Introdução ao cálculo de probabilidades no futebol

Daniel Takata Gomes

7a. Semana de Iniciação Científica - ENCE/IBGE

7 de novembro de 2023

Sumário

Sumário

Sumário

Sumário

Sumário

Sumário

Arquivos em:

https://github.com/daniel-takata/minicurso-semic

Atenção: os dados estão atualizados até a rodada 31 do Campeonato Brasileiro 2023 (não constam os resultados da rodada do último fim de semana).

TNT SPORTS NOTÍCIAS BLOGS VÍDEOS TABELAS AO VIVO

FUTEBOL BRASILEIRO

Mesmo com derrota, Botafogo mantém alta probabilidade de ser campeão; veja números

Glorioso perdeu de virada para o Palmeiras, que aumentou suas chances de levantar o troféu do Brasileirão

Parece óbvio que tais probabilidades são calculadas através de simulações, baseadas nas probabilidades dos resultados dos jogos.

Parece óbvio que tais probabilidades são calculadas através de simulações, baseadas nas probabilidades dos resultados dos jogos.

O que não se comenta muito é: como as probabilidades dos resultados dos jogos são calculadas?

Neste minicurso, será feita uma introdução aos modelos probabilísticos para resultados de jogos de futebol, tendo como base alguns dos trabalhos mais citados que envolvem o tema, tais como:

Neste minicurso, será feita uma introdução aos modelos probabilísticos para resultados de jogos de futebol, tendo como base alguns dos trabalhos mais citados que envolvem o tema, tais como:

- Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica* 36, n. 3, pp. 109-118.

Neste minicurso, será feita uma introdução aos modelos probabilísticos para resultados de jogos de futebol, tendo como base alguns dos trabalhos mais citados que envolvem o tema, tais como:

- Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica* 36, n. 3, pp. 109-118.
- Dixon, M. J., Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics* 46, n. 2, pp. 265-280.

Neste minicurso, será feita uma introdução aos modelos probabilísticos para resultados de jogos de futebol, tendo como base alguns dos trabalhos mais citados que envolvem o tema, tais como:

- Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica* 36, n. 3, pp. 109-118.
- Dixon, M. J., Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics* 46, n. 2, pp. 265-280.
- Baker, R. D., McHale, I. G. (2015). Time varying ratings in association football: the all-time greatest team is... *Journal of the Royal Statistical Society Series A* 178, n. 2, pp. 481-492.

Pensamento inicial: resultados de jogos de futebol são realizações de eventos aleatórios.

Pensamento inicial: resultados de jogos de futebol são realizações de eventos aleatórios.

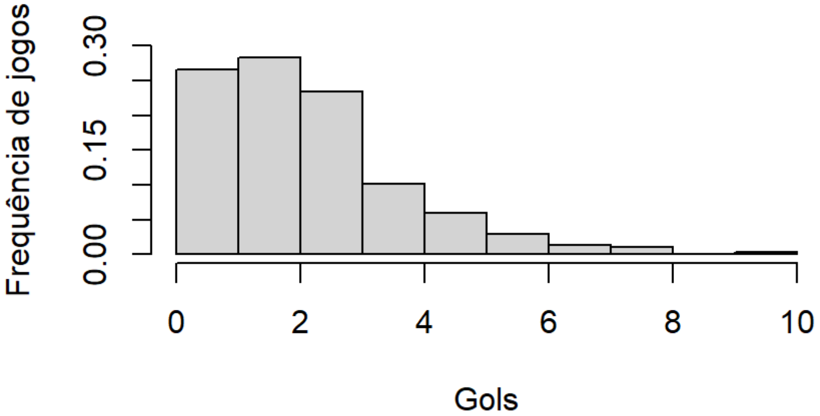
Como proceder com a modelagem de tais eventos?

Pensamento inicial: resultados de jogos de futebol são realizações de eventos aleatórios.

Como proceder com a modelagem de tais eventos?

É possível construir um modelo inicial utilizando conceitos básicos de probabilidade.

Histograma do número de gols por partida - Campeonato Brasileiro 2023



A distribuição do número de gols é visivelmente assimétrica, o que não é surpresa.

A distribuição do número de gols é visivelmente assimétrica, o que não é surpresa.

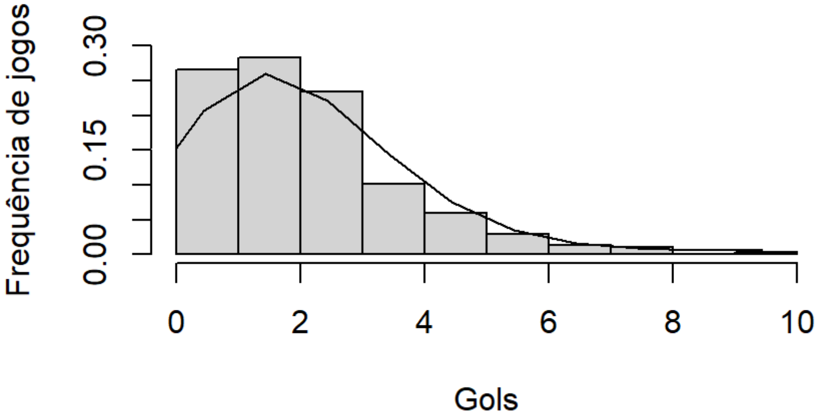
Um gol em uma partida de futebol é considerado um evento raro. Por isso, há muitos jogos com poucos gols e poucos jogos com muitos gols.

A distribuição do número de gols é visivelmente assimétrica, o que não é surpresa.

Um gol em uma partida de futebol é considerado um evento raro. Por isso, há muitos jogos com poucos gols e poucos jogos com muitos gols.

Média de gols por jogo: 2,54.

Histograma do número de gols por partida - Campeonato Brasileiro 2023



Se a probabilidade de um ataque de um time resultar em um gol é p (pequena), e o número de ataques (por suposição independentes) em uma partida é razoavelmente grande, o número de gols terá distribuição Binomial.

Se a probabilidade de um ataque de um time resultar em um gol é p (pequena), e o número de ataques (por suposição independentes) em uma partida é razoavelmente grande, o número de gols terá distribuição Binomial.

Nessas circunstâncias, pode-se pensar que a distribuição de Poisson é uma boa aproximação para o número de gols marcados por um time em uma partida.

Se a probabilidade de um ataque de um time resultar em um gol é p (pequena), e o número de ataques (por suposição independentes) em uma partida é razoavelmente grande, o número de gols terá distribuição Binomial.

Nessas circunstâncias, pode-se pensar que a distribuição de Poisson é uma boa aproximação para o número de gols marcados por um time em uma partida.

Esse é o ponto de partida da grande maioria dos modelos utilizados para cálculo de probabilidades de resultados de jogos de futebol.

Se X é uma variável aleatória com distribuição de Poisson, sua função de probabilidade é dada por:

Se X é uma variável aleatória com distribuição de Poisson, sua função de probabilidade é dada por:

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

Se X é uma variável aleatória com distribuição de Poisson, sua função de probabilidade é dada por:

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

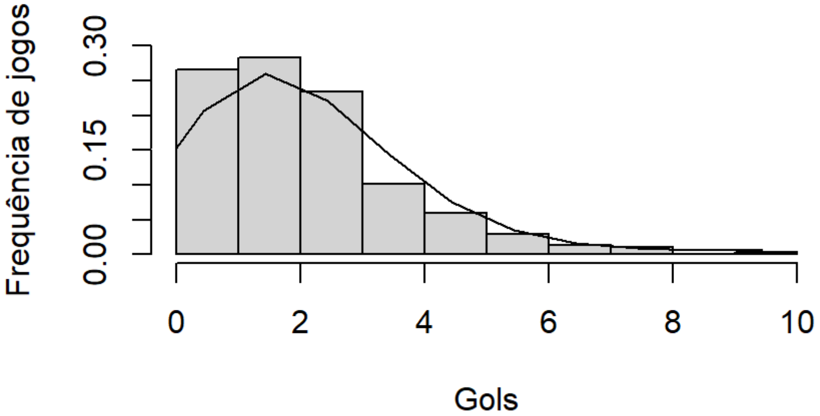
em que $X = 0, 1, 2, 3, \ldots$ e $\lambda > 0$. Em particular, $E(X) = \lambda$.

Na prática, a distribuição de Poisson tem mostrado um ótimo modelo teórico para número de gols em jogos de futebol.

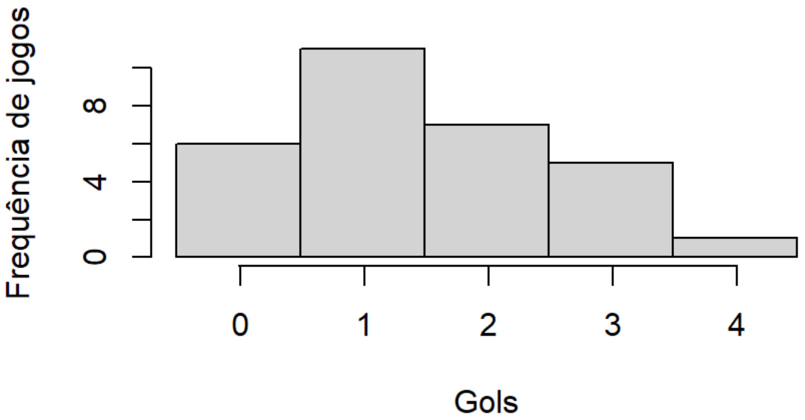
Na prática, a distribuição de Poisson tem mostrado um ótimo modelo teórico para número de gols em jogos de futebol.

De fato, a função de probabilidade da distribuição de Poisson, considerando o valor para o parâmetro λ sendo 2,54 (justamente a média de gols do campeonato), parece se ajustar bem aos dados observados.

Histograma do número de gols por partida - Campeonato Brasileiro 2023



Histograma do número de gols por partida - Flamengo - Brasileiro 2023



Média de gols por jogo = 1,47

Assumindo 1,47 como o valor mais plausível para λ de acordo com os dados (afinal, λ é a média da distribuição de Poisson), temos, utilizando a distribuição teórica:

Assumindo 1,47 como o valor mais plausível para λ de acordo com os dados (afinal, λ é a média da distribuição de Poisson), temos, utilizando a distribuição teórica:

Número de gols marcados	Número de jogos	Frequência observada	Frequência esperada pela Poisson
0	6	20.00%	23.07%
1	11	36.67%	33.84%
2	7	23.33%	24.81%
3	5	16.67%	12.13%
4	1	3.33%	4.45%

Modelo simples:

Modelo simples:

X_i: v.a. número de gols do time i em um jogo, $i = 1, 2, \ldots, 20$.

Modelo simples:

X_i: v.a. número de gols do time i em um jogo, $i = 1, 2, \ldots, 20$.

Considere o time i jogando contra o time j.

Modelo simples:

X_i: v.a. número de gols do time i em um jogo, $i = 1, 2, \ldots, 20$.

Considere o time i jogando contra o time j.

O placar final será determinado pelo número de gols do time i (X_i) e pelo número de gols do time j (X_j).

Modelo simples:

X_i: v.a. número de gols do time i em um jogo, $i = 1, 2, \ldots, 20$.

Considere o time i jogando contra o time j.

O placar final será determinado pelo número de gols do time i (X_i) e pelo número de gols do time j (X_j).

Placar observado: (x_i, x_j).

Modelo simples:

X_i: v.a. número de gols do time i em um jogo, $i = 1, 2, \ldots, 20$.

Considere o time i jogando contra o time j.

O placar final será determinado pelo número de gols do time i (X_i) e pelo número de gols do time j (X_j).

Placar observado: (x_i, x_j).

Por suposição, X_i e X_j são independentes.

Seria isso um problema?

Seria isso um problema?

"All models are wrong, but some are useful" (George Box)

$X_i \sim Poisson(\lambda_i),$

$$X_i \sim Poisson(\lambda_i),$$

em que λ_i é o número esperado (média) de gols do time i.

$X_i \sim Poisson(\lambda_i)$,

em que λ_i é o número esperado (média) de gols do time i.

O valor de λ_i mais plausível a ser considerado será a média de gols observada nos jogos realizados.

$X_i \sim Poisson(\lambda_i)$,

em que λ_i é o número esperado (média) de gols do time i.

O valor de λ_i mais plausível a ser considerado será a média de gols observada nos jogos realizados.

Vamos denotar esse valor observado por $\hat{\lambda}_i$.

Média de gols por partida de cada time no Campeonato Brasileiro 2023

Time	$\hat{\lambda}$	**Time**	$\hat{\lambda}$
AmericaMG	1.19	Fluminense	1.35
AthleticoPR	1.45	Fortaleza	1.17
AtleticoMG	1.26	Goias	1.07
Bahia	1.23	Gremio	1.68
Botafogo	1.60	Internacional	1.16
Corinthians	1.19	Palmeiras	1.68
Coritiba	1.13	RedBullBragantino	1.41
Cruzeiro	0.97	Santos	1.16
Cuiaba	1.03	SaoPaulo	1.17
Flamengo	1.47	Vasco	1.00

Para o cálculo da probabilidade de um time A vencer um time B, é preciso calcular a probabilidade de o time A marcar mais gols que o time B.

Para o cálculo da probabilidade de um time A vencer um time B, é preciso calcular a probabilidade de o time A marcar mais gols que o time B.

Para isso, utiliza-se o seguinte resultado:

Para o cálculo da probabilidade de um time A vencer um time B, é preciso calcular a probabilidade de o time A marcar mais gols que o time B.

Para isso, utiliza-se o seguinte resultado:

$$P(A > B) = \sum_{k=0}^{\infty} P(A > B | B = k) P(B = k)$$

$$= \sum_{k=0}^{\infty} P(A \geq k+1) P(B = k) = \sum_{k=0}^{\infty} \left(\sum_{l=k+1}^{\infty} \frac{\lambda_A^l e^{-\lambda_A}}{l!} \right) \frac{\lambda_B^k e^{-\lambda_B}}{k!}.$$

Por exemplo, cálculo da probabilidade de o time i vencer o time j por 1 a 0:

$$P(X_i = 1, X_j = 0) = P(X_i = 1)P(X_j = 0) = \frac{e^{-\hat{\lambda}_i}\hat{\lambda}_i^1}{1!}\frac{e^{-\hat{\lambda}_j}\hat{\lambda}_j^0}{0!}.$$

Supondo que o time i seja o Internacional e o time j seja o Fluminense, então $\hat{\lambda}_i = 1,16$ e $\hat{\lambda}_i = 1,35$.

Supondo que o time i seja o Internacional e o time j seja o Fluminense, então $\hat{\lambda}_i = 1,16$ e $\hat{\lambda}_i = 1,35$.

$$P(X_i = 1, X_j = 0) = P(X_i = 1)P(X_j = 0) = 1,16e^{-1,16}e^{-1,35} =$$

$= 0{,}0943 = 9{,}43\%.$

Supondo que o time i seja o Internacional e o time j seja o Fluminense, então $\hat{\lambda}_i = 1,16$ e $\hat{\lambda}_i = 1,35$.

$$P(X_i = 1, X_j = 0) = P(X_i = 1)P(X_j = 0) = 1,16e^{-1,16}e^{-1,35} =$$

$= 0{,}0943 = 9{,}43\%$.

Pode-se fazer o mesmo cálculo para os outros resultados possíveis.

		Fluminense							
	Gols	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	0	8.13%	10.97%	7.41%	3.33%	1.12%	0.30%	0.07%	0.01%
	1	9.43%	12.73%	8.59%	3.87%	1.30%	0.35%	0.08%	0.02%
	2	5.47%	7.38%	4.98%	2.24%	0.76%	0.20%	0.05%	0.01%
	3	2.11%	2.85%	1.93%	0.87%	0.29%	0.08%	0.02%	0.00%
Inter	**4**	0.61%	0.83%	0.56%	0.25%	0.08%	0.02%	0.01%	0.00%
	5	0.14%	0.19%	0.13%	0.06%	0.02%	0.01%	0.00%	0.00%
	6	0.03%	0.04%	0.03%	0.01%	0.00%	0.00%	0.00%	0.00%
	7	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

.

		Fluminense							
	Gols	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Inter	**0**	8.13%	10.97%	7.41%	3.33%	1.12%	0.30%	0.07%	0.01%
	1	9.43%	12.73%	8.59%	3.87%	1.30%	0.35%	0.08%	0.02%
	2	5.47%	7.38%	4.98%	2.24%	0.76%	0.20%	0.05%	0.01%
	3	2.11%	2.85%	1.93%	0.87%	0.29%	0.08%	0.02%	0.00%
	4	0.61%	0.83%	0.56%	0.25%	0.08%	0.02%	0.01%	0.00%
	5	0.14%	0.19%	0.13%	0.06%	0.02%	0.01%	0.00%	0.00%
	6	0.03%	0.04%	0.03%	0.01%	0.00%	0.00%	0.00%	0.00%
	7	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

	Fluminense								
	Gols	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Inter	**0**	8.13%	10.97%	7.41%	3.33%	1.12%	0.30%	0.07%	0.01%
	1	9.43%	12.73%	8.59%	3.87%	1.30%	0.35%	0.08%	0.02%
	2	5.47%	7.38%	4.98%	2.24%	0.76%	0.20%	0.05%	0.01%
	3	2.11%	2.85%	1.93%	0.87%	0.29%	0.08%	0.02%	0.00%
	4	0.61%	0.83%	0.56%	0.25%	0.08%	0.02%	0.01%	0.00%
	5	0.14%	0.19%	0.13%	0.06%	0.02%	0.01%	0.00%	0.00%
	6	0.03%	0.04%	0.03%	0.01%	0.00%	0.00%	0.00%	0.00%
	7	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Probabilidade de vitória do Internacional: 32,09%.

		Fluminense							
	Gols	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Inter	**0**	8.13%	10.97%	7.41%	3.33%	1.12%	0.30%	0.07%	0.01%
	1	9.43%	12.73%	8.59%	3.87%	1.30%	0.35%	0.08%	0.02%
	2	5.47%	7.38%	4.98%	2.24%	0.76%	0.20%	0.05%	0.01%
	3	2.11%	2.85%	1.93%	0.87%	0.29%	0.08%	0.02%	0.00%
	4	0.61%	0.83%	0.56%	0.25%	0.08%	0.02%	0.01%	0.00%
	5	0.14%	0.19%	0.13%	0.06%	0.02%	0.01%	0.00%	0.00%
	6	0.03%	0.04%	0.03%	0.01%	0.00%	0.00%	0.00%	0.00%
	7	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

		Fluminense							
	Gols	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Inter	**0**	8.13%	10.97%	7.41%	3.33%	1.12%	0.30%	0.07%	0.01%
	1	9.43%	12.73%	8.59%	3.87%	1.30%	0.35%	0.08%	0.02%
	2	5.47%	7.38%	4.98%	2.24%	0.76%	0.20%	0.05%	0.01%
	3	2.11%	2.85%	1.93%	0.87%	0.29%	0.08%	0.02%	0.00%
	4	0.61%	0.83%	0.56%	0.25%	0.08%	0.02%	0.01%	0.00%
	5	0.14%	0.19%	0.13%	0.06%	0.02%	0.01%	0.00%	0.00%
	6	0.03%	0.04%	0.03%	0.01%	0.00%	0.00%	0.00%	0.00%
	7	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Probabilidade de vitória do Fluminense: 41,11%.

Fluminense

Inter	Gols	0	1	2	3	4	5	6	7
	0	8.13%	10.97%	7.41%	3.33%	1.12%	0.30%	0.07%	0.01%
	1	9.43%	12.73%	8.59%	3.87%	1.30%	0.35%	0.08%	0.02%
	2	5.47%	7.38%	4.98%	2.24%	0.76%	0.20%	0.05%	0.01%
	3	2.11%	2.85%	1.93%	0.87%	0.29%	0.08%	0.02%	0.00%
	4	0.61%	0.83%	0.56%	0.25%	0.08%	0.02%	0.01%	0.00%
	5	0.14%	0.19%	0.13%	0.06%	0.02%	0.01%	0.00%	0.00%
	6	0.03%	0.04%	0.03%	0.01%	0.00%	0.00%	0.00%	0.00%
	7	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

		Fluminense							
	Gols	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Inter	**0**	8.13%	10.97%	7.41%	3.33%	1.12%	0.30%	0.07%	0.01%
	1	9.43%	12.73%	8.59%	3.87%	1.30%	0.35%	0.08%	0.02%
	2	5.47%	7.38%	4.98%	2.24%	0.76%	0.20%	0.05%	0.01%
	3	2.11%	2.85%	1.93%	0.87%	0.29%	0.08%	0.02%	0.00%
	4	0.61%	0.83%	0.56%	0.25%	0.08%	0.02%	0.01%	0.00%
	5	0.14%	0.19%	0.13%	0.06%	0.02%	0.01%	0.00%	0.00%
	6	0.03%	0.04%	0.03%	0.01%	0.00%	0.00%	0.00%	0.00%
	7	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Probabilidade de empate: 26,79%.

Uma desvantagem dessa abordagem é que o modelo probabilístico não leva em conta qual é o time mandante e qual é o time visitante.

Uma desvantagem dessa abordagem é que o modelo probabilístico não leva em conta qual é o time mandante e qual é o time visitante.

Para qualquer time, jogar em seu estádio pode representar uma vantagem.

Uma desvantagem dessa abordagem é que o modelo probabilístico não leva em conta qual é o time mandante e qual é o time visitante.

Para qualquer time, jogar em seu estádio pode representar uma vantagem.

Neste campeonato, quase todos os times têm médias de gols maiores quando são mandantes.

Uma desvantagem dessa abordagem é que o modelo probabilístico não leva em conta qual é o time mandante e qual é o time visitante.

Para qualquer time, jogar em seu estádio pode representar uma vantagem.

Neste campeonato, quase todos os times têm médias de gols maiores quando são mandantes.

Tal padrão observa-se frequentemente em campeonatos de futebol.

Time	Média de gols como...		Time	Média de gols como...	
	Mandante	Visitante		Mandante	Visitante
AmericaMG	1.13	1.25	**Fluminense**	1.75	0.93
AthleticoPR	1.81	1.07	**Fortaleza**	1.64	0.73
AtleticoMG	1.31	1.20	**Goias**	1.20	0.93
Bahia	1.60	0.88	**Gremio**	2.00	1.38
Botafogo	2.00	1.14	**Internacional**	1.63	0.67
Corinthians	1.44	0.93	**Palmeiras**	1.73	1.63
Coritiba	1.00	1.25	**RedBullBragantino**	1.53	1.29
Cruzeiro	0.60	1.27	**Santos**	1.67	0.69
Cuiaba	1.13	0.87	**SaoPaulo**	1.69	0.60
Flamengo	1.36	1.56	**Vasco**	1.13	0.94

Por isso, pode-se fazer uma modificação no modelo probabilístico para contemplar o fato de se jogar como mandante.

Por isso, pode-se fazer uma modificação no modelo probabilístico para contemplar o fato de se jogar como mandante.

Basicamente, a mudança é que, agora, cada time passará a ter dois parâmetros λ: um quando jogar como mandante e outro quando jogar como visitante.

Modelo:

Modelo:

X_i: v.a. número de gols do time i quando o time é mandante.

 Modelo considerando mando de campo

Modelo:

X_i: v.a. número de gols do time i quando o time é mandante.

Y_j: v.a. número de gols do time j quando o time é visitante.

Modelo:

X_i: v.a. número de gols do time i quando o time é mandante.

Y_j: v.a. número de gols do time j quando o time é visitante.

Placar observado de um jogo do time i, mandante, contra o time j, visitante: (x_i, y_j), $i, j = 1, 2, \ldots, 20, i \neq j$.

Modelo:

X_i: v.a. número de gols do time i quando o time é mandante.

Y_j: v.a. número de gols do time j quando o time é visitante.

Placar observado de um jogo do time i, mandante, contra o time j, visitante: (x_i, y_j), $i, j = 1, 2, \ldots, 20, i \neq j$.

Por suposição, X_i e Y_j são independentes.

 Modelo considerando mando de campo

$X_i \sim Poisson(\lambda_i^M),$

 Modelo considerando mando de campo

$X_i \sim Poisson(\lambda_i^M)$,

em que λ_i^M é o número esperado (média) de gols do time i quando joga como mandante.

 Modelo considerando mando de campo

$X_i \sim Poisson(\lambda_i^M)$,

em que λ_i^M é o número esperado (média) de gols do time i quando joga como mandante.

$Y_j \sim Poisson(\lambda_j^V)$,

$X_i \sim Poisson(\lambda_i^M)$,

em que λ_i^M é o número esperado (média) de gols do time i quando joga como mandante.

$Y_j \sim Poisson(\lambda_j^V)$,

em que λ_j^V é o número esperado (média) de gols do time j quando joga como mandante.

 Modelo considerando mando de campo

$X_i \sim Poisson(\lambda_i^M)$,

em que λ_i^M é o número esperado (média) de gols do time i quando joga como mandante.

$Y_j \sim Poisson(\lambda_j^V)$,

em que λ_j^V é o número esperado (média) de gols do time j quando joga como mandante.

Os valores de λ_i^M e λ_j^V mais plausíveis serão as médias de gols observadas nos jogos realizados em que os times i e j atuem como mandante e visitante, respectivamente.

Supondo que o time i, mandante, seja o Internacional e o time j, visitante, seja o Fluminense, então $\hat{\lambda}_i^M = 1,63$ e $\hat{\lambda}_j^V = 0,93$.

 Modelo considerando mando de campo

Supondo que o time i, mandante, seja o Internacional e o time j, visitante, seja o Fluminense, então $\hat{\lambda}_i^M = 1,63$ e $\hat{\lambda}_j^V = 0,93$.

$$P(X_i = 1, X_j = 0) = P(X_i = 1)P(X_j = 0) = 1,63e^{-1{,}63}e^{-0{,}93} =$$

$= 0{,}1206 = 12{,}06\%$.

 Modelo considerando mando de campo

Supondo que o time i, mandante, seja o Internacional e o time j, visitante, seja o Fluminense, então $\hat{\lambda}_i^M = 1,63$ e $\hat{\lambda}_j^V = 0,93$.

$$P(X_i = 1, X_j = 0) = P(X_i = 1)P(X_j = 0) = 1,63e^{-1,63}e^{-0,93} =$$

$= 0{,}1206 = 12{,}06\%$.

Pode-se fazer o mesmo cálculo para os outros resultados possíveis.

	Fluminense								
	Gols	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Inter	**0**	7.73%	7.19%	3.34%	1.04%	0.24%	0.04%	0.01%	0.00%
	1	12.60%	11.72%	5.45%	1.69%	0.39%	0.07%	0.01%	0.00%
	2	10.27%	9.55%	4.44%	1.38%	0.32%	0.06%	0.01%	0.00%
	3	5.58%	5.19%	2.41%	0.75%	0.17%	0.03%	0.01%	0.00%
	4	2.27%	2.11%	0.98%	0.30%	0.07%	0.01%	0.00%	0.00%
	5	0.74%	0.69%	0.32%	0.10%	0.02%	0.00%	0.00%	0.00%
	6	0.20%	0.19%	0.09%	0.03%	0.01%	0.00%	0.00%	0.00%
	7	0.05%	0.04%	0.02%	0.01%	0.00%	0.00%	0.00%	0.00%

		Fluminense							
	Gols	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	0	7.73%	7.19%	3.34%	1.04%	0.24%	0.04%	0.01%	0.00%
	1	12.60%	11.72%	5.45%	1.69%	0.39%	0.07%	0.01%	0.00%
	2	10.27%	9.55%	4.44%	1.38%	0.32%	0.06%	0.01%	0.00%
	3	5.58%	5.19%	2.41%	0.75%	0.17%	0.03%	0.01%	0.00%
Inter	**4**	2.27%	2.11%	0.98%	0.30%	0.07%	0.01%	0.00%	0.00%
	5	0.74%	0.69%	0.32%	0.10%	0.02%	0.00%	0.00%	0.00%
	6	0.20%	0.19%	0.09%	0.03%	0.01%	0.00%	0.00%	0.00%
	7	0.05%	0.04%	0.02%	0.01%	0.00%	0.00%	0.00%	0.00%

	Fluminense								
	Gols	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	0	7.73%	7.19%	3.34%	1.04%	0.24%	0.04%	0.01%	0.00%
	1	12.60%	11.72%	5.45%	1.69%	0.39%	0.07%	0.01%	0.00%
	2	10.27%	9.55%	4.44%	1.38%	0.32%	0.06%	0.01%	0.00%
	3	5.58%	5.19%	2.41%	0.75%	0.17%	0.03%	0.01%	0.00%
Inter	**4**	2.27%	2.11%	0.98%	0.30%	0.07%	0.01%	0.00%	0.00%
	5	0.74%	0.69%	0.32%	0.10%	0.02%	0.00%	0.00%	0.00%
	6	0.20%	0.19%	0.09%	0.03%	0.01%	0.00%	0.00%	0.00%
	7	0.05%	0.04%	0.02%	0.01%	0.00%	0.00%	0.00%	0.00%

Probabilidade de vitória do Internacional: 53,78%.

		Fluminense							
	Gols	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Inter	**0**	7.73%	7.19%	3.34%	1.04%	0.24%	0.04%	0.01%	0.00%
	1	12.60%	11.72%	5.45%	1.69%	0.39%	0.07%	0.01%	0.00%
	2	10.27%	9.55%	4.44%	1.38%	0.32%	0.06%	0.01%	0.00%
	3	5.58%	5.19%	2.41%	0.75%	0.17%	0.03%	0.01%	0.00%
	4	2.27%	2.11%	0.98%	0.30%	0.07%	0.01%	0.00%	0.00%
	5	0.74%	0.69%	0.32%	0.10%	0.02%	0.00%	0.00%	0.00%
	6	0.20%	0.19%	0.09%	0.03%	0.01%	0.00%	0.00%	0.00%
	7	0.05%	0.04%	0.02%	0.01%	0.00%	0.00%	0.00%	0.00%

		Fluminense							
	Gols	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Inter	**0**	7.73%	7.19%	3.34%	1.04%	0.24%	0.04%	0.01%	0.00%
	1	12.60%	11.72%	5.45%	1.69%	0.39%	0.07%	0.01%	0.00%
	2	10.27%	9.55%	4.44%	1.38%	0.32%	0.06%	0.01%	0.00%
	3	5.58%	5.19%	2.41%	0.75%	0.17%	0.03%	0.01%	0.00%
	4	2.27%	2.11%	0.98%	0.30%	0.07%	0.01%	0.00%	0.00%
	5	0.74%	0.69%	0.32%	0.10%	0.02%	0.00%	0.00%	0.00%
	6	0.20%	0.19%	0.09%	0.03%	0.01%	0.00%	0.00%	0.00%
	7	0.05%	0.04%	0.02%	0.01%	0.00%	0.00%	0.00%	0.00%

Probabilidade de vitória do Fluminense: 24,71%.

		Fluminense							
	Gols	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Inter	**0**	7.73%	7.19%	3.34%	1.04%	0.24%	0.04%	0.01%	0.00%
	1	12.60%	11.72%	5.45%	1.69%	0.39%	0.07%	0.01%	0.00%
	2	10.27%	9.55%	4.44%	1.38%	0.32%	0.06%	0.01%	0.00%
	3	5.58%	5.19%	2.41%	0.75%	0.17%	0.03%	0.01%	0.00%
	4	2.27%	2.11%	0.98%	0.30%	0.07%	0.01%	0.00%	0.00%
	5	0.74%	0.69%	0.32%	0.10%	0.02%	0.00%	0.00%	0.00%
	6	0.20%	0.19%	0.09%	0.03%	0.01%	0.00%	0.00%	0.00%
	7	0.05%	0.04%	0.02%	0.01%	0.00%	0.00%	0.00%	0.00%

Gols		Fluminense 0	1	2	3	4	5	6	7
Inter	0	7.73%	7.19%	3.34%	1.04%	0.24%	0.04%	0.01%	0.00%
	1	12.60%	11.72%	5.45%	1.69%	0.39%	0.07%	0.01%	0.00%
	2	10.27%	9.55%	4.44%	1.38%	0.32%	0.06%	0.01%	0.00%
	3	5.58%	5.19%	2.41%	0.75%	0.17%	0.03%	0.01%	0.00%
	4	2.27%	2.11%	0.98%	0.30%	0.07%	0.01%	0.00%	0.00%
	5	0.74%	0.69%	0.32%	0.10%	0.02%	0.00%	0.00%	0.00%
	6	0.20%	0.19%	0.09%	0.03%	0.01%	0.00%	0.00%	0.00%
	7	0.05%	0.04%	0.02%	0.01%	0.00%	0.00%	0.00%	0.00%

Probabilidade de empate: 21,47%.

Note que a probabilidade de vitória do Internacional aumentou consideravelmente, de 32,09% para 53,78%, pois agora o fator mando de campo está sendo levado em consideração.

Note que a probabilidade de vitória do Internacional aumentou consideravelmente, de 32,09% para 53,78%, pois agora o fator mando de campo está sendo levado em consideração.

Agora entraremos em um modelo com mais parâmetros, um pouco mais refinado.

O modelo proposto por Maher (1982) tem sido utilizado, com adaptações e variações, como um dos principais para o cálculo de probabilidades de resultados de jogos de futebol.

O modelo proposto por Maher (1982) tem sido utilizado, com adaptações e variações, como um dos principais para o cálculo de probabilidades de resultados de jogos de futebol.

Sua vantagem é que o número esperado de gols de um determinado time pode mudar, dependendo do seu adversário - o que faz sentido.

X_{ij}: v.a. número de gols do time i (mandante) jogando contra o time j.

X_{ij}: v.a. número de gols do time i (mandante) jogando contra o time j.

Y_{ij}: v.a. número de gols do time j jogando contra o time i (mandante).

 Modelo de Maher (1982)

X_{ij}: v.a. número de gols do time i (mandante) jogando contra o time j.

Y_{ij}: v.a. número de gols do time j jogando contra o time i (mandante).

Placar observado: (x_{ij}, y_{ij}).

 Modelo de Maher (1982)

X_{ij}: v.a. número de gols do time i (mandante) jogando contra o time j.

Y_{ij}: v.a. número de gols do time j jogando contra o time i (mandante).

Placar observado: (x_{ij}, y_{ij}).

Por suposição, X_{ij} e Y_{ij} são independentes.

 Modelo de Maher (1982)

$$X_{ij} \sim Poisson(\alpha_i \beta_j)$$

 Modelo de Maher (1982)

$$X_{ij} \sim Poisson(\alpha_i \beta_j)$$

$$Y_{ij} \sim Poisson(\gamma_i \delta_j)$$

 Modelo de Maher (1982)

$X_{ij} \sim Poisson(\alpha_i \beta_j)$

$Y_{ij} \sim Poisson(\gamma_i \delta_j)$

α_i: parâmetro de força do ataque do time i quando é mandante.

$X_{ij} \sim Poisson(\alpha_i \beta_j)$

$Y_{ij} \sim Poisson(\gamma_i \delta_j)$

α_i: parâmetro de força do ataque do time i quando é mandante.
β_j: parâmetro de fragilidade da defesa do time j quando não é mandante.

 Modelo de Maher (1982)

$X_{ij} \sim Poisson(\alpha_i \beta_j)$

$Y_{ij} \sim Poisson(\gamma_i \delta_j)$

α_i: parâmetro de força do ataque do time i quando é mandante.
β_j: parâmetro de fragilidade da defesa do time j quando não é mandante.
γ_i: parâmetro de fragilidade da defesa do time i quando é mandante.

$X_{ij} \sim Poisson(\alpha_i \beta_j)$

$Y_{ij} \sim Poisson(\gamma_i \delta_j)$

α_i: parâmetro de força do ataque do time i quando é mandante.
β_j: parâmetro de fragilidade da defesa do time j quando não é mandante.
γ_i: parâmetro de fragilidade da defesa do time i quando é mandante.
β_j: parâmetro de força do ataque do time j quando não é mandante.

 Modelo de Maher (1982)

Para os placares dos jogos dos times mandantes, a função de log-verossimilhança será

Para os placares dos jogos dos times mandantes, a função de log-verossimilhança será

$$\log L(\underset{\sim}{\alpha}, \underset{\sim}{\beta}) = \sum_{i} \sum_{j \neq i} (-\alpha_i \beta_j + x_{ij} \log(\alpha_i \beta_j - \log(x_{ij}!)).$$

Para os placares dos jogos dos times mandantes, a função de log-verossimilhança será

$$\log L(\underset{\sim}{\alpha}, \underset{\sim}{\beta}) = \sum_{i} \sum_{j \neq i} (-\alpha_i \beta_j + x_{ij} \log(\alpha_i \beta_j - \log(x_{ij}!)).$$

Sua derivada será

Para os placares dos jogos dos times mandantes, a função de log-verossimilhança será

$$\log L(\underset{\sim}{\alpha}, \underset{\sim}{\beta}) = \sum_i \sum_{j \neq i} (-\alpha_i \beta_j + x_{ij} \log(\alpha_i \beta_j - \log(x_{ij}!)).$$

Sua derivada será

$$\frac{\partial \log L}{\partial \alpha_i} = \sum_{j \neq i} \left(-\beta_j + \frac{x_{ij}}{\alpha_i} \right).$$

Para os placares dos jogos dos times mandantes, a função de log-verossimilhança será

$$\log L(\underset{\sim}{\alpha}, \underset{\sim}{\beta}) = \sum_{i} \sum_{j \neq i} (-\alpha_i \beta_j + x_{ij} \log(\alpha_i \beta_j - \log(x_{ij}!)).$$

Sua derivada será

$$\frac{\partial \log L}{\partial \alpha_i} = \sum_{j \neq i} \left(-\beta_j + \frac{x_{ij}}{\alpha_i} \right).$$

Estimadores de máxima verossimilhança:

$$\hat{\alpha}_i = \frac{\sum_{j \neq i} x_{ij}}{\sum_{j \neq i} \hat{\beta}_j} \text{ e } \hat{\beta}_j = \frac{\sum_{i \neq j} x_{ij}}{\sum_{i \neq j} \hat{\alpha}_i}.$$

Para os placares dos jogos dos times mandantes, a função de log-verossimilhança será

$$\log L(\underset{\sim}{\alpha}, \underset{\sim}{\beta}) = \sum_i \sum_{j \neq i} (-\alpha_i \beta_j + x_{ij} \log(\alpha_i \beta_j - \log(x_{ij}!)).$$

Sua derivada será

$$\frac{\partial \log L}{\partial \alpha_i} = \sum_{j \neq i} \left(-\beta_j + \frac{x_{ij}}{\alpha_i} \right).$$

Estimadores de máxima verossimilhança:

$$\hat{\alpha}_i = \frac{\sum_{j \neq i} x_{ij}}{\sum_{j \neq i} \hat{\beta}_j} \text{ e } \hat{\beta}_j = \frac{\sum_{i \neq j} x_{ij}}{\sum_{i \neq j} \hat{\alpha}_i}.$$

De maneira similar, $\underset{\sim}{\gamma}$ e $\underset{\sim}{\delta}$ podem ser determinados a partir de y_{ij}.

	mandante	alpha	beta_emv$beta	gamma_emv$gamma	delta_emv$delta
1	AmericaMG	0.8700835	1.6782105	1.7186895	1.1596444
2	AthleticoPR	1.4246981	0.8613756	1.0501831	0.8930977
3	AtleticoMG	1.0121805	0.4692781	0.8215533	0.9920743
4	Bahia	1.2037303	1.2785438	0.8690954	0.7736405
5	Botafogo	1.5468597	0.5294916	0.6994222	0.8759476
6	Corinthians	1.1398592	1.0386351	0.8690954	0.7736405
7	Coritiba	0.7785343	1.9495918	1.4207054	1.1399486
8	Cruzeiro	0.4373851	0.6397333	0.6472209	1.0372235
9	Cuiaba	0.8773180	0.6994903	1.0979242	0.7275808
10	Flamengo	0.9477772	1.1696603	0.6599770	1.3657188
11	Fluminense	1.4052200	1.2908602	0.8110395	0.7711665
12	Fortaleza	1.1318954	0.8966659	0.7458740	0.6037493
13	Goias	0.8936019	1.0733668	1.2178061	0.7888413
14	Gremio	1.5245546	1.5386871	0.8323569	1.2132577
15	Internacional	1.2799313	0.9029743	1.1447005	0.5611466
16	Palmeiras	1.2680395	0.7124718	0.7225544	1.4252197
17	RedBullBragantino	1.1240481	0.7548075	0.5863870	0.9793803
18	Santos	1.2692586	1.5200278	1.3211153	0.6234329
19	SaoPaulo	1.3387488	1.0484800	0.7412004	0.4938501
20	Vasco	0.8477809	1.1642135	0.9884548	0.8344041

 Modelo de Maher (1982)

Obviamente, os parâmetros são estáticos, e representam a força "média" dos times ao longo de um período (no caso, ao longo de um campeonato completo).

Exemplo: cálculo dos parâmetros da distribuição de Poisson de um jogo entre Internacional (mandante) e Fluminense.

Exemplo: cálculo dos parâmetros da distribuição de Poisson de um jogo entre Internacional (mandante) e Fluminense.

$\lambda_{INT} = 1,28 \times 1,29 = 1,65$

Exemplo: cálculo dos parâmetros da distribuição de Poisson de um jogo entre Internacional (mandante) e Fluminense.

$\lambda_{INT} = 1,28 \times 1,29 = 1,65$

$\lambda_{FLU} = 1,14 \times 0.77 = 0,88$

Para o cálculo de probabilidades:

Para o cálculo de probabilidades:

$$P(A > B) = \sum_{k=0}^{\infty} P(A > B | B = k) P(B = k)$$

$$= \sum_{k=0}^{\infty} P(A \geq k+1) P(B = k) = \sum_{k=0}^{\infty} \left(\sum_{l=k+1}^{\infty} \frac{\lambda_A^l e^{-\lambda_A}}{l!} \right) \frac{\lambda_B^k e^{-\lambda_B}}{k!}.$$

Para o cálculo de probabilidades:

$$P(A > B) = \sum_{k=0}^{\infty} P(A > B|B = k)P(B = k)$$

$$= \sum_{k=0}^{\infty} P(A \geq k+1)P(B = k) = \sum_{k=0}^{\infty} \left(\sum_{l=k+1}^{\infty} \frac{\lambda_A^l e^{-\lambda_A}}{l!} \right) \frac{\lambda_B^k e^{-\lambda_B}}{k!}.$$

Prob. de vitória do Internacional: 55,54%

Para o cálculo de probabilidades:

$$P(A > B) = \sum_{k=0}^{\infty} P(A > B|B = k)P(B = k)$$

$$= \sum_{k=0}^{\infty} P(A \geq k+1)P(B = k) = \sum_{k=0}^{\infty} \left(\sum_{l=k+1}^{\infty} \frac{\lambda_A^l e^{-\lambda_A}}{l!} \right) \frac{\lambda_B^k e^{-\lambda_B}}{k!}.$$

Prob. de vitória do Internacional: 55,54%
Prob. de empate: 24,45%

Para o cálculo de probabilidades:

$$P(A > B) = \sum_{k=0}^{\infty} P(A > B | B = k) P(B = k)$$

$$= \sum_{k=0}^{\infty} P(A \geq k+1) P(B = k) = \sum_{k=0}^{\infty} \left(\sum_{l=k+1}^{\infty} \frac{\lambda_A^l e^{-\lambda_A}}{l!} \right) \frac{\lambda_B^k e^{-\lambda_B}}{k!}.$$

Prob. de vitória do Internacional: 55,54%
Prob. de empate: 24,45%
Prob. de vitória do Fluminense: 20,01%

Após considerar diferentes estruturas de modelos, Maher concluiu que não são necessários tantos parâmetros.

Após considerar diferentes estruturas de modelos, Maher concluiu que não são necessários tantos parâmetros.

Ele chegou a um modelo que tem uma capacidade muito boa de ajuste e tem quase metade do número de parâmetros

 Modelo de Maher (1982)

Após considerar diferentes estruturas de modelos, Maher concluiu que não são necessários tantos parâmetros.

Ele chegou a um modelo que tem uma capacidade muito boa de ajuste e tem quase metade do número de parâmetros

Na parametrização anterior, se n times são considerados, o modelo possui $2(2n-1)$ parâmetros.

 Modelo de Maher (1982)

Após considerar diferentes estruturas de modelos, Maher concluiu que não são necessários tantos parâmetros.

Ele chegou a um modelo que tem uma capacidade muito boa de ajuste e tem quase metade do número de parâmetros

Na parametrização anterior, se n times são considerados, o modelo possui $2(2n-1)$ parâmetros.

Agora, o modelo possuirá $2n$ parâmetros.

 Modelo de Maher (1982)

Esse modelo considera, para todos os times, que a força de ataque do time visitante é igual à de quando ele é mandante multiplicada por uma constante, digamos η.

Esse modelo considera, para todos os times, que a força de ataque do time visitante é igual à de quando ele é mandante multiplicada por uma constante, digamos η.

O mesmo vale para a fragilidade da defesa, e a mesma constante η é considerada.

Esse modelo considera, para todos os times, que a força de ataque do time visitante é igual à de quando ele é mandante multiplicada por uma constante, digamos η.

O mesmo vale para a fragilidade da defesa, e a mesma constante η é considerada.

Ou seja, considera-se que $\delta_i = \eta\alpha_i$ e $\gamma_j = \eta\beta_j$.

 Modelo de Maher (1982)

Esse modelo considera, para todos os times, que a força de ataque do time visitante é igual à de quando ele é mandante multiplicada por uma constante, digamos η.

O mesmo vale para a fragilidade da defesa, e a mesma constante η é considerada.

Ou seja, considera-se que $\delta_i = \eta\alpha_i$ e $\gamma_j = \eta\beta_j$.

Nessa parametrização, η tipicamente tem valor menor que um.

 Modelo de Maher (1982)

$X_{ij} \sim Poisson(\alpha_i \beta_j)$

 Modelo de Maher (1982)

$X_{ij} \sim Poisson(\alpha_i \beta_j)$
$Y_{ij} \sim Poisson(\eta^2 \alpha_j \beta_i)$

 Modelo de Maher (1982)

$X_{ij} \sim Poisson(\alpha_i \beta_j)$
$Y_{ij} \sim Poisson(\eta^2 \alpha_j \beta_i)$

Estimadores de máxima verossimilhança:

 Modelo de Maher (1982)

$X_{ij} \sim Poisson(\alpha_i \beta_j)$
$Y_{ij} \sim Poisson(\eta^2 \alpha_j \beta_i)$

Estimadores de máxima verossimilhança:

$$\hat{\alpha}_i = \frac{\sum_{j \neq i}(x_{ij} + y_{ji})}{(1 + \eta^2) \sum_{j \neq i} \hat{\beta}_j},$$

$X_{ij} \sim Poisson(\alpha_i \beta_j)$
$Y_{ij} \sim Poisson(\eta^2 \alpha_j \beta_i)$

Estimadores de máxima verossimilhança:

$$\hat{\alpha}_i = \frac{\sum_{j \neq i}(x_{ij} + y_{ji})}{(1+\eta^2)\sum_{j \neq i} \hat{\beta}_j},$$

$$\hat{\beta}_j = \frac{\sum_{i \neq j}(x_{ij} + y_{ji})}{(1+\eta^2)\sum_{i \neq j} \hat{\alpha}_i}, \text{ e}$$

$X_{ij} \sim Poisson(\alpha_i \beta_j)$
$Y_{ij} \sim Poisson(\eta^2 \alpha_j \beta_i)$

Estimadores de máxima verossimilhança:

$$\hat{\alpha}_i = \frac{\sum_{j \neq i}(x_{ij} + y_{ji})}{(1+\eta^2)\sum_{j \neq i} \hat{\beta}_j},$$

$$\hat{\beta}_j = \frac{\sum_{i \neq j}(x_{ij} + y_{ji})}{(1+\eta^2)\sum_{i \neq j} \hat{\alpha}_i}, \text{ e}$$

$$\hat{\eta}^2 = \frac{\sum_i \sum_{j \neq i} y_{ij}}{\sum_i \sum_{j \neq i} x_{ij}}.$$

	mandante	alpha	beta_emv$beta
1	AmericaMG	1.450611	1.3553604
2	AthleticoPR	1.695782	0.7573290
3	AtleticoMG	1.445568	0.5003501
4	Bahia	1.443064	0.8755340
5	Botafogo	1.777452	0.4852850
6	Corinthians	1.395456	0.7699877
7	Coritiba	1.373783	1.3724256
8	Cruzeiro	1.038729	0.5136943
9	Cuiaba	1.164005	0.7017352
10	Flamengo	1.657986	0.7562933
11	Fluminense	1.593276	0.8593222
12	Fortaleza	1.273464	0.6630364
13	Goias	1.217947	0.9098725
14	Gremio	1.988049	0.9780546
15	Internacional	1.361334	0.8106173
16	Palmeiras	1.936404	0.5729919
17	RedBullBragantino	1.523905	0.5435749
18	Santos	1.391585	1.1444176
19	SaoPaulo	1.353985	0.7272858
20	Vasco	1.214646	0.8684126

$$\hat{\eta}^2 = 0,72$$

Nesse modelo, os parâmetros têm uma interpretação interessante.

 Modelo de Maher (1982)

Nesse modelo, os parâmetros têm uma interpretação interessante.

Seja $\lambda_{ij} = \alpha_i \beta_j$ o parâmetro do time mandante e $\lambda_{ji} = \eta^2 \alpha_j \beta_i$ o parâmetro do time visitante.

Nesse modelo, os parâmetros têm uma interpretação interessante.

Seja $\lambda_{ij} = \alpha_i \beta_j$ o parâmetro do time mandante e $\lambda_{ji} = \eta^2 \alpha_j \beta_i$ o parâmetro do time visitante.

Excluindo-se o efeito de se jogar em casa (ou seja $\eta = 1$), se $\lambda_{ij} = \lambda_{ji}$, os números esperados de gols dos dois times são iguais. Logo, pode-se pensar que os times têm forças iguais se $\alpha_i \beta_j = \alpha_j \beta_i$.

Nesse modelo, os parâmetros têm uma interpretação interessante.

Seja $\lambda_{ij} = \alpha_i \beta_j$ o parâmetro do time mandante e $\lambda_{ji} = \eta^2 \alpha_j \beta_i$ o parâmetro do time visitante.

Excluindo-se o efeito de se jogar em casa (ou seja $\eta = 1$), se $\lambda_{ij} = \lambda_{ji}$, os números esperados de gols dos dois times são iguais. Logo, pode-se pensar que os times têm forças iguais se $\alpha_i \beta_j = \alpha_j \beta_i$.

Se $\alpha_i \beta_j > \alpha_j \beta_i$, considera-se que o time i é mais forte que o time j, ou, equivalentemente, se $\alpha_i / \beta_i > \alpha_j / \beta_j$.

Nesse modelo, os parâmetros têm uma interpretação interessante.

Seja $\lambda_{ij} = \alpha_i \beta_j$ o parâmetro do time mandante e $\lambda_{ji} = \eta^2 \alpha_j \beta_i$ o parâmetro do time visitante.

Excluindo-se o efeito de se jogar em casa (ou seja $\eta = 1$), se $\lambda_{ij} = \lambda_{ji}$, os números esperados de gols dos dois times são iguais. Logo, pode-se pensar que os times têm forças iguais se $\alpha_i \beta_j = \alpha_j \beta_i$.

Se $\alpha_i \beta_j > \alpha_j \beta_i$, considera-se que o time i é mais forte que o time j, ou, equivalentemente, se $\alpha_i / \beta_i > \alpha_j / \beta_j$.

Logo, α_i / β_i pode ser usado como medida de força do time i.

mandante	alpha	beta_emv$beta	forca	posicao	posicao_campeonato
Botafogo	1.777452	0.4852850	3.662697	1	1
Palmeiras	1.936404	0.5729919	3.379461	2	2
AtleticoMG	1.445568	0.5003501	2.889113	3	5
RedBullBragantino	1.523905	0.5435749	2.803487	4	3
AthleticoPR	1.695782	0.7573290	2.239161	5	7
Flamengo	1.657986	0.7562933	2.192253	6	6
Gremio	1.988049	0.9780546	2.032657	7	4
Cruzeiro	1.038729	0.5136943	2.022076	8	16
Fortaleza	1.273464	0.6630364	1.920655	9	9
SaoPaulo	1.353985	0.7272858	1.861697	10	10
Fluminense	1.593276	0.8593222	1.854108	11	8
Corinthians	1.395456	0.7699877	1.812309	12	12
Internacional	1.361334	0.8106173	1.679379	13	13
Cuiaba	1.164005	0.7017352	1.658753	14	11
Bahia	1.443064	0.8755340	1.648210	15	14
Vasco	1.214646	0.8684126	1.398697	16	17
Goias	1.217947	0.9098725	1.338591	17	18
Santos	1.391585	1.1444176	1.215977	18	15
AmericaMG	1.450611	1.3553604	1.070277	19	20
Coritiba	1.373783	1.3724256	1.000989	20	19

Exemplo: cálculo dos parâmetros da distribuição de Poisson de um jogo entre Internacional (mandante) e Fluminense.

Exemplo: cálculo dos parâmetros da distribuição de Poisson de um jogo entre Internacional (mandante) e Fluminense.

$\lambda_{INT} = 1,36 \times 0,86 = 1,17$

 Modelo de Maher (1982)

Exemplo: cálculo dos parâmetros da distribuição de Poisson de um jogo entre Internacional (mandante) e Fluminense.

$\lambda_{INT} = 1,36 \times 0,86 = 1,17$

$\lambda_{FLU} = 0,72 \times 1,59 \times 0.81 = 0,93$

Para o cálculo de probabilidades:

 Modelo de Maher (1982)

Para o cálculo de probabilidades:

Prob. de vitória do Internacional: 41,24%

Para o cálculo de probabilidades:

Prob. de vitória do Internacional: 41,24%
Prob. de empate: 29,65%

 Modelo de Maher (1982)

Para o cálculo de probabilidades:

Prob. de vitória do Internacional: 41,24%
Prob. de empate: 29,65%
Prob. de vitória do Fluminense: 29,11%

 Modelo de Maher (1982)

O ajuste é surpreendentemente bom para um modelo tão simpes!

O ajuste é surpreendentemente bom para um modelo tão simpes!

Principalmente levando-se em conta que o número de gols de uma equipe em uma partida não deve ser independente do número de gols da outra equipe.

Dixon e Coles (1997), em outro trabalho muito citado, incorporaram a estrutura de dependência em um modelo mais elaborado.

 Modelo de Maher (1982)

Dixon e Coles (1997), em outro trabalho muito citado, incorporaram a estrutura de dependência em um modelo mais elaborado.

Os autores sugerem mais uma melhoria no modelo, no sentido de que as partidas mais recentes podem ter mais peso na estimação dos parâmetros.

Dixon e Coles (1997), em outro trabalho muito citado, incorporaram a estrutura de dependência em um modelo mais elaborado.

Os autores sugerem mais uma melhoria no modelo, no sentido de que as partidas mais recentes podem ter mais peso na estimação dos parâmetros.

Com essa abordagem, é possível até mesmo elaborar um modelo em que os parâmetros variam ao longo do tempo, conforme trabalharam Baker e McHale (2015).

Dixon e Coles (1997), em outro trabalho muito citado, incorporaram a estrutura de dependência em um modelo mais elaborado.

Os autores sugerem mais uma melhoria no modelo, no sentido de que as partidas mais recentes podem ter mais peso na estimação dos parâmetros.

Com essa abordagem, é possível até mesmo elaborar um modelo em que os parâmetros variam ao longo do tempo, conforme trabalharam Baker e McHale (2015).

Além disso, é possível considerar outras informações, como desfalques dos times, trocas de técnico etc.

Dixon e Coles (1997), em outro trabalho muito citado, incorporaram a estrutura de dependência em um modelo mais elaborado.

Os autores sugerem mais uma melhoria no modelo, no sentido de que as partidas mais recentes podem ter mais peso na estimação dos parâmetros.

Com essa abordagem, é possível até mesmo elaborar um modelo em que os parâmetros variam ao longo do tempo, conforme trabalharam Baker e McHale (2015).

Além disso, é possível considerar outras informações, como desfalques dos times, trocas de técnico etc.

São refinamentos na modelagem que não serão tratados aqui neste minicurso.

Para finalizar, vamos calcular a probabilidade de cada time ser campeão!

 Modelo de Maher (1982)

Para finalizar, vamos calcular a probabilidade de cada time ser campeão!

Como temos as probabilidades dos resultados de cada jogo, é possível simular os resultados de cada um dos jogos restantes.

Para finalizar, vamos calcular a probabilidade de cada time ser campeão!

Como temos as probabilidades dos resultados de cada jogo, é possível simular os resultados de cada um dos jogos restantes.

Simulamos milhares de cenários a partir da 32a. rodada e observamos o time campeão em cada um. A probabilidade estimada de um dado time ser campeão será o percentual de vezes que o time termina em primeiro lugar nesses milhares de cenários.

Resultados de 10.000 simulações:

Resultados de 10.000 simulações:

Athlético-PR: 0,01%

Resultados de 10.000 simulações:

Athlético-PR: 0,01%
Flamengo: 0,18%

Resultados de 10.000 simulações:

Athlético-PR: 0,01%
Flamengo: 0,18%
Atlético-MG: 0,47%

Resultados de 10.000 simulações:

Athlético-PR: 0,01%
Flamengo: 0,18%
Atlético-MG: 0,47%
Grêmio: 0,83%

Resultados de 10.000 simulações:

Athlético-PR: 0,01%
Flamengo: 0,18%
Atlético-MG: 0,47%
Grêmio: 0,83%
Red Bull Bragantino: 8,26%

 Modelo de Maher (1982)

Resultados de 10.000 simulações:

Athlético-PR: 0,01%
Flamengo: 0,18%
Atlético-MG: 0,47%
Grêmio: 0,83%
Red Bull Bragantino: 8,26%
Palmeiras: 12,33%

Resultados de 10.000 simulações:

Athlético-PR: 0,01%
Flamengo: 0,18%
Atlético-MG: 0,47%
Grêmio: 0,83%
Red Bull Bragantino: 8,26%
Palmeiras: 12,33%
Botafogo: 77,92%

Atualização de última hora! Após a 32a. rodada:

Atualização de última hora! Após a 32a. rodada:

Atlético-MG: 0,17%

 Modelo de Maher (1982)

Atualização de última hora! Após a 32a. rodada:

Atlético-MG: 0,17%
Flamengo: 0,77%

Atualização de última hora! Após a 32a. rodada:

Atlético-MG: 0,17%
Flamengo: 0,77%
Grêmio: 1,35%

Atualização de última hora! Após a 32a. rodada:

Atlético-MG: 0,17%
Flamengo: 0,77%
Grêmio: 1,35%
Red Bull Bragantino: 18,24%

Atualização de última hora! Após a 32a. rodada:

Atlético-MG: 0,17%
Flamengo: 0,77%
Grêmio: 1,35%
Red Bull Bragantino: 18,24%
Palmeiras: 24,37%

Atualização de última hora! Após a 32a. rodada:

Atlético-MG: 0,17%
Flamengo: 0,77%
Grêmio: 1,35%
Red Bull Bragantino: 18,24%
Palmeiras: 24,37%
Botafogo: 55,10%

 Modelo de Maher (1982)

Dando maior peso para as últimas rodadas:

 Modelo de Maher (1982)

Dando maior peso para as últimas rodadas:

Atlético-MG: 0,21%

Dando maior peso para as últimas rodadas:

Atlético-MG: 0,21%
Flamengo: 0,30%

 Modelo de Maher (1982)

Dando maior peso para as últimas rodadas:

Atlético-MG: 0,21%
Flamengo: 0,30%
Grêmio: 2,11%

Dando maior peso para as últimas rodadas:

Atlético-MG: 0,21%
Flamengo: 0,30%
Grêmio: 2,11%
Red Bull Bragantino: 26,78%

Dando maior peso para as últimas rodadas:

Atlético-MG: 0,21%
Flamengo: 0,30%
Grêmio: 2,11%
Red Bull Bragantino: 26,78%
Palmeiras: 34,81%

Dando maior peso para as últimas rodadas:

Atlético-MG: 0,21%
Flamengo: 0,30%
Grêmio: 2,11%
Red Bull Bragantino: 26,78%
Palmeiras: 34,81%
Botafogo: 35,78%

Neste minicurso, exploramos alguns dos modelos mais conhecidos, baseados na distribuição de Poisson, para o cálculo de probabilidades no futebol.

Neste minicurso, exploramos alguns dos modelos mais conhecidos, baseados na distribuição de Poisson, para o cálculo de probabilidades no futebol.

Tais cálculos podem ser utilizados para mensurar as chances de seu time ser campeão ou ser rebaixado, que tanto aparecem na imprensa ao final dos campeonatos.

Neste minicurso, exploramos alguns dos modelos mais conhecidos, baseados na distribuição de Poisson, para o cálculo de probabilidades no futebol.

Tais cálculos podem ser utilizados para mensurar as chances de seu time ser campeão ou ser rebaixado, que tanto aparecem na imprensa ao final dos campeonatos.

Também são úteis no contexto de apostas esportivas, em que as cotações são calculadas baseadas justamente nas probabilidades.

- Baker, R. D., McHale, I. G. (2015). Time varying ratings in association football: the all-time greatest team is... *Journal of the Royal Statistical Society Series A* 178, n. 2, pp. 481-492.
- Dixon, M. J., Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics* 46, n. 2, pp. 265-280.
- Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica* 36, n. 3, pp. 109-118.
- Meyer, P. L. (1983). *Probabilidade: Aplicações à Estatística*, 2a. ed., Livros Técnicos e Científicos Editora SA.

Obrigado!