

PREVISÃO PROBABILÍSTICA DA PREMIER LEAGUE COM MODELOS DE POISSON

Matheus Farias Sant'Anna¹, Igor Campos de Almeida Lima²

¹Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil
(matheusfariassantanna99@gmail.com)

Resumo: Comparam-se o modelo de Poisson independente, o modelo de Poisson bivariado e o modelo Dixon-Coles na previsão 1X2 da Premier League 2024/2025. Usaram-se 304 jogos no ajuste e 76 no teste. O Dixon-Coles reduziu as perdas em menos de 0,1%, sem diferença significativa nos testes pareados. O modelo de Poisson independente mostrou-se a alternativa mais parcimoniosa.

Palavras-chave: Modelos de Poisson; modelo de Poisson bivariado; modelo Dixon-Coles; previsão probabilística de futebol; Premier League.

INTRODUÇÃO

A previsão de partidas de futebol pode ser formulada a partir das contagens de gols do mandante e do visitante. A distribuição conjunta dos placares permite obter probabilidades de vitória do mandante, empate e vitória do visitante, preservando informação que se perde quando a análise se limita à classe de maior probabilidade. Modelos de Poisson são recorrentes nesse contexto por combinarem interpretação direta, estimação relativamente simples e probabilidades para todos os placares possíveis (Maher, 1982; Lee, 1997; Goddard, 2005).

A especificação independente representa as intensidades de gols por efeitos de ataque, vulnerabilidade defensiva e mando. O modelo de Poisson bivariado acrescenta um componente compartilhado, cuja intensidade determina covariância não negativa entre as contagens. O modelo Dixon-Coles modifica localmente as probabilidades de 0-0, 1-0, 0-1 e 1-1, mantendo as marginais de Poisson (Dixon e Coles, 1997; Karlis e Ntzoufras, 2003). A flexibilidade adicional somente se justifica quando produz ganho preditivo fora da amostra.

A avaliação probabilística requer regras de pontuação próprias. Log-loss, Brier Score e Ranked Probability Score (RPS) utilizam o vetor completo de probabilidades, enquanto a acurácia registra apenas a coincidência entre o resultado observado e a classe mais provável. A validação deve respeitar a ordem temporal das partidas para impedir o uso de informação futura na estimação (Gneiting e Raftery, 2007; Constantinou e Fenton, 2012; Tsokos et al., 2019).

O objetivo foi comparar o modelo de Poisson independente, o modelo de Poisson bivariado e o modelo Dixon-Coles na Premier League 2024/2025. Foram examinados desempenho fora da amostra, diferença pareada entre Dixon-Coles e Poisson independente, calibração, significância contra uma hipótese nula por permutação e parâmetros estimados das equipes.

MATERIAL E MÉTODOS

A base de dados reuniu as 380 partidas da Premier League 2024/2025, identificada como divisão E0 no portal Football-Data. Foram utilizadas a data, as equipes mandante e visitante, os gols marcados e o resultado final. Após a ordenação cronológica, as primeiras 304 partidas (80%) constituíram o conjunto de treinamento e as 76 restantes (20%), o conjunto de teste. Não foram incluídas informações produzidas durante ou após os jogos, como finalizações, posse de bola ou expulsões. Essa delimitação preservou uma avaliação prospectiva e permitiu atribuir as diferenças de desempenho às estruturas probabilísticas comparadas (FOOTBALL-DATA.CO.UK, s.d.; Tsokos et al., 2019).

Para a partida m , consideraram-se as contagens de gols do mandante e do visitante. No modelo de Poisson independente, essas variáveis são condicionalmente independentes, dadas as intensidades específicas do confronto. As Equações (1) e (2) definem as respectivas distribuições marginais.

$$G_{H,m} \sim \text{Poisson}(\lambda_{H,m}). \quad (1)$$

$$G_{A,m} \sim \text{Poisson}(\lambda_{A,m}). \quad (2)$$

Nas Equações (1) e (2), os parâmetros de intensidade representam os números esperados de gols do mandante e do visitante. Como essas intensidades

dependem das equipes envolvidas e da condição de mando, empregou-se uma parametrização log-linear. Para uma partida entre uma equipe mandante e uma visitante, as intensidades são especificadas pelas Equações (3) e (4).

$$\log \lambda_{H,m} = \mu + h + \alpha_{i_m} + \delta_{j_m}. \quad (3)$$

$$\log \lambda_{A,m} = \mu + \alpha_{j_m} + \delta_{i_m}. \quad (4)$$

Nas Equações (3) e (4), μ representa o nível médio de gols da competição; h , a vantagem média de mando; α , a força ofensiva específica de cada equipe; e δ , sua vulnerabilidade defensiva. A ligação logarítmica garante intensidades positivas e permite interpretar os efeitos na escala multiplicativa. A estimação utilizou uma equipe de referência; para a apresentação, os efeitos de ataque e defesa foram recentrados para soma nula, com compensação do intercepto, sem alterar as intensidades ou as probabilidades previstas (Maher, 1982; Lee, 1997; Goddard, 2005).

Para admitir dependência entre as duas contagens, ajustou-se o modelo de Poisson bivariado. A Equação (5) representa os gols como soma de componentes específicos e de um componente compartilhado, enquanto a Equação (6) explicita a covariância decorrente dessa construção.

$$G_{H,m} = U_{1,m} + U_{3,m}, \quad G_{A,m} = U_{2,m} + U_{3,m}. \quad (5)$$

$$\text{Cov}(G_{H,m}, G_{A,m}) = \lambda_3, \quad \lambda_3 \geq 0. \quad (6)$$

Na Equação (5), U_1 , U_2 e U_3 são componentes de Poisson mutuamente independentes. Os dois primeiros correspondem às parcelas específicas de cada equipe, e U_3 participa simultaneamente das duas contagens. Sua intensidade λ_3 é, portanto, a covariância indicada na Equação (6). Como $\lambda_3 \geq 0$, a formulação clássica representa apenas dependência não negativa; quando λ_3 se aproxima de zero, o modelo converge ao caso independente (Karlis e Ntzoufras, 2003; McHale e Scarf, 2007).

O modelo Dixon-Coles trata a dependência de forma localizada. A Equação (7) multiplica a massa conjunta da Poisson independente pelo fator τ , cuja definição por partes é apresentada na Equação (8). A correção atua diretamente apenas nos placares 0-0, 0-1, 1-0 e 1-1 (Dixon e Coles, 1997).

$$P_{DC}(x, y) = \tau(x, y; \lambda_H, \lambda_A, \rho) P_{Pois}(x; \lambda_H) P_{Pois}(y; \lambda_A). \quad (7)$$

$$\tau(x, y) = \begin{cases} 1 - \lambda_H \lambda_A \rho, & x = 0, y = 0, \\ 1 + \lambda_H \rho, & x = 0, y = 1, \\ 1 + \lambda_A \rho, & x = 1, y = 0, \\ 1 - \rho, & x = 1, y = 1, \\ 1, & \text{nos demais placares.} \end{cases} \quad (8)$$

Nas Equações (7) e (8), x e y são os gols do mandante e do visitante, e ρ controla a redistribuição de probabilidade entre os quatro placares de baixa contagem. Quando $\rho = 0$, $\tau = 1$ e o modelo coincide com a Poisson independente. Mantidas as

intensidades, valores negativos de ρ aumentam as massas de 0-0 e 1-1 e reduzem as de 0-1 e 1-0; valores positivos produzem o movimento inverso. Nos demais placares, τ permanece igual a um.

Os parâmetros dos três modelos foram estimados por máxima verossimilhança exclusivamente no treinamento. Para cada partida do teste, calculou-se a distribuição conjunta dos placares em uma grade de zero a dez gols por equipe; as probabilidades de vitória do mandante, empate e vitória do visitante foram obtidas pela soma das massas correspondentes e normalizadas para totalizar um. A avaliação probabilística utilizou as medidas definidas nas Equações (9), (10) e (11).

$$\text{LogLoss} = -\frac{1}{N} \sum_{m=1}^N \log p_{m,y_m}. \quad (9)$$

$$\text{Brier} = \frac{1}{N} \sum_{m=1}^N \sum_{c \in \{H,D,A\}} (p_{m,c} - o_{m,c})^2. \quad (10)$$

$$\text{RPS}_m = \frac{1}{K-1} \sum_{k=1}^{K-1} (F_{m,k} - O_{m,k})^2. \quad (11)$$

Na Equação (9), N é o número de partidas de teste e o termo dentro do logaritmo corresponde à probabilidade atribuída ao resultado observado; a log-loss penaliza intensamente previsões excessivamente confiantes e incorretas. Na Equação (10), comparam-se as probabilidades das três classes com os indicadores do resultado observado; adotou-se o Brier Score multiclasse sem divisão pelo número de classes. Na Equação (11), comparam-se as distribuições acumuladas prevista e observada segundo a ordem H-D-A. Valores menores indicam melhor desempenho nas três perdas. A acurácia foi reportada como medida complementar da decisão pela classe de maior probabilidade, sem substituir a avaliação do vetor probabilístico (Brier, 1950; Epstein, 1969; Gneiting e Raftery, 2007).

A comparação direta entre o modelo Dixon-Coles e o modelo de Poisson independente foi baseada nas perdas calculadas nas mesmas 76 partidas. A Equação (12) define, para cada métrica q , a diferença pareada por partida e sua média amostral.

$$d_m^{(q)} = L_{DC,m}^{(q)} - L_{PI,m}^{(q)}, \quad \bar{d}^{(q)} = \frac{1}{N} \sum_{m=1}^N d_m^{(q)}. \quad (12)$$

Na Equação (12), as duas perdas são calculadas para a mesma partida e para a mesma métrica. Como log-loss, Brier e RPS são minimizados, diferenças negativas favorecem o Dixon-Coles. A hipótese de diferença nula foi examinada por intervalo t de 95%, teste t pareado, teste de Wilcoxon, teste de Diebold-Mariano com correção para pequena amostra e inversão aleatória de sinais com 99.999 réplicas. Os testes bilaterais constituíram a análise principal, e os p -valores foram ajustados pelo método de Holm em cada família de comparações (Diebold e Mariano, 1995; Harvey et al., 1997).

Os testes de permutação avaliaram cada modelo contra uma hipótese nula sem informação específica

das identidades das equipes. Em cada réplica, os pares de gols do treinamento foram redistribuídos em bloco entre as partidas, mantendo fixas as equipes e as posições de mandante e visitante. O modelo foi reajustado e avaliado no conjunto de teste original. Para as métricas de perda, o p-valor de cauda inferior foi calculado pela Equação (13).

$$\hat{p} = \frac{I + \sum_{b=1}^B 1(M_b \leq M_{obs})}{B+1}. \quad (13)$$

Na Equação (13), o numerador conta as réplicas com desempenho pelo menos tão favorável quanto o observado, $B = 999$ e $1(\cdot)$ é a função indicadora. O acréscimo de uma unidade ao numerador e ao denominador incorpora a configuração observada e evita p-valores iguais a zero. Para a acurácia, cuja melhoria corresponde a valores maiores, empregou-se a cauda superior (Hope, 1968; Ernst, 2004; Phipson e Smyth, 2010).

A calibração foi examinada separadamente para vitória do mandante (H), empate (D) e vitória do visitante (A). As probabilidades previstas de cada classe foram divididas em cinco faixas com números semelhantes de observações, e as frequências observadas receberam intervalos de Wilson de 95%. Como síntese descritiva, calculou-se o Expected Calibration Error pela Equação (14).

$$ECE = \sum_{c,b} \frac{n_{cb}}{3N} |f_{cb} - \bar{p}_{cb}|. \quad (14)$$

Na Equação (14), cada termo combina o número de observações da classe e da faixa, a frequência observada e a probabilidade média prevista. O denominador $3N$ corresponde às três probabilidades avaliadas em cada partida. Valores menores indicam menor discrepância média entre probabilidades e frequências observadas. Como o resultado depende da quantidade e da construção das faixas, o ECE foi interpretado como diagnóstico descritivo, e não como teste formal de calibração (Wilson, 1927).

O processamento foi realizado em linguagem R (R Core Team, 2026). A base em formato XLSX foi importada com o pacote readxl (Wickham; Bryan, 2026), enquanto a preparação, transformação e estruturação dos dados empregaram dplyr (Wickham et al., 2026), tidyr (Wickham; Vaughan; Girlich, 2026), tibble (Müller; Wickham, 2026) e purrr (Wickham; Henry, 2026). O tratamento das datas foi efetuado com lubridate (Grolemund; Wickham, 2011). As etapas iterativas e os testes de permutação foram executados em paralelo por meio dos pacotes parallel, foreach, doParallel e doSNOW, mantendo-se dois núcleos lógicos livres para as demais operações do sistema (Microsoft; Weston, 2022a; Microsoft Corporation; Weston, 2022b, 2022c; R Core Team, 2026). Os gráficos foram elaborados com ggplot2 (Wickham, 2016), e as planilhas contendo resultados, diagnósticos e métricas foram produzidas com openxlsx (Schauberger; Walker, 2025).

RESULTADOS E DISCUSSÃO

A temporada apresentou média de 1,513 gol do mandante e 1,421 do visitante. Vitórias do mandante corresponderam a 40,79% das partidas, empates a 24,47% e vitórias do visitante a 34,74%. No teste ocorreram 32, 18 e 26 resultados dessas classes, respectivamente; a regra que sempre seleccionasse a classe majoritária obteria acurácia de 42,11%.

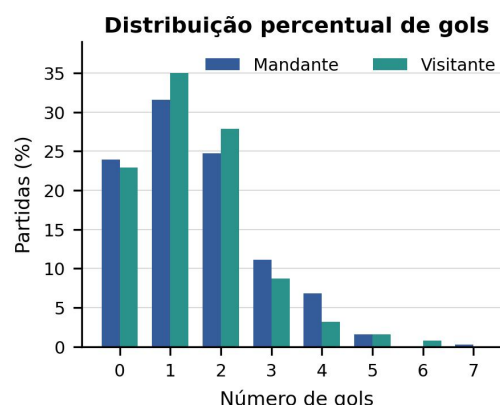


Figura 1. Distribuição percentual das contagens de gols por condição de mando.

As contagens concentraram-se entre zero e dois gols. Um gol foi a categoria modal para mandantes (31,6%) e visitantes (35,0%), enquanto ocorrências acima de quatro gols foram raras. O padrão é compatível com o uso de modelos de contagem, sem constituir, isoladamente, prova de adequação da distribuição de Poisson.

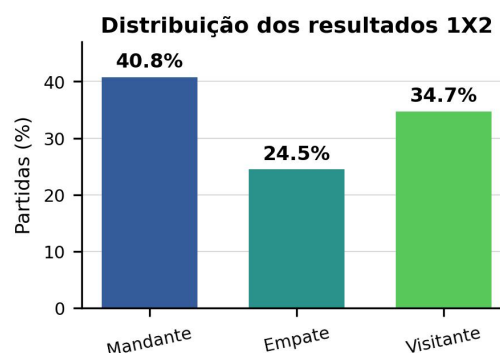


Figura 2. Distribuição dos resultados 1X2 na temporada 2024/2025.

O desequilíbrio entre as três classes foi moderado. A vitória do mandante foi a categoria mais frequente, mas representou menos da metade das partidas; os empates formaram a classe minoritária. Essa composição reforça a necessidade de avaliar o vetor probabilístico, e não apenas a classe selecionada pelo máximo.

Em relação às perdas probabilísticas no conjunto de teste ($N = 76$), o modelo Poisson independente

apresentou Log-loss = 1,001564, Brier = 0,606144 e RPS = 0,211349. Já o modelo Poisson bivariado, apresentou Log-loss = 1,001564, Brier = 0,606143 e RPS = 0,211349. Por fim, o modelo Dixon-Coles retornou Log-loss = 1,000890, Brier = 0,05823 e RPS = 0,211200. A Tabela 1 mostra o desempenho categórico no conjunto de teste, para cada um dos modelos avaliados.

Tabela 1. Perdas probabilísticas no conjunto de teste (N = 76).

Modelo	Acertos	Exatidão
Poisson independente	36	0,473684
Poisson bivariado	36	0,473684
Dixon-Coles	36	0,473684

Os três modelos acertaram 36 partidas e produziram a mesma acurácia de 47,37%. O Dixon-Coles apresentou as menores perdas, mas as reduções absolutas frente à Poisson independente foram 0,000674 na log-loss, 0,000321 no Brier e 0,000149 no RPS. O modelo de Poisson bivariado diferiu da independente apenas em casas decimais posteriores.

Dixon-Coles versus Poisson independente

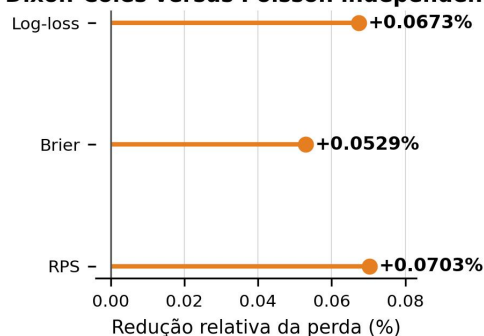


Figura 3. Melhoria relativa do Dixon-Coles em relação à Poisson independente.

Os ganhos relativos foram 0,0673% na log-loss, 0,0529% no Brier e 0,0703% no RPS. A igualdade da acurácia indica que a redistribuição de probabilidade não alterou a classe mais provável em nenhuma partida.

A Tabela 2 apresenta as diferenças pareadas entre as perdas do modelo Dixon-Coles e do modelo de Poisson independente nas mesmas 76 partidas de teste. Como a diferença foi calculada pela subtração Dixon-Coles menos Poisson independente, valores negativos indicam vantagem do primeiro modelo; também são apresentados os intervalos de confiança de 95% e os ganhos relativos correspondentes.

Tabela 2. Perdas probabilísticas no conjunto de teste (N = 76).

Métrica	d médio	IC 95%	Ganho %
Log-loss	-0,000674	[-0,004076; 0,002728]	0,0673
Brier	-0,000321	[-0,002078; 0,001437]	0,0529
RPS	-0,000149	[-0,000482; 0,000185]	0,0703

A Tabela 3 mostra a significância dos testes empregados (teste t, teste de Wilcoxon, teste de Diebold-Mariano e teste de sinais).

Tabela 2. Valores-p bilaterais da comparação pareada.

Métrica	t	Wilcoxon	DM	Sinais
Log-loss	0,6940	0,1964	0,6960	0,6952
Brier	0,7171	0,2307	0,7189	0,7173
RPS	0,3779	0,9897	0,3811	0,3776

O ajuste de Holm manteve todos os resultados acima de 0,05. Todos os intervalos incluíram zero. O menor p-valor bilateral foi 0,1964, no Wilcoxon da log-loss; após ajuste de Holm, nenhuma comparação apresentou evidência de diferença a 5%. Portanto, a vantagem numérica do Dixon-Coles não foi acompanhada por evidência de superioridade preditiva direta. Menor perda observada não equivale, por si só, a diferença estatisticamente detectável.

A Figura 4 apresenta os efeitos de ataque estimados no conjunto de treinamento pelo modelo de Poisson

independente.

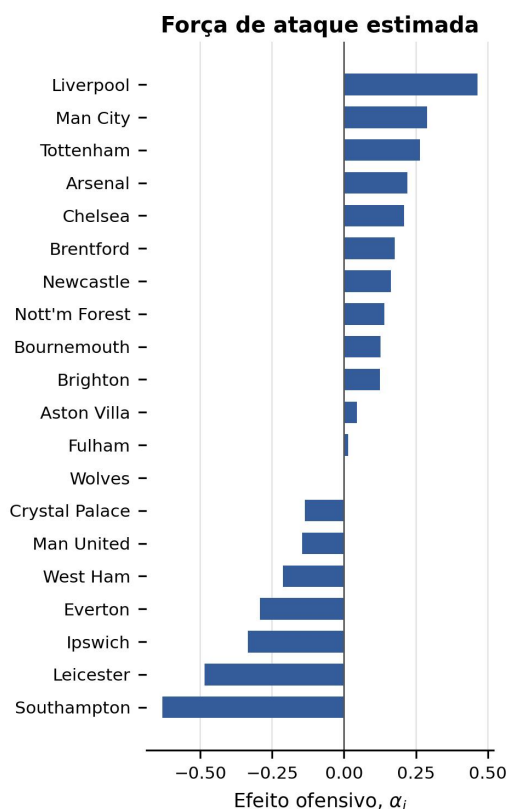


Figura 4. Efeitos de ataque estimados no treinamento pela Poisson independente.

Como os coeficientes foram recentrados em torno de zero, valores positivos indicam equipes com força ofensiva acima da média da competição, enquanto valores negativos representam desempenho ofensivo relativo inferior. Esses parâmetros sintetizam a contribuição de cada equipe para a intensidade esperada de gols, controladas a vulnerabilidade defensiva do adversário e a condição de mando. Liverpool apresentou o maior efeito ofensivo ($\alpha = 0,4647$; $\exp(\alpha) = 1,592$), seguido por Manchester City, Tottenham, Arsenal e Chelsea. Southampton, Leicester e Ipswich ocuparam as últimas posições. Os coeficientes descrevem a janela de treinamento e não constituem medidas causais nem rankings independentes do calendário.

Vulnerabilidade defensiva estimada

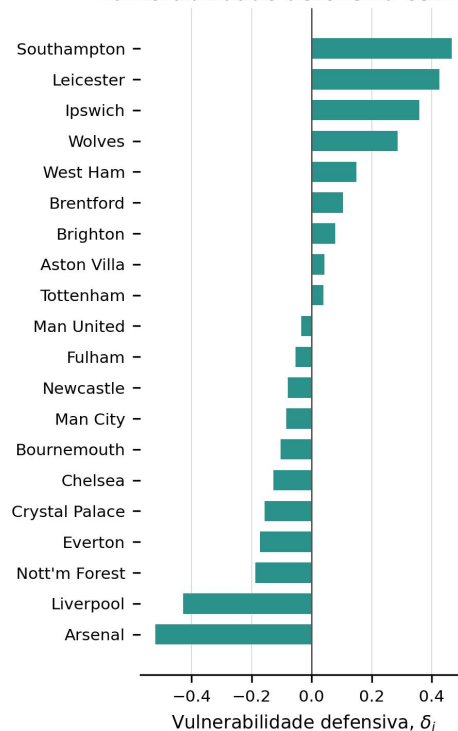


Figura 5. Vulnerabilidades defensivas estimadas no treinamento pela Poisson independente.

Valores maiores de δ indicam aumento da intensidade de gols do adversário. Arsenal apresentou a menor vulnerabilidade defensiva ($\delta = -0,5213$; $\exp(\delta) = 0,594$), seguido por Liverpool e Nottingham Forest. Southampton, Leicester e Ipswich apresentaram os maiores valores. A leitura conjunta de α e δ separa contribuição ofensiva e exposição defensiva no período de estimação.

No modelo de Poisson bivariado, λ_3 foi estimado em $1,84 \times 10^{-6}$, equivalente a $1,24 \times 10^{-6}$ da média de gols do treinamento. A covariância bruta entre as contagens foi $-0,2573$ e a covariância residual após o ajuste ataque-defesa-mando foi $-0,0460$. Como a formulação clássica impõe $\lambda_3 \geq 0$, a verossimilhança conduziu o componente compartilhado ao limite inferior. O resultado explica a equivalência prática com a Poisson independente e evidencia a limitação para dependência residual negativa (Karlis e Ntzoufras, 2003; McHale e Scarf, 2007).

No Dixon-Coles, estimou-se $\rho = -0,02858$. Mantidas as intensidades, o sinal negativo transfere massa de 0-1 e 1-0 para 0-0 e 1-1. A diferença média foi de 0,2919 ponto percentual em módulo para cada um desses quatro placares.

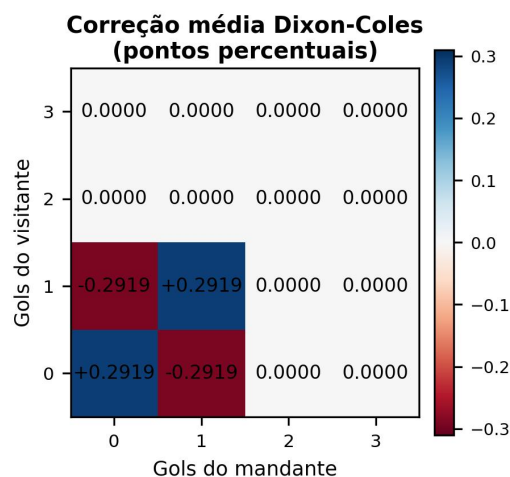


Figura 6. Redistribuição média de probabilidade produzida pelo Dixon-Coles.

A correção permaneceu restrita aos placares previstos pela formulação original, o que explica o pequeno ganho nas regras probabilísticas e a ausência de mudança na decisão 1X2.

A calibração foi sintetizada pelo *Expected Calibration Error* (ECE), calculado globalmente e separadamente para vitórias do mandante, empates e vitórias do visitante. O indicador expressa a discrepância média entre as probabilidades previstas e as frequências observadas nos intervalos de calibração, de modo que valores menores representam maior concordância. A Tabela 4 apresenta esses resultados para os três modelos, devendo-se considerar que o ECE constitui medida descritiva sensível ao número de faixas adotado e ao tamanho do conjunto de teste

Tabela 4. Expected Calibration Error global e por classe.

Métrica	Global	A	D	H
Log-loss	0,08865	0,11155	0,08706	0,06735
Brier	0,08865	0,11155	0,08706	0,06735
RPS	0,09349	0,11131	0,09984	0,06932

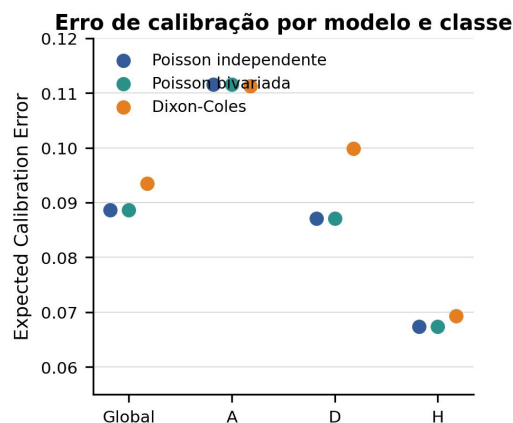


Figura 7. Redistribuição média de probabilidade produzida pelo Dixon-Coles.

Os modelos independente e bivariado apresentaram ECE global de 0,08865, e o Dixon-Coles, 0,09349. As diferenças foram pequenas e não constituem teste formal de calibração. A classe visitante concentrou o maior desvio nos modelos de referência; no Dixon-Coles, o ECE de empates aumentou para 0,09984. Os intervalos de Wilson obtidos nas cinco faixas foram amplos em várias regiões, compatíveis com o número reduzido de partidas de teste.

Nenhum modelo selecionou empate como classe de maior probabilidade, embora essa classe recebesse probabilidades positivas. A sensibilidade foi 53,85% para vitórias do visitante, 0% para empates e 68,75% para vitórias do mandante. A ausência de previsões categóricas de empate não implica probabilidade nula, mas evidencia o efeito da regra de decisão pelo máximo.

As Figuras 8 e 9 mostram os resultados das calibrações dos modelos de Dixon-Coles e de Poisson independente, com intervalos de Wilson.

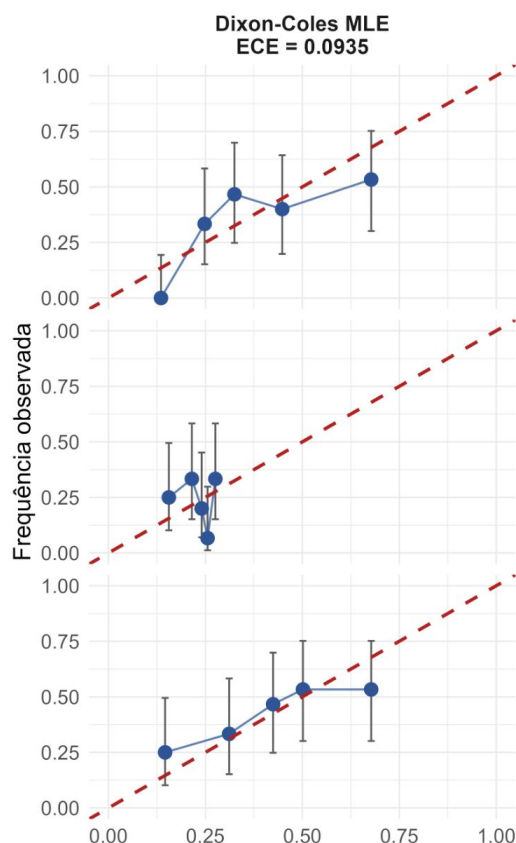


Figura 8. Calibração por classe do Dixon-Coles, com intervalos de Wilson de 95%.

No Dixon-Coles, as frequências observadas acompanharam parcialmente a diagonal de calibração, mas os intervalos foram largos, sobretudo para empates e para faixas de maior probabilidade. A incerteza impede interpretar desvios locais como evidência conclusiva de má calibração.

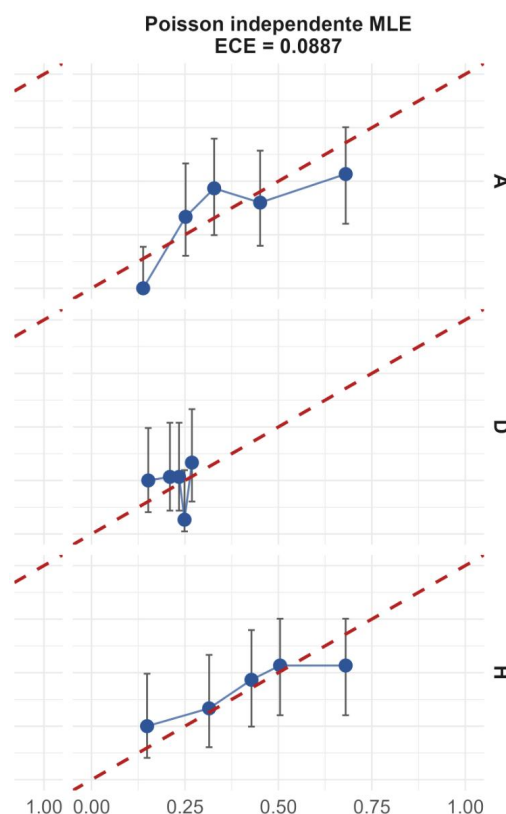


Figura 9. Calibração por classe da Poisson independente, com intervalos de Wilson de 95%..

A Poisson independente apresentou padrão visual próximo ao do modelo de Poisson bivariado, coerente com a estimativa de λ_3 próxima de zero. As diferenças em relação ao Dixon-Coles foram pequenas e concentradas nas probabilidades de baixa contagem, sem melhoria global do ECE.

Tabela 5. Valores-p dos testes de permutação contra a hipótese nula.

Modelo	Log-loss	Brier	RPS	Exatidão
Poisson independente	0,002	0,003	0,002	0,074
Poisson bivariado	0,004	0,005	0,004	0,091
Dixon-Coles	0,001	0,002	0,002	0,098

As três regras probabilísticas apresentaram valores-p inferiores a 0,01 nos três modelos, enquanto a acurácia permaneceu acima de 0,05. Assim, as probabilidades exploraram informação específica das equipes ausente sob a permutação, mas a decisão categórica não se distinguiu da referência nula ao nível de 5%. Esse resultado não contradiz a comparação pareada: superar uma distribuição nula não implica superar outro modelo concorrente.

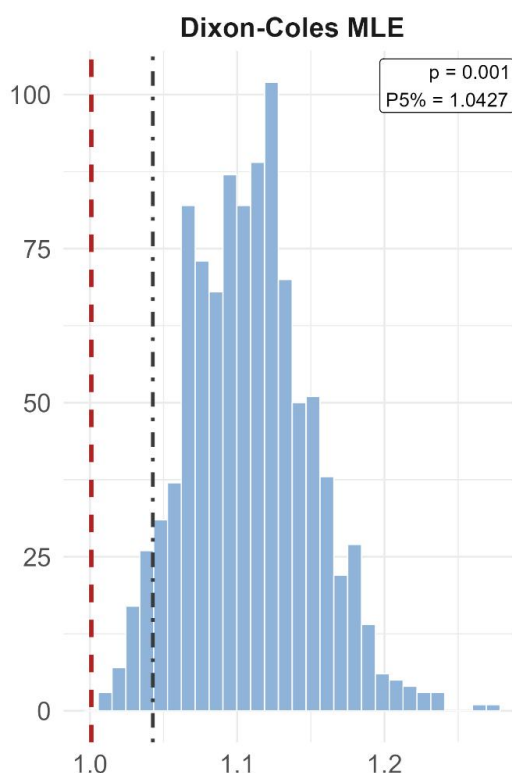


Figura 10. Distribuição nula da log-loss para o Dixon-Coles; linhas: valor observado e percentil de 5%.

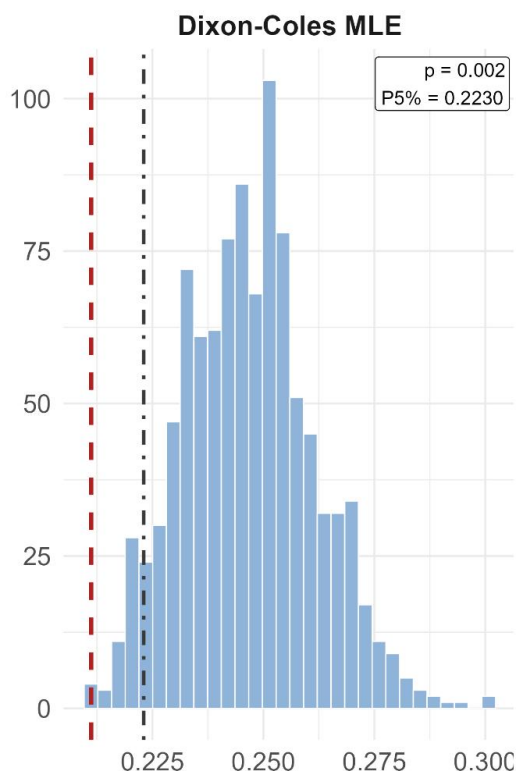


Figura 11. Distribuição nula do RPS para o Dixon-Coles; linhas: valor observado e percentil de 5%.

No Dixon-Coles, os valores observados de log-loss e RPS permaneceram à esquerda do percentil de 5%

das respectivas distribuições nulas, em concordância com $p = 0,001$ e $p = 0,002$. O mesmo padrão foi registrado para os outros modelos, conforme mostrado na Tabela 5.

CONCLUSÃO

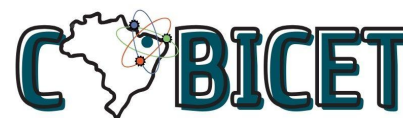
Os três modelos produziram a mesma acurácia e probabilidades muito semelhantes no teste da Premier League 2024/2025. O Dixon-Coles obteve as menores perdas, porém os ganhos foram inferiores a 0,1% e não foram sustentados pelos testes pareados. O modelo de Poisson bivariado convergiu ao limite $\lambda_3 \approx 0$, compatível com covariância residual não positiva e com a restrição $\lambda_3 \geq 0$. Os testes de permutação indicaram informação preditiva nas probabilidades, mas não na acurácia. Diante da equivalência empírica, a Poisson independente constituiu a alternativa mais parcimoniosa. A generalização permanece limitada a uma liga, uma temporada e uma única origem temporal. Trabalhos futuros devem investigar a incorporação de múltiplas temporadas e validação por origem móvel.

AGRADECIMENTOS

Escreva os agradecimentos aqui.

REFERÊNCIAS

- BRIER, G. W. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, v. 78, n. 1, p. 1–3, 1950. DOI: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.
- CONSTANTINOU, A. C.; FENTON, N. E. Solving the problem of inadequate scoring rules for assessing probabilistic football forecast models. *Journal of Quantitative Analysis in Sports*, v. 8, n. 1, p. 1–14, 2012. DOI: 10.1515/1559-0410.1418.
- DIEBOLD, F. X.; MARIANO, R. S. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, v. 13, n. 3, p. 253–263, 1995. DOI: 10.1080/07350015.1995.10524599.
- DIXON, M. J.; COLES, S. G. Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, v. 46, n. 2, p. 265–280, 1997. DOI: 10.1111/1467-9876.00065.
- EPSTEIN, E. S. A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, v. 8, n. 6, p. 985–987, 1969. DOI: 10.1175/1520-0450(1969)008<0985:ASSFPF>2.0.CO;2.
- ERNST, M. D. Permutation methods: a basis for exact inference. *Statistical Science*, v. 19, n. 4, p. 676–685, 2004. DOI: 10.1214/088342304000000396.



- FOOTBALL-DATA.CO.UK. European football results and betting odds. Base de resultados e probabilidades de partidas de futebol. Acesso em: 17 jul. 2026.
- GNEITING, T.; RAFTERY, A. E. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, v. 102, n. 477, p. 359–378, 2007. DOI: 10.1198/016214506000001437.
- GODDARD, J. Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, v. 21, n. 2, p. 331–340, 2005. DOI: 10.1016/j.ijforecast.2004.08.002.
- GROLEMUND, G.; WICKHAM, H. Dates and times made easy with lubridate. *Journal of Statistical Software*, v. 40, n. 3, p. 1–25, 2011.
- HARVEY, D.; LEYBOURNE, S.; NEWBOLD, P. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, v. 13, n. 2, p. 281–291, 1997. DOI: 10.1016/S0169-2070(96)00719-4.
- HOPE, A. C. A. A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 30, n. 3, p. 582–598, 1968. DOI: 10.1111/j.2517-6161.1968.tb00759.x.
- KARLIS, D.; NTZOUFRAS, I. Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, v. 52, n. 3, p. 381–393, 2003. DOI: 10.1111/1467-9884.00366.
- LEE, A. J. Modeling scores in the Premier League: is Manchester United really the best? *Chance*, v. 10, n. 1, p. 15–19, 1997. DOI: 10.1080/09332480.1997.10554791.
- MAHER, M. J. Modelling association football scores. *Statistica Neerlandica*, v. 36, n. 3, p. 109–118, 1982. DOI: 10.1111/j.1467-9574.1982.tb00782.x.
- MCHALE, I.; SCARF, P. Modelling soccer matches using bivariate discrete distributions with general dependence structure. *Statistica Neerlandica*, v. 61, n. 4, p. 432–445, 2007. DOI: 10.1111/j.1467-9574.2007.00368.x.
- MICROSOFT; WESTON, S. doParallel: foreach parallel adaptor for the parallel package. R package, 2022a. DOI: 10.32614/CRAN.package.doParallel.
- MICROSOFT; WESTON, S. doSNOW: foreach parallel adaptor for the snow package. R package, 2022b. DOI: 10.32614/CRAN.package.doSNOW.
- MICROSOFT; WESTON, S. foreach: provides foreach looping construct. R package, 2022c. DOI: 10.32614/CRAN.package.foreach.
- MÜLLER, K.; WICKHAM, H. tibble: simple data frames. R package, 2026. DOI: 10.32614/CRAN.package.tibble.
- PHIPSON, B.; SMYTH, G. K. Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. *Statistical Applications in Genetics and Molecular Biology*, v. 9, n. 1, art. 39, 2010. DOI: 10.2202/1544-6115.1585.
- R CORE TEAM. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2026.
- SCHAUBERGER, P.; WALKER, A. openxlsx: read, write and edit xlsx files. R package, 2025. DOI: 10.32614/CRAN.package.openxlsx.
- TSOKOS, A. et al. Modeling outcomes of soccer matches. *Machine Learning*, v. 108, p. 77–95, 2019. DOI: 10.1007/s10994-018-5741-1.
- WICKHAM, H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag, 2016.
- WICKHAM, H.; BRYAN, J. readxl: read Excel files. R package, 2026. DOI: 10.32614/CRAN.package.readxl.
- WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; MÜLLER, K.; VAUGHAN, D. dplyr: a grammar of data manipulation. R package, 2026. DOI: 10.32614/CRAN.package.dplyr.
- WICKHAM, H.; HENRY, L. purrr: functional programming tools. R package, 2026. DOI: 10.32614/CRAN.package.purrr.
- WICKHAM, H.; VAUGHAN, D.; GIRLICH, M. tidyr: tidy messy data. R package, 2025. DOI: 10.32614/CRAN.package.tidyr.
- WILSON, E. B. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, v. 22, n. 158, p. 209–212, 1927. DOI: 10.1080/01621459.1927.10502953.