ch2Ano	3	4	4	4	4	4	4	4,0000
Ch1Ger	3	4	5	4	5	3	4	4,0000
Ch2Ger	3	5	4	4	5	4	4	4,1429

Fonte: Dos Autores

**APÊNDICE E** - Tabelas de resultado para o teste de Kruskal–Wallis e tabela de Dunn post-hoc para TA (Memória).

Tabela 10 – Resultados do teste de Kruskal–Wallis para as classes de Memória

<b>Métrica</b>	<b>H</b>	<b>p-valor</b>	<b>Conclusão</b>
TA	3.205	0.5242	Não significativo
MD	11.867	0.0184	Significativo
MC <sub>F</sub> 1	18.256	0.0011	Significativo

Tabela 11 – Dunn (Holm) post-hoc — TA — fator Memória

<b>Grupo 1</b>	<b>Grupo 2</b>	<b>z</b>	<b>p-raw</b>	<b>p-holm</b>	<b>Signif.</b>
Anual	Geral	0.120	0.9046	0.9046	Não
Anual	Nula	1.548	0.1217	0.3652	Não
Anual	Rodada	0.626	0.5313	1.0000	Não
Anual	Semestre	0.104	0.9169	1.0000	Não
Geral	Nula	1.516	0.1294	0.3873	Não
Geral	Rodada	0.544	0.5864	1.0000	Não
Geral	Semestre	-0.009	0.9926	0.9926	Não
Nula	Rodada	-0.976	0.3290	0.9850	Não
Nula	Semestre	-1.452	0.1464	0.4392	Não
Rodada	Semestre	-0.522	0.6019	1.0000	Não

Tabela 12 – Dunn (Holm) post-hoc — MD — fator Memória.

<b>Grupo 1</b>	<b>Grupo 2</b>	<b>z</b>	<b>p-raw</b>	<b>p-holm</b>	<b>Signif.</b>
Anual	Geral	-0.449	0.6530	1.0000	Não
Anual	Nula	-2.463	0.0138	0.0828	Não
Anual	Rodada	-0.625	0.5313	1.0000	Não
Anual	Semestre	-0.104	0.9169	1.0000	Não
Geral	Nula	-3.399	0.0007	0.0068	Sim
Geral	Rodada	-0.544	0.5864	1.0000	Não
Geral	Semestre	-0.009	0.9926	0.9926	Não
Nula	Rodada	0.976	0.3290	0.9850	Não
Nula	Semestre	1.452	0.1464	0.4392	Não
Rodada	Semestre	0.522	0.6019	1.0000	Não

Tabela 13 – Dunn (Holm) post-hoc — MC\_F1 — fator Memória.

<b>Grupo 1</b>	<b>Grupo 2</b>	<b>z</b>	<b>p-raw</b>	<b>p-holm</b>	<b>Signif.</b>
Anual	Geral	0.794	0.4274	1.0000	Não
Anual	Nula	0.328	0.7430	1.0000	Não
Anual	Rodada	2.026	0.0429	0.3855	Não
Anual	Semestre	0.187	0.8511	1.0000	Não
Geral	Nula	-0.466	0.6419	1.0000	Não
Geral	Rodada	2.668	0.0076	0.0684	Não
Geral	Semestre	-0.393	0.6947	1.0000	Não
Nula	Rodada	2.341	0.0192	0.2304	Não
Nula	Semestre	-0.142	0.8879	1.0000	Não
Rodada	Semestre	-3.540	0.0004	0.0040	Sim

**APÊNDICE F** - Tabelas de resultado para o teste de Kruskal–Wallis (Família).

Tabela 14 – Resultados do teste de Kruskal–Wallis para as classes de Família

<b>Métrica</b>	<b>H</b>	<b>p-valor</b>	<b>Conclusão</b>
TA	18.070	<b>0.0118</b>	Significativo
MD	10.570	0.1033	Não significativo
MC <sub>F1</sub>	2.388	0.9353	Não significativo

APÊNDICE G - Tabelas de Dunn post-hoc para (Família).

Tabela 15 – Dunn (Holm) post-hoc — TA — fator Família

Grupo 1	Grupo 2	z	p-raw	p-holm	Signif.
Chance I	Chance II	-0.188	0.8505	1.0000	Não
Chance I	Controle	-0.301	0.7630	1.0000	Não
Chance I	SD 0	-2.393	0.0167	0.2315	Não
Chance I	SD 1	-2.456	0.0140	0.1946	Não
Chance I	UFMG original	0.624	0.5328	1.0000	Não
Chance I	UFMG sem sor.	-0.452	0.6511	1.0000	Não
Chance I	UFMG sor. normal	-0.528	0.5977	1.0000	Não
Chance II	Controle	0.113	0.9099	1.0000	Não
Chance II	SD 0	-2.612	0.0090	0.0900	Não
Chance II	SD 1	-2.728	0.0064	0.0640	Não
Chance II	UFMG original	0.339	0.7345	1.0000	Não
Chance II	UFMG sem sor.	-0.264	0.7919	1.0000	Não
Chance II	UFMG sor. normal	-0.339	0.7345	1.0000	Não
Controle	SD 0	-2.881	0.0039	<b>0.0369</b>	Sim
Controle	SD 1	-3.179	0.0015	<b>0.0149</b>	Sim
Controle	UFMG original	0.226	0.8211	1.0000	Não
Controle	UFMG sem sor.	-0.151	0.8802	1.0000	Não
Controle	UFMG sor. normal	-0.226	0.8211	1.0000	Não
SD 0	SD 1	0.000	1.0000	1.0000	Não
SD 0	UFMG original	0.226	0.8211	1.0000	Não
SD 0	UFMG sem sor.	0.414	0.6785	1.0000	Não
SD 0	UFMG sor. normal	0.565	0.5718	1.0000	Não
SD 1	UFMG original	0.226	0.8211	1.0000	Não
SD 1	UFMG sem sor.	0.339	0.7345	1.0000	Não
SD 1	UFMG sor. normal	0.566	0.5718	1.0000	Não
UFMG original	UFMG sem sor.	-0.414	0.6785	1.0000	Não
UFMG original	UFMG sor. normal	-0.339	0.7345	1.0000	Não
UFMG sem sor.	UFMG sor. normal	-0.075	0.9399	1.0000	Não

Tabela 16 – Dunn (Holm) post-hoc — MD — fator Familia

<b>Grupo 1</b>	<b>Grupo 2</b>	<b>z</b>	<b>p-raw</b>	<b>p-holm</b>	<b>Signif.</b>
Chance I	Chance II	-0.113	0.9099	1.0000	Não
Chance I	Controle	-0.301	0.7630	1.0000	Não
Chance I	SD 0	-0.578	0.5632	1.0000	Não
Chance I	SD 1	-0.528	0.5977	1.0000	Não
Chance I	UFMG original	-0.226	0.8211	1.0000	Não
Chance I	UFMG sem sor.	0.301	0.7630	1.0000	Não
Chance I	UFMG sor. normal	0.226	0.8211	1.0000	Não
Chance II	Controle	0.452	0.6511	1.0000	Não
Chance II	SD 0	-0.339	0.7345	1.0000	Não
Chance II	SD 1	-0.301	0.7630	1.0000	Não
Chance II	UFMG original	-0.264	0.7919	1.0000	Não
Chance II	UFMG sem sor.	0.263	0.7919	1.0000	Não
Chance II	UFMG sor. normal	0.113	0.9099	1.0000	Não
Controle	SD 0	0.226	0.8211	1.0000	Não
Controle	SD 1	0.226	0.8211	1.0000	Não
Controle	UFMG original	0.339	0.7345	1.0000	Não
Controle	UFMG sem sor.	0.151	0.8802	1.0000	Não
Controle	UFMG sor. normal	0.226	0.8211	1.0000	Não
SD 0	SD 1	0.000	1.0000	1.0000	Não
SD 0	UFMG original	-0.226	0.8211	1.0000	Não
SD 0	UFMG sem sor.	-0.414	0.6785	1.0000	Não
SD 0	UFMG sor. normal	-0.565	0.5718	1.0000	Não
SD 1	UFMG original	-0.339	0.7345	1.0000	Não
SD 1	UFMG sem sor.	-0.339	0.7345	1.0000	Não
SD 1	UFMG sor. normal	-0.566	0.5718	1.0000	Não
UFMG original	UFMG sem sor.	0.414	0.6785	1.0000	Não
UFMG original	UFMG sor. normal	0.339	0.7345	1.0000	Não
UFMG sem sor.	UFMG sor. normal	0.075	0.9399	1.0000	Não

Tabela 17 – Dunn (Holm) post-hoc — MC\_F1 — fator Família

<b>Grupo 1</b>	<b>Grupo 2</b>	<b>z</b>	<b>p-raw</b>	<b>p-holm</b>	<b>Signif.</b>
Chance I	Chance II	-0.188	0.8505	1.0000	Não
Chance I	Controle	-0.301	0.7630	1.0000	Não
Chance I	SD 0	-0.452	0.6511	1.0000	Não
Chance I	SD 1	-0.528	0.5977	1.0000	Não
Chance I	UFMG original	0.624	0.5328	1.0000	Não
Chance I	UFMG sem sor.	-0.263	0.7919	1.0000	Não
Chance I	UFMG sor. normal	-0.527	0.5977	1.0000	Não
Chance II	Controle	0.113	0.9099	1.0000	Não
Chance II	SD 0	-0.188	0.8505	1.0000	Não
Chance II	SD 1	-0.226	0.8211	1.0000	Não
Chance II	UFMG original	-0.339	0.7345	1.0000	Não
Chance II	UFMG sem sor.	-0.142	0.8879	1.0000	Não
Chance II	UFMG sor. normal	-0.339	0.7345	1.0000	Não
Controle	SD 0	-0.789	0.4292	1.0000	Não
Controle	SD 1	-0.789	0.4292	1.0000	Não
Controle	UFMG original	0.226	0.8211	1.0000	Não
Controle	UFMG sem sor.	-0.327	0.7430	1.0000	Não
Controle	UFMG sor. normal	-0.332	0.7397	1.0000	Não
SD 0	SD 1	0.000	1.0000	1.0000	Não
SD 0	UFMG original	0.414	0.6785	1.0000	Não
SD 0	UFMG sem sor.	0.795	0.4266	1.0000	Não
SD 0	UFMG sor. normal	0.986	0.3235	1.0000	Não
SD 1	UFMG original	0.566	0.5718	1.0000	Não
SD 1	UFMG sem sor.	0.795	0.4266	1.0000	Não
SD 1	UFMG sor. normal	0.986	0.3235	1.0000	Não
UFMG original	UFMG sem sor.	-0.414	0.6785	1.0000	Não
UFMG original	UFMG sor. normal	-0.566	0.5718	1.0000	Não
UFMG sem sor.	UFMG sor. normal	-0.075	0.9399	1.0000	Não