



**LUCIANO RIBEIRO GALVÃO**

**ESTIMAÇÃO DE RATING NO FUTEBOL: CAMPEONATO  
BRASILEIRO DE 2017**

**LAVRAS – MG**

**2018**

**LUCIANO RIBEIRO GALVÃO**

**ESTIMAÇÃO DE RATING NO FUTEBOL: CAMPEONATO BRASILEIRO DE 2017**

Dissertação a ser apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Professor Júlio Sílvio de Sousa Bueno Filho  
Orientador

**LAVRAS – MG**  
**2018**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca  
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Galvão, Luciano Ribeiro.

Estimação de *Rating* no Futebol: Campeonato Brasileiro de  
2017 / Luciano Ribeiro Galvão. - 2018.

50 p. : il.

Orientador(a): Júlio Sílvio de Sousa Bueno Filho.

.  
Dissertação (mestrado acadêmico) - Universidade Federal de  
Lavras, 2018.

Bibliografia.

1. Campeonato Brasileiro. 2. Predição. 3. Distribuição de  
Poisson. I. Bueno Filho, Júlio Sílvio de Sousa. . II. Título.

**LUCIANO RIBEIRO GALVÃO**

**ESTIMAÇÃO DE RATING NO FUTEBOL: CAMPEONATO BRASILEIRO DE 2017**

Dissertação a ser apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 06 de julho de 2018.

Prof. Dr. Paulo Henrique Sales Guimarães UFLA  
Prof. Dr. Deive Ciro de Oliveira Unifal

Professor Júlio Sílvio de Sousa Bueno Filho  
Orientador

**LAVRAS – MG  
2018**

*Dedico: À minha mãe Andrea, por todo apoio, ensinamentos e cuidado. Ao meu irmão  
Leandro pelo exemplo de caráter e dedicação*

## **AGRADECIMENTOS**

À minha família pelo apoio incondicional, em todas as horas.

Aos meus amigos do mestrado, pelos momentos divididos e desafios enfrentados.

Aos professores do departamento de Estatísticas de Universidade Federal de Lavras, pelos ensinamentos, em especial à meu orientador, prof. Júlio Bueno, que me acompanhou desde a graduação, e que ao fim desse processo se tornou um grande amigo, com ensinamentos não só profissionais mas também para a vida.

À todos os funcionários do departamento de Estatística À Universidade Federal de Lavras, que se tornou uma segunda casa.

## RESUMO

Observando os jogos da série A do campeonato brasileiro de 2017, ajustou-se modelos Poisson com preditores lineares para a esperança do número de gols do time mandante e do time visitante. Estimou-se parâmetros referentes à força de ataque ( $\gamma$ ) e de defesa ( $\delta$ ) de cada time. Foi utilizado um efeito de mando de campo comum ( $\mu$ ) para todos os clubes. O primeiro turno da competição (dos dois turnos completos) foi usado para a estimação inicial dos *ratings*, que por sua vez foram reestimados no segundo turno, jogo a jogo. Pretende-se com esse estudo apresentar um modelo descritivo que permita tanto estimar quanto prever as probabilidades associadas aos placares e, ou as associadas aos resultados (vitória, empate ou derrota) para o mandante e para o visitante em cada partida. Como aspectos adicionais podem-se prever a classificação final do campeonato e resultados derivados como as probabilidades de classificação para a Copa Libertadores da América ou de ser rebaixado para a segunda divisão da competição. O modelo proposto se mostrou flexível em sua versão de análise de verossimilhança, mas de suas propriedades depreende-se que a implementação paramétrica bayesiana pode incorporar outras características desejáveis do modelo.

**Palavras-chave:** Campeonato Brasileiro. Força de Ataque. Força de Defesa. Distribuição de Poisson. Predição.

## ABSTRACT

Observing games from 2017 Brazilian Football Championship League series A, we adjusted Poisson models with linear predictors for the expectation of the number of goals from home and away teams. We estimated attacking ( $\gamma$ ) and defense ( $\delta$ ) strength parameters for each team. A common home advantage effect ( $\mu$ ) was used for all teams. The first leg of the double round-robin was used to estimate initial ratings. Those parameters were then re-estimated at each round of the second leg. We intend to present a descriptive model that could estimate as well as predict scores and results from each game. Additional features of the final table could also be predicted as probability of playing South American Champions League or being demoted to series B of the competition. Model has been shown flexible in the pure likelihood analysis version, but some of desirable features of parametric Bayesian inference could enhance some of its properties and will be object of future research.

**Keywords:** Brazilian football. Attacking Strength. Defensive Strength. League Championship. Poisson distribution. Prediction.



## LISTA DE FIGURAS

Figura 4.1 – Diagrama de dispersão para gols pró e $\hat{\gamma}_i$ . . . . .	27
Figura 4.2 – Diagrama de dispersão para gols contra e $\hat{\delta}_i$ . . . . .	28
Figura 4.3 – Estimativas para os parâmetros de ataque ( $\hat{\gamma}_i$ ) produzidas a cada rodada do segundo turno para todos os times do Campeonato Brasileiro série A de 2017	32
Figura 4.4 – Estimativas para os parâmetros de defesa ( $\hat{\delta}_i$ ) produzidas a cada rodada do segundo turno para todos os times do Campeonato Brasileiro série A de 2017	33
Figura 4.5 – Diagrama de Dispersão para Gols Pró e $\hat{\gamma}_i$ . . . . .	34
Figura 4.6 – Diagrama de Dispersão para Gols Contra e $\hat{\delta}_i$ . . . . .	34
Figura 4.7 – Elipses contendo os <i>ratings</i> relativos dos times no primeiro turno do Campeonato Brasileiro de 2017 . . . . .	36
Figura 4.8 – Elipses contendo os <i>ratings</i> relativos dos times no segundo turno do Campeonato Brasileiro de 2017 . . . . .	36

## LISTA DE TABELAS

Tabela 3.1 – Relação de times que participaram da primeira divisão do campeonato Brasileiro de 2017 . . . . .	20
Tabela 3.2 – Representação dos parâmetros . . . . .	22
Tabela 4.1 – Estimativa dos <i>ratings</i> relativos a ataque para os times do primeiro turno do Brasileirão de 2017. . . . .	25
Tabela 4.2 – Estimativa dos <i>ratings</i> relativos a defesa para os times do primeiro turno do Brasileirão de 2017. . . . .	26
Tabela 4.3 – Valores de Correlação entre os Parâmetros $\hat{\gamma}$ e $\hat{\delta}$ com Gols Pró e Gols contra no Primeiro Turno . . . . .	28
Tabela 4.4 – Estimativa dos <i>ratings</i> relativos a ataque para os times do segundo turno do Brasileirão de 2017. . . . .	30
Tabela 4.5 – Estimativa dos <i>ratings</i> relativos a defesa para os times do segundo turno do Brasileirão de 2017. . . . .	31
Tabela 4.6 – Valores de Correlação entre os Parâmetros $\hat{\gamma}$ e $\hat{\delta}$ com gols pró e gols contra, sucessivamente: Segundo Turno . . . . .	35
Tabela 4.7 – Probabilidades de vitória, empate e derrota para o time mandante. . . . .	39
Tabela 4.8 – Probabilidades para cada possível combinação de gols marcados em uma partida, considerando $\gamma_A = \gamma_B = \delta_A = \delta_B = 1$ ) e $\mu = -0.18$ . Os valores na matriz triangular superior representam vitórias do time mandante e na matriz triangular inferior, do visitante. . . . .	39
Tabela 4.9 – Probabilidades para cada possível placar no jogo Corinthians vs São Paulo. Estimativas associadas: $\hat{\gamma}_{COR} = 0.1279$ , $\hat{\gamma}_{SPA} = -0,0895$ , $\hat{\delta}_{COR} = 0.9380$ , $\hat{\delta}_{SPA} = 1.6966$ e $\mu = 0.24$ . A matriz triangular superior representa vitórias do mandante e a inferior do visitante. Na diagonal resultados de empate. . . . .	40
Tabela 4.10 – Probabilidades Vitória, Empate e Derrota para o mandante no jogo Corinthians vs São Paulo. Estimativas associadas: $\hat{\gamma}_{COR} = 0.9379$ , $\hat{\gamma}_{SPA} = 1,6366$ , $\hat{\delta}_{COR} = 0.1241$ , $\hat{\delta}_{SPA} = -0.0894$ e $\mu = 0.24$ . . . . .	40
Tabela 4.11 – Probabilidades de gols marcados numa partida entre Palmeiras e Fluminense : $\hat{\gamma}_{PAL} = -0,1896$ , $\hat{\gamma}_{FLU} = 0,4495$ , $\hat{\delta}_{PAL} = 1,3614$ , $\hat{\delta}_{FLU} = 0,2936$ e $\mu = 0,24$ ; Triângulo superior para visitante e inferior pra mandante . . . . .	41

Tabela 4.12 – Probabilidades de vitória, empate e derrota para time Mandante : $PAL_m$ x $FLU_v$ . . . . .	41
Tabela 4.13 – alguns resultados surpreendentes do segundo turno e probabilidades associadas. O valor de referência para testar a hipótese de que os dados se adequam ao modelo é $\chi^2 = 5,99$ para $\nu = 2g.l$ . . . . .	42
Tabela 6.1 – Tabela de Classificação dos Times no Final do Primeiro Turno do Brasileiro de 2017 . . . . .	47
Tabela 6.2 – Tabela de Classificação dos Times no Final do Segundo Turno do Brasileiro de 2017 . . . . .	48

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	10
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	12
<b>2.1</b>	<b>O futebol</b>	12
<b>2.1.1</b>	<b>O futebol no Brasil</b>	12
<b>2.1.2</b>	<b>Campeonato Brasileiro</b>	13
<b>2.2</b>	<b>Estimação e Modelagem</b>	14
<b>2.2.1</b>	<b>Resultados de jogos: placar dos jogos</b>	16
<b>2.3</b>	<b>Análise dos ratings e probabilidades : Descritiva dos dados</b>	18
<b>3</b>	<b>MATERIAL E MÉTODOS</b>	20
<b>3.1</b>	<b>Material</b>	20
<b>3.2</b>	<b>Métodos</b>	21
<b>4</b>	<b>Resultados e Discussão</b>	24
<b>4.1</b>	<b>Primeiro Turno</b>	24
<b>4.2</b>	<b>Segundo Turno</b>	28
<b>4.3</b>	<b>Probabilidades dos placares e resultados</b>	38
<b>4.3.1</b>	<b>Considerações gerais</b>	43
<b>5</b>	<b>Conclusão</b>	44
<b>5.1</b>	<b>Referências</b>	45
<b>6</b>	<b>Anexo</b>	47

## 1 INTRODUÇÃO

A previsão de resultados tem aplicações em vários tipos de esportes e jogos. Esta atividade é mais importante em esportes e jogos organizados em federações profissionais e geram audiência de mídia em todo o mundo, como é o caso do futebol. É comum que o interesse seja maior em medir o desempenho dos times ou atletas favoritos do público, mesmo sem conhecimento estatístico específico. As informações disponibilizadas por métodos de análise variados, através de revistas, programas de televisão e sites da internet, em geral têm justificativa estatística pouco clara e são apresentadas de forma simplificada.

O futebol tem lugar de destaque entre os esportes mais populares devido ao grande número de fãs e praticantes em todo o mundo. O esporte movimenta anualmente altas cifras em contratações milionárias de atletas, patrocínios, comissões técnicas e também em sites de casas de apostas em vários países. A empresa de consultoria ATKearney, estimou que o futebol movimenta anualmente algo entre R\$ 455 bilhões e R\$ 577 bilhões, e que seu "PIB" cresce mais que de países em desenvolvimento, à impressionante taxa de 8% ao ano (Globo, 2017).

A aplicação de métodos estatísticos a partidas de futebol tem importância também para os próprios times e organizações. Supõe-se que de posse das informações geradas pelas análises é possível conduzir melhores estratégias para competições específicas. Isto pode estender-se a informações sobre os atletas do próprio time e de times adversários ou se limitar a previsões de curto prazo sobre o resultado de uma partida.

Na literatura atual é comum encontrar métodos de estimação paramétrica para dados referentes a partidas de futebol usando abordagens variadas. Estes métodos partem de um ponto comum que é a definição da distribuição que melhor se ajusta a dois tipos de observações: os gols marcados numa partida ou o resultado geral (vitória, empate ou derrota do mandante). As distribuições mais usadas no último caso são a binomial (ignorando empates) e a multinomial. No caso do número de gols é em geral adotada a distribuição Poisson ou variações para dar conta da correlação entre os gols na mesma partida e da variação extra-Poisson (usando por exemplo a distribuição binomial negativa).

No presente estudo, optou-se pela distribuição de Poisson para a modelagem dos número de gols para cada time numa partida. O objetivo foi produzir e apresentar estimativas de *ratings*, que são parâmetros associadas à força relativa de ataque e de defesa dos times, além de um parâmetro associado ao mando de campo.

Os *ratings* assumem valores reais para cada time ou participante de determinada competição. Os *ratings* considerados neste processo são sempre relativos aos demais participantes, ou seja, as estimativas dependem do grupo considerado. São possíveis muitas aplicações para as estimativas obtidas, seja para categorizar os participantes ou times da competição, seja para previsões de curto prazo de resultados ou números de gols em partidas, por exemplo.

Os processos de estimação mais comumente empregados para os *ratings* são o de Mínimos Quadrados Ponderados (MQO), o de Mínimos Quadrados Ordinários (MQP), os métodos de inferência Bayesiana e os baseados apenas na verossimilhança. Para este estudo será conduzido a estimação dos *ratings* pelo Método de Máxima Verossimilhança (MMV).

O estudo aborda especificamente o campeonato brasileiro de futebol do ano de 2017, porém as técnicas utilizadas podem ser facilmente estendidas a outros campeonatos ou modalidades esportivas.

## **Objetivo**

O objetivo deste trabalho é estimar os *ratings* relativos a cada time participante na primeira divisão do Campeonato Brasileiro de 2017, e por meio desses valores fazer as comparações entre previsões e resultados observados com números crescentes de rodadas do campeonato. Como forma de validar o modelo pretende-se cotejar os resultados obtidos com o ajuste das rodadas do primeiro turno com os resultados observados a cada rodada a partir do início do segundo turno da competição.

## 2 REFERENCIAL TEÓRICO

### 2.1 O futebol

O futebol é conhecido e praticado mundialmente, sendo atualmente um esporte bastante popular. No último levantamento a *Fédération Internationale de Football Association* (FIFA) realizado em 2007, constatou que 256 milhões de pessoas o praticam regularmente em todo mundo, o número sobe para 270 milhões caso se considere as equipes técnicas também envolvidas.

A origem do futebol remete à China há aproximadamente 3000 anos, num jogo similar chamado *tsu chu*, praticado com uma bola feita de couro de animais que consistia em apenas chutar com máxima força esta bola por hastes de madeiras fixadas na terra, era inicialmente uma prática de treinamento militar e não esportiva. Também é possível encontrar ilustrações que demonstram que um jogo similar era comum na Grécia antiga, na qual recebeu o nome de *Harpastun* (Davidson's, 2000).

Em meados de 1300 na Europa, foi considerada uma prática mal vista, recebendo até mesmo um tratado de proibição do então rei Edward da Inglaterra. Porém, também na Inglaterra e a despeito da proibição se tornou um esporte muito popular e em 1800 já eram disputados campeonatos, inclusive com regras similares às praticadas atualmente. Foi também na Inglaterra que recebeu o nome de *association football* sendo chamado apenas de *football*, como é atualmente chamado, exceto nos Estados Unidos, na Austrália e no Canadá, em que é chamado de *soccer* para diferenciar dos demais tipos de futebol (*American football, Rugby football, Australian football, etc.*).

Foi em 1863 que a primeira associação de futebol internacional foi criada, a FA (*Football Association*). A entidade estabeleceu o conjunto de regras que passaram a ser as oficiais deste esporte e organizou a primeira edição da FA Cup, disputada em 1871 e que ainda persiste como o mais antigo torneio oficial de futebol. Em 1904 a FIFA foi criada e desde então é a autoridade máxima no esporte.

#### 2.1.1 O futebol no Brasil

O futebol é um esporte global, mas os praticantes e amantes do esporte no Brasil julgam que este é o país de referência no esporte, revelando vários atletas de destaque internacional,

tendo vários times importantes em nível nacional e internacional e realizando campeonatos profissionais disputados todos os anos.

O marco inicial do futebol no Brasil foi por meio de Charles William Miller, filho de escoceses nascido no Brás, em São Paulo, 1874, e que após uma estadia em Southampton, na Inglaterra, onde o futebol já era popular, retorna ao Brasil. Nesta ocasião, trouxe duas bolas de couro e deu início à prática na cidade de São Paulo, inicialmente apenas entre imigrantes britânicos.

Em 1894 foi criado o São Paulo Athletic Club, o primeiro clube poliesportivo do Brasil em que se praticou o futebol, porém ainda sem a participação massiva de brasileiros, o que só ocorreu tardiamente, pois o futebol era considerado um esporte de elite e somente membros da pequena comunidade britânica frequentavam o clube.

Entre agosto e setembro de 1910 um famoso time amador inglês (Corinthian Football Club) visitou o Brasil, vencendo o Fluminense por 10 a 1, um combinado carioca por 8 a 1 e um selecionado brasileiro por 5 a 2. Em São Paulo venceram o Associação Atlética das Palmeiras (que não tem relação com a Sociedade Esportiva Palmeiras) por 2 a 0, o Club Atlético Paulistano por 5 a 0 e o São Paulo Athletic Club (com Charles Miller) por 8 a 0. Os placares elásticos inspiraram vários pequenos times, dentre os quais se destaca uma importante iniciativa de popularização do futebol brasileiro, iniciada em setembro de 1910, quando trabalhadores ferroviários de São Paulo se juntaram a imigrantes e operários que jogavam informalmente na várzea e fundaram o Sport Club Corinthians Paulista, cujo nome homenageia os ingleses.

Em 1914 foi fundada a Federação Brasileira de Sports, desta maneira o Brasil já teria um time nacional para competir, e assim nesse ano foi campeão da Copa Roca, que era uma competição que envolvia times da América do Sul.

Em 1923 times menores uniram-se para que pessoas de baixa renda também pudessem ter acesso ao esporte, ainda sob incentivo do governo brasileiro para a prática do futebol por uma fatia maior da população.

### **2.1.2 Campeonato Brasileiro**

Em 1971 o primeiro Campeonato Brasileiro na forma de uma liga envolvendo clubes de todo país foi organizado, que visava suprir a demanda de uma competição nacional, que havia em outros países do mundo, ao contrário do Brasil, em que somente ocorriam competições regionais e estaduais. Durante o processo de consolidação o campeonato recebeu nomes como



Copa do Brasil, Taça de Prata, Taça de Ouro, Copa União, Copa Brasil, Campeonato Nacional de Clubes e Copa João Havelange.

Desde 1994 o campeonato Brasileiro mantém o mesmo formato de pontuação, sendo 3 pontos para o time vitorioso, um ponto cada, no caso de empates, e o time que não vence deixa de pontuar na partida. Em 2003 começou a seguir o sistema de turno e retorno (*round-robin* duplo), no qual todos os times tem a oportunidade ter mando de campo contra cada adversário, e de mesmo modo jogar como visitante.

Este trabalho tem como objeto de estudo o Campeonato Brasileiro devido seu alto nível competitivo e a forma acirrada com que as equipes o disputam, muitas definindo este como foco principal de seus treinamentos e preparo dos atletas. Atualmente, o Brasileirão, que é como o Campeonato Brasileiro é comumente chamado, é dividido em série A, B e C estas com vinte times cada, e também a série D, com quantidade variada de times dependendo do número de inscritos.

O estudo tem como foco a série A desta competição, devido a ser um *round-robin* duplo, o que representa um delineamento que permite melhor estimação dos parâmetros dos clubes. Outro ponto crucial é a importância da série A do campeonato, em que os 4 primeiros colocados são classificados para a Copa Libertadores da América, competição continental Sul Americana equivalente à liga mais rica do mundo (a *Champions League* disputada por toda Europa). Por outro lado, os 4 últimos colocados deixam a série A para a série B... Assim esta competição conta com a participação de 20 times que se confrontam em 10 jogos por rodada, resultando em 380 jogos divididos em dois turnos de 190 jogos cada.

Para o modelo de análise a ser empregado não haveria restrições em estudar o delineamento das séries B, C ou D do brasileiro ou qualquer outro campeonato, mas por simplicidade e potencial de impacto, além de se esperar jogos mais homogêneos com placares menos elásticos, nos limitamos à série A.

## **2.2 Estimação e Modelagem**

No presente estudo estabeleceu-se um conjunto de parâmetros que resulta em um sistema de *ratings* para estimar a força relativa dos times na competição. *Rating* é um valor real que é atribuído com o objetivo de comparação entre indivíduos ou grupos. Esta forma de comparação varia de acordo como o objetivo de cada processo de estimação, neste trabalho

intenta-se que esses valores representem o desempenho de um time com relação aos demais e diferenças no ataque e defesa durante o campeonato.

O rating é estimado considerando somente os resultados observados dos jogos dos times na competição, sendo desconsiderada sua história pregressa. Os ratings representam forças relativas de ataque, de defesa e de mando de campo, para os contendores. Ignora-se o fato de que a força do time oscila com as variações na equipe montada para a competição e demais fatores perturbadores, que são muitos.

Pretende-se estabelecer um modelo de predição referente à condição atual do time e sua relação de força com os demais envolvidos no campeonato. A modelagem irá medir comparativamente os vinte clubes que compõe a série A da competição nos quesitos ataque e defesa, incluindo um parâmetro para medir a vantagem do mando de campo. Considera-se para efeito de estimação somente o placar de cada jogo, além do fator mando de campo, considerando como potencial fator de influência sobre o resultado dos jogos.

Como observado em Silva e Moreira (2008), o mando de campo está relacionado a presença massiva da torcida, familiaridade dos jogadores com a condição do gramado e até a facilidade de deslocamento até a partida. Por estas e demais questões envolvidas, o mando de campo é considerado fator importante.

Há várias outras questões potencialmente perturbadoras do processo de estimação que não são consideradas. Uma questão é que a FIFA no Brasil, desde 2003 define que as dimensões do campo podem variar de 90 a 120 metros no comprimento, e na largura essa medida pode ir de 45 a 90 metros, o que causa um nível de mudança muito grande quando se joga em estádios de dimensões diferentes de gramado, tais como as próprias condições deste. De acordo com Courneya e Carron (1992), Nevill e Holder (1999) e por Pollard (2002) foi visto que variações na dimensão e condições gerais do campo afetam significativamente o desempenho dos atletas na partida, nos aspectos de orientação espacial, velocidade na corrida, distância até a bola e outros fatores inerentes ao uso efetivo do gramado.

Porém, depois das reformas realizadas nos principais estádios do país para receber a Copa do Mundo FIFA de 2014, e com a resolução da CBF (Comissão Brasileira de Futebol) em 2016, todos os jogos disputados no Brasileirão devem ser em campos com as medidas padrões da FIFA, portanto, será desconsiderado este fator no processo de modelagem, incluindo todos no padrão de dimensões internacionais. Assim mando de campo irá apenas diferenciar qual time está jogando em casa ou sendo visitante, desconsiderando o local do estádio, condições gerais,

público médio e outros fatores serão desconsiderados como independentes e serão associados apenas ao parâmetro já definido.

Como é visto em Souza e Gamerman (2004), inicialmente tratou-se os dados com um modelo estático para o processo de estimação dos *ratings*, no trabalho os autores partiram para um processo dinâmico a partir de uma priori de normalidade dos dados referentes aos gols marcados em cada jogo. No estudo que foi conduzido para o Brasileirão do ano de 2017 considera-se um modelo estático porém que é recalculado em cada jogo a partir da 4 rodada do campeonato, considerando o mando de campo apenas para as partidas do segundo turno (rodada 20 adiante).

### **2.2.1 Resultados de jogos: placar dos jogos**

A distribuição que melhor se ajusta à variável aleatória número de gols marcados numa partida (pelo mandante e visitante) varia de acordo com os autores, a maioria usam a distribuição de Poisson e alguns propõem distribuições como a binomial negativa ou outras que acomodem a variação extra-Poisson. Em Moroney(1956) e Reep et al. (1971), o número de gols marcados em uma partida de futebol por determinado time segue uma distribuição Binomial negativa. O mesmo pode ser observado com o modelo preditivo de Maher (1982) e Ridder et al. (1995). Em Karlis (2008) a modelagem de gols marcados por times mandantes e visitantes em partidas de futebol é modelado segundo a distribuição de Skellam, obtendo bons ajustes.

No trabalho de Stevenson (1988), onde se estuda os resultados do campeonato inglês no período de 1946 a 1984, descobriu-se que jogos anteriores à década de 70 obtém melhor ajuste seguindo a distribuição Binomial negativa, enquanto os jogos pós a década de 70 são melhor ajustados quando se usa a distribuição de Poisson ou a Binomial Negativa, em Ridder et Al.(1994) tem-se as mesmas conclusões para o período pós 1970.

Em seu trabalho sobre modelagem no futebol usando ambas as distribuições, Dimitri (2000) consegue ajustes com a Binomial Negativa e também com a Poisson, porém considerou a qualidade do ajuste com a Binomial Negativa inferior, e recomenda o uso de Poisson para este tipo de estudo.

De modo geral trabalhos recentes sobre o tema sugerem que resultado de cada jogo, no que se refere ao número de gols segue uma distribuição de Poisson. Souza e Gamerman (2004) em seu trabalho sobre o campeonato brasileiro de 2002 usaram modelos dinâmicos para fazer previsões acerca do numero de gols dos times na competição, para tal fizeram o ajuste

desta variável aleatória seguindo uma Poisson e verificaram por testes de bondade, concluindo que tem ajuste adequado. Assim como em Araújo(2015) e OLIVEIRA FILHO et al. (2017), também visto em Santos (2013) e Anderson e Sally (2013) que estudaram não só campeonatos nacionais mas também o Campeonato Europeu (*UEFA*) e a Copa do Mundo FIFA, chegando na mesma conclusão sobre a distribuição dos gols por partida.

Portanto com base na literatura definiu-se para o presente estudo o uso da distribuição de Poisson para a modelagem conjunta dos gols feitos por cada time na estimação dos parâmetros de interesse.

Em Santos (2013) o uso de modelagem não-Gaussiana via técnica de verossimilhança para dados referentes aos confrontos entre Brasil e Argentina em edições da Copa do Mundo no período de 1914 à 2008 usa-se a técnica de máxima verossimilhança, assim como é visto em Gamerman, Santos e Franco (2013) para os modelos preditivos baseados em distribuições não normais. O uso da verossimilhança é consagrada tanto em análises frequentistas bem como em Bayesianas portanto optou-se no presente estudo por conduzir estimações dos parâmetros via método da Máxima Verossimilhança.

O uso do logaritmo da verossimilhança não apresenta perda de sentido prático nos dados pois é uma transformação monótona bem como a própria verossimilhança no qual não altera a natureza dos dados. Esta técnica envolve iterações nas equações de estimação dos parâmetros. Será apresentado uma breve descrição do método de Máxima Verossimilhança como pode ser visto em Casella e Berger (2010).

Considere  $f(x|\theta)$  a função de probabilidade ou função densidade de probabilidade conjunta de uma amostra  $\mathbf{X} = (X_1, \dots, X_n)$ . Se  $\mathbf{X} = \mathbf{x}$  é observado, a função de  $\theta$  definida por

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$$

Sendo  $\mathbf{X}$  um vetor aleatório discreto, então  $L(\theta|\mathbf{x}) = \mathbb{P}_\theta(\mathbf{X} = \mathbf{x})$ . Suponha que a função de verossimilhança em dois pontos paramétricos  $\theta_1$  e  $\theta_2$  satisfaçam a seguinte desigualdade

$$\mathbb{P}_{\theta_1}(\mathbf{X} = \mathbf{x}) = L(\theta_1|\mathbf{x}) > L(\theta_2|\mathbf{x}) = \mathbb{P}_{\theta_2}(\mathbf{X} = \mathbf{x})$$

Pode-se concluir que é mais provável que a amostra observada tenha ocorrido se  $\theta = \hat{\theta}_1$  do que se  $\theta = \hat{\theta}_2$ .

O método da máxima verossimilhança pode ser usado de modo direto por meio de produtos, dos  $n$  termos com que se esteja trabalhando. No caso de amostras, geralmente são independentes e identicamente distribuídas, neste caso :

Amostra da Variável aleatória  $X_n$ , onde  $X_n = (x_1, x_2, x_3, \dots, x_n)$ , com  $n$  termos *iid*.

$$L\{X_n|\theta\} = L\{X = x_1|\theta\}L\{X = x_2|\theta\} \dots L\{X = x_n|\theta\}$$

$$L\{X_n|\theta\} = \prod_{i=1}^n L\{X = x_i|\theta\}$$

Para  $f(x, \theta)$ , sendo  $\theta$  o parâmetro desconhecido. Esta deve ser interpretada como uma função de  $\theta$  no qual o estimador de máxima verossimilhança de  $\theta$  é o valor que maximiza  $L(\theta; x_1, \dots, x_n)$ .

Com o objetivo de facilitar a execução dos cálculos, neste trabalho a técnica usada será uma transformação da verossimilhança para a Log-verossimilhança. Com a aplicação do logaritmo da verossimilhança transforma-se os produtos dos termos em somatórias de cálculo mais simples. O método não afeta a condução da análise com visto nas equações a seguir:

Amostra da Variável aleatória  $X_n$ , onde  $X_n = (x_1, x_2, x_3, \dots, x_n)$ , com  $n$  termos *independentes e identicamente distribuídos*.

$$l = \log L\{X_n|\theta\} = l\{X_n|\theta\}$$

$$l\{X_n|\theta\} = \log\{X = x_1|\theta\} + \log\{X = x_2|\theta\} + \dots + \log\{X = x_n|\theta\}$$

$$l\{X_n|\theta\} = \log \sum_{i=1}^n \{X = x_i|\theta\}$$

Para  $f(x, \theta)$ , sendo  $\theta$  o parâmetro desconhecido. Esta deve ser interpretada como uma função de  $\theta$  onde o estimador de máxima verossimilhança de  $\theta$  é o valor que maximiza  $L(\theta; x_1, \dots, x_n)$ .

### 2.3 Análise dos ratings e probabilidades : Descritiva dos dados

A verificação dos *ratings* pode ser feita quando se compara os valores obtidos no primeiro turno, considerando que caso o campeonato terminasse neste ponto, quais os times que se classificariam para a Libertadores, os times que seriam rebaixados e também o vencedor da competição obviamente com o maior *rating* dentre os vinte times. O mesmo pode ser feito para o segundo turno, porém com maior precisão uma vez que considera-se a pontuação da rodada anterior como ponto de partida, e também o mando de campo.

As esperanças dos resultados obtidos com base na estimativa de Máxima Verossimilhança podem ser comparadas com seu valor observado a fim de verificar a qualidade do modelo obtido (Gameran e Souza, 2004). A partir da obtenção dos resultados do processo de estima-

ção dos de  $\gamma_i$  e  $\delta_i$  de cada time, pode-se verificar as probabilidades associadas, de empate derrota e vitória, bem como as chances de placares, esse resultados podem ser verificados atestando a qualidade do modelo e da distribuição adotada para a variável aleatória em estudo. Estes testes de ajuste serão condizidos neste trabalho como testes de bondade do modelo, com exceção dos testes para ajuste de distribuição que considerou-se suficiente as informações disponíveis na literatura que atestam acerca do ajuste adequado deste tipo de variável aleatória.

### 3 MATERIAL E MÉTODOS

#### 3.1 Material

O foco deste trabalho é a estimação dos parâmetros para a condução da análise de resultados (gols pró e gols contra o time mandante) de partidas de futebol e definição do *rating* dos clubes e o mando de campo, o que permite calcular probabilidades para placares ou resultados como vitória, empate e derrota.

Para tal usou-se dados referentes ao campeonato Brasileiro de 2017, a competição que conta com 20 times disputando 38 rodadas divididas em 19 rodadas por turno com 10 jogos por rodada, totalizando 190 jogos somente no primeiro turno e ao final do campeonato serão 380 jogos realizados.

Os dados foram obtidos no site Uol esporte, onde foram coletados os números de gols marcados por cada time na condição de mandante para cada jogo, e o mesmo para cada time em cada partida na condição de visitante. Como comentado anteriormente neste estudo, o histórico dos campeonatos anteriores é propositalmente ignorado, uma vez que considera-se que os *ratings* de anos anteriores terão pouco ou nenhum efeito sobre a competição atual, já que a conformação dos times variam grandemente

Na Tabela 3.1 tem-se a lista de times participantes na série A do campeonato brasileiro de 2017 seguidos de sua sigla, assim serão identificados os times adiante neste trabalho.

Tabela 3.1 – Relação de times que participaram da primeira divisão do campeonato Brasileiro de 2017

Nome oficial da agremiação	Sigla	Nome oficial da agremiação	Sigla
Associação Atlética Ponte Preta	PON	Cruzeiro Esporte Clube	CRU
Associação Chapecoense de Futebol	CHA	Esporte Clube Bahia	BAH
Atlético Clube Goianiense	ACG	Esporte Clube Vitória	VIT
Avai Futebol Clube	AVA	Fluminense Football Club	FLU
Botafogo de Futebol e Regatas	BOT	Grêmio Foot-Ball Porto Alegrense	GRE
Club de Regatas Vasco da Gama	VAS	Santos Futebol Clube	SAN
Clube Atlético Mineiro	CAM	São Paulo Futebol Clube	SPA
Clube Atlético Paranaense	CAP	Sociedade Esportiva Palmeiras	PAL
Clube de Regatas do Flamengo	FLA	Sport Club Corinthians Paulista	COR
Coritiba Foot Ball Club	CTB	Sport Club do Recife	SPT

a variável aleatória analisada foi o número de gols pró e gols contra o mandante. O delineamento consiste em um torneio competido (*round robin*) duplo com os times jogando entre si duas vezes, em um turno como mandante e em outro como visitante. do ponto de vista do delineamento trata-se de um bloco incompleto balanceado resolvido, com blocos de tamanho 2 (um jogo com 2 times) e superblocos de tamanho 20 (uma rodada com 10 jogos). Como o

delineamento respeita o mando de campo pode-se dizer que é um delineamento em blocos do tipo linha e coluna. Para fins de modelagem será ignorado o potencial efeito aleatório de jogo, o que simplifica sobremaneira a análise.

Durante o campeonato os times são divididos na tabela de classificação de acordo com a pontuação obtida, esta divisão foi feita em três regiões denominadas: grupo dos quatro melhor posicionados (G4), zona de rebaixamento, onde os quatro menos pontuados se agrupam (Z4), e grupo dos que não se posicionam em nenhum dos dois anteriores, permanecem na região média da tabela de classificação (RM). Essa nomenclatura também será adotada neste trabalho. Note que de ano para ano varia o número de clubes classificados para a copa Libertadores da América, mas não levamos isto em conta na apresentação.

### 3.2 Métodos

Para este trabalho o software R foi utilizado para toda a fase de programação, equações de estimação e inferência estatística. O *script* desenvolvido será disponibilizado virtualmente.

Usou-se valores nulos para dar início cada um dos parâmetros durante o primeiro turno estes recalculados com o desempenho de cada time ao final desta fase da competição, substituindo os valores iniciais por estes recalibrados, enquanto no segundo turno, os valores iniciais foram o valores médios finais dos *ratings* observados no final da primeiro turno para cada time, assim ao fim da competição os parâmetros médios foram estimados e estes representam o desempenho geral do time no Brasileirão de 2017 em critérios de ataque e defesa.

O processo de estimação dos parâmetros do modelos seguindo uma distribuição de Poisson, consiste na aplicação das seguintes derivadas parciais nas equações de verossimilhança para cada um dos parâmetros como mostrado a seguir:

$$\frac{dL(\gamma|x_i)}{d\gamma} = 0 \rightarrow \hat{\gamma}_i$$

$$\frac{dL(\delta|x_j)}{d\delta} = 0 \rightarrow \hat{\delta}_j$$

$$\frac{dL(\mu|x_{ij})}{d\mu} = 0 \rightarrow \hat{\mu}_{ij}$$

Adotaremos os parâmetros individuais  $\gamma_i$ ,  $\delta_i$  e o parâmetro coletivo  $\mu$  para representar o poder de ataque, força defensiva e mando de campo, respectivamente, definidos na tabela 3.2:



Tabela 3.2 – Representação dos parâmetros

Parâmetro	Valor
$\gamma_i$	representa a força de ataque do time $i$
$\delta_j$	representa a força de defesa do time $j$
$\mu$	representa para efeito comum de mando de campo

Os parâmetros estimados são atribuídos a um vetor  $\hat{\theta}$ , para o cálculo das equações e definição do preditor linear:

$$(\hat{\eta}) : \hat{\theta} = \begin{bmatrix} \hat{\gamma} \\ \hat{\delta} \\ \hat{\mu} \end{bmatrix}$$

$$Y_m \sim \text{Poisson}(\lambda_m = \eta_m = g(\lambda_m) = (\gamma_m - \delta_v + \mu))$$

$$Y_v \sim \text{Poisson}(\lambda_v = \eta_v = g(\lambda_v) = (\gamma_m - \delta_v))$$

$$g(.) = \text{log}(.)$$

Em que  $g(.) = \text{log}(.)$  é o logaritmo natural

No exemplo, o índice  $m$  denota o mando de campo e o índice  $v$  visitante,  $y_m$  e  $y_v$  são os gols pró e contra em um determinado jogo, ou seja, o preditor linear baseado nos *ratings* equivale a uma função  $g(.)$  da esperança condicional dos gols pró e gols contra em cada jogo.

Como considerou-se adequada a distribuição de Poisson, o temos que  $\lambda = e^{\eta_i}$ ; em que

$$\eta = \begin{bmatrix} \eta_m \\ \eta_v \end{bmatrix}$$

é o preditor linear que generaliza as equações anteriores para todos os resultados do turno.

Obtido os valores dos parâmetros  $\hat{\gamma}$  e  $\hat{\delta}$  foi conduzido o estudo da correlação dos parâmetros de ataque e defesa com gols pro e gols respectivamente. As correlações foram testadas segundo o teste *t de Student* para o coeficiente de correlação e interpretadas.

Para que os resultados fossem mais facilmente interpretados organizou-se os *ratings* do segundo turno em função dos grupos e zonas de classificação (G4, RM e Z4), para cada time em cada uma das 19 rodadas, fazendo uma linha de suavização para cada um. Para isso uso-se a função `ggplot()` do pacote `ggplot2`. Ainda de modo didático definiu-se as regiões de

confiança bivariada para os parâmetros  $\gamma$  e  $\delta$  simultaneamente, com o pacote *mixtools*. Os diâmetros das elipses foram reduzidos na ordem de 20 para 1 afim de evitar o confundimento com as áreas, uma vez que são muitos times envolvidos e as elipses tiveram grande raio devido a amplitude das variâncias. Embora isto reduza o poder de interpretação destas elipses, dado que a confiança abrangida na região é baixa, o recurso pode permitir visualizar as estimativas do clube no conjunto ataque-defesa e não apenas com um parâmetro individual.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Primeiro Turno

Nesta seção serão apresentados os resultados dos processos de estimação para os parâmetros de defesa e ataque de cada time ao final do primeiro e segundo turno. Esta divisão teve propósito de melhor exibir e diferenciar o desempenho dos times, considerando o mando de campo apenas na segunda etapa da competição. Os resultados foram divididos em *ratings* de ataque e defesa, onde apresenta-se para cada time da primeira divisão os valores de seus parâmetros.

De modo geral os *ratings* de ataque dos times exibidos na tabela 4.1, se organizaram numa ordem semelhante ao observado no campeonato e sua classificação no primeiro turno, ou seja, times com *ratings* de ataque mais altos comparativamente obtiveram resultados melhores no campeonato do que times que obtiveram *ratings* inferiores. Os resultados dos processos de estimação podem ser observados na tabela 4.1, nesta tabela tem-se a E.g (Estimativa do parâmetro de ataque ( $\hat{\gamma}$ ), o  $S\hat{\gamma}$  (erro de estimativa parâmetro), o valor calculado seguindo uma distribuição Normal padronizada (valor - z) e a probabilidade de significância de  $\gamma$  num teste informal (valor- p(%)):

Tabela 4.1 – Estimativa dos *ratings* relativos a ataque para os times do primeiro turno do Brasileirão de 2017.

Ordem	Clube	$\hat{\gamma}_i$	$s_{\hat{\gamma}_i}$	$z_c$	$P(z < z_c)(\%)$
1	COR	0,6523	0,5041	0,42	67,4
2	GRE	0,2776	0,5075	-0,32	74,9
3	SAN	1,0296	0,4061	1,46	14,4
4	PAL	0,1823	0,5784	-0,44	65,9
5	FLA	0,4055	0,5164	-0,06	95,2
6	SPT	0,2776	0,5075	-0,32	74,9
7	CRU	0,7129	0,4499	0,61	54,1
8	CAP	0,3646	0,5627	-0,13	89,6
9	CTB	-0,0408	0,5401	-0,89	37,3
10	FLU	0,9708	0,4606	1,16	24,6
11	BOT	0,1823	0,5773	-0,44	65,9
12	VAS	0,6931	0,4773	0,53	59,6
13	BAH	0,1823	0,4935	-0,52	60,3
14	CAM	0,3646	0,5411	-0,14	88,8
15	PON	0,5878	0,4472	0,33	74,1
16	CHA	0,2624	0,4696	-0,38	70,3
17	SPA	0,4447	0,5570	0,01	99,2
18	VIT	0,5878	0,2357	0,63	52,8
19	AVA	0,0488	0,5563	-0,7	48,3
20	ACG	0,5878	0,4472	0,33	74,1

Os times Corinthians, Grêmio, e Santos foram estimados com altos valores para *ratings* de ataque, fato que contribuiu para que estes se mantivessem em boas colocações na primeira fase, sendo Corinthians o líder da competição neste turno. Bons *ratings* de ataque são interessantes pois o ideal num campeonato de longa duração são posições de equilíbrio, próximos da zona de classificação e longe da zona de rebaixamento, e um time forte ofensivamente, tem mais chances de pontuar no campeonato.

Os erros de estimativa foram altos para o parâmetro de ataque pois como no primeiro turno, cada time teve apenas um confronto com os demais, a estimativa foi definida com uma repetição apenas, porém ainda assim observou-se poucos valores inesperados quando consideramos a probabilidade associada a estimação num teste de hipótese informal.

A tabela 4.2 foi construída com os *ratings* de defesa do segundo turno e como foi observado nos valores relativos aos parâmetros de ataque, os parâmetros de defesa são importantes na classificação dos times e da posição que ocupam na tabela, esta relação foi observada para times com altos valores para poder defensivo, obtendo boas colocações na competição ao fim do primeiro turno.

Tabela 4.2 – Estimativa dos *ratings* relativos a defesa para os times do primeiro turno do Brasileirão de 2017.

Ordem	Clubes	$\hat{\delta}_i$	$s_{\hat{\delta}_i}$	$z_c$	$P(z < z_c)(\%)$
1	COR	1,3342	0,4298	0,05	96,0
2	GRE	1,4857	0,4616	0,37	71,1
3	SAN	0,4461	0,3974	-2,19	2,80
4	PAL	1,7146	0,4972	0,80	42,3
5	FLA	1,1348	0,4714	-0,38	70,3
6	SPT	1,4857	0,4616	0,37	71,1
7	CRU	0,7880	0,4255	-1,24	21,5
8	CAP	1,9095	0,4534	1,31	19,0
9	CTB	1,8042	0,4972	0,98	32,7
10	FLU	0,7926	0,4095	-1,20	7,41
11	BOT	2,0919	0,4714	1,65	9,89
12	VAS	1,3579	0,4150	0,10	92,0
13	BAH	1,3187	0,4714	0,01	99,2
14	CAM	1,7554	0,4534	0,97	33,2
15	PON	0,7701	0,4346	-1,25	21,1
16	CHA	1,4210	0,4463	0,24	81,0
17	SPA	1,8295	0,4463	1,15	25,0
18	VIT	0,5878	0,2357	-3,08	0,21
19	AVA	1,4914	0,5149	0,34	73,3
20	ACG	0,7701	0,4346	-1,25	21,1

A tabela 4.2 mostra que os times com estimativas maiores para o *rating* de defesa foi mais importante que a de ataque, sendo que a combinação de valores altos tanto para ataque quanto para defesa se encontra nos times das primeiras colocações da tabela, o que foi verificado para o Corinthians, Santos e Grêmio que tem valores mais altos que a média em ambas estimativas. A classificação do campeonato que estes times que comparativamente foram mais eficientes defensivamente mostrou realmente um bom desempenho ao longo da primeira rodada, a exemplo do Corinthians que teve um *rating* de defesa alto, e assim manteve na liderança isolada nesta etapa da competição.

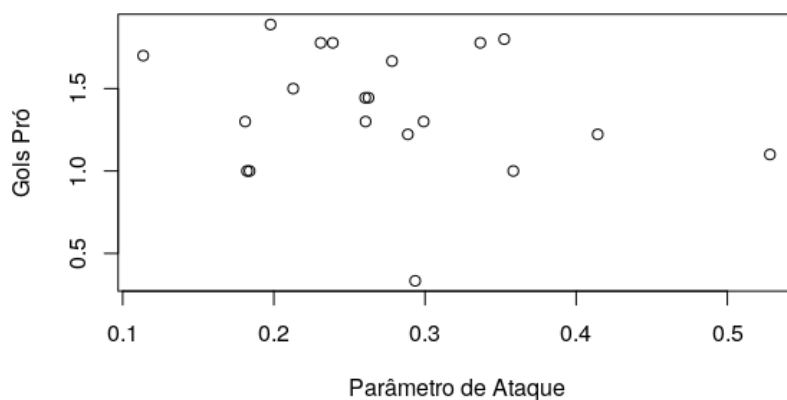
Times como Avaí e São Paulo obtiveram *ratings* altos devido a variações no número de gols, que geraram erros de estimativa altos, confundindo a valores mais precisos associados a estes times, uma vez que no campeonato foi visto que estes times não se destacaram.

Os erros de estimativa foram altos para o parâmetro de ataque e também de defesa pois como no primeiro turno cada time teve apenas um confronto com os demais, a estimativa foi definida com uma repetição apenas, porém ainda assim observou-se poucos valores inesperados quando consideramos a probabilidade associada a estimação.

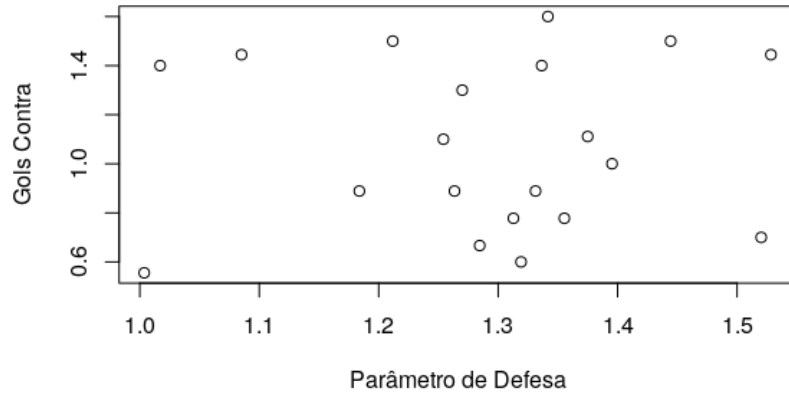
Os valores obtidos para os *ratings* foram testados quando a sua aplicabilidade, de modo prático, intui-se que o poder ofensivo de um time poderia ter uma correlação alta e direta com o número de gols marcados, de mesmo modo, o parâmetro de defesa terá relação inversa com os gols sofridos, assim estas correlações foram estabelecidas para o primeiro turno.

As correlações foram testadas ao nível de 5% de significância para gols pró e força de ataque e gols contra e força defensiva. Como auxílio visual tem-se a seguir os diagramas de dispersão para gols pró e gols contra com relação a  $\gamma$  e  $\delta$  respectivamente, como visto nas figuras 4.1 e 4.2:

Figura 4.1 – Diagrama de dispersão para gols pró e  $\hat{\gamma}_i$



No diagrama de dispersão da figura 4.1 pode-se notar uma concentração dos dados a direita superior, bem como uma tendência à associação positiva, porém as conclusões foram baseadas num teste adequado de t de Student.

Figura 4.2 – Diagrama de dispersão para gols contra e  $\hat{\delta}_i$ 

A figura 4.2 mostra que os pontos apesar de muito dispersos se concentraram no primeiro quadrante do gráfico, mostrando que os valores foram de modo geral baixos para este parâmetro.

Tabela 4.3 – Valores de Correlação entre os Parâmetros  $\hat{\gamma}$  e  $\hat{\delta}$  com Gols Pró e Gols contra no Primeiro Turno

	$\gamma$	$\delta$
<b>Gols Pró</b>	-0,21	0,05
<b>Gols Contra</b>	-0,06	0,03

Os valores não foram significativos no nível de 5%

O teste *t de Student* os valores para a correlação entre gols pró e o *rating* de ataque não foram significativos, bem como os valores de correlação do *rating* de defesa para os gols contra. Também foram testadas as correlações entre o parâmetro de ataque e os gols sofridos, e entre o parâmetro de defesa e os gols marcados, como esperado, estas últimas correlações foram baixas mesmo não significativas.

As estimativas  $\hat{\gamma}$  não obtiveram correlação direta com os gols pró, bem como  $\hat{\delta}$  não pôde ter sua correlação inversa com gols pró verificada, com era esperado inicialmente. Esses resultados podem ter sido observados devido ao alto valor residual das estimações conduzidas, possivelmente devido a subparametrização do modelo escolhido.

## 4.2 Segundo Turno

Os times participantes da primeira divisão do Campeonato Brasileiro de 2017 demonstraram, no segundo turno, bastante variação em relação a primeira etapa da competição como

será visto nas análises deste trabalho. Atribuiu-se esta variação a duração do campeonato, que tem um ano de jogos, a mudança nos estádios onde foram disputados os jogos, pois times que foram mandantes no primeiro turno, foram visitantes no segundo, e este fator interfere no desempenho médio dos times de maneira significativa.

Os *ratings* calculados no primeiro turno foram obtidos a partir de valores nulos e calibrados de acordo com o desempenho médio geral dos times considerando todas as 19 rodadas do campeonato, gerando 20 resultados referentes a força relativa de ataque, 20 valores da força defensiva dos times e um parâmetro fixo que expressa o efeito do mando de campo. Sobre os *ratings* dos times, foi observado de modo geral que valores extremos positivos e negativos tem pouca importância sobre o placar final, que implica diretamente na posição que o time ocupa durante a competição. Times com valores equilibrados para ambos quesitos, garantem melhor desempenho geral, e conseqüentemente posições mais altas a tabela de classificação.

Os Resultados apresentados são valores médios calculados ao fim do campeonato de 2017, obtidos ao final de cada rodada e assim usados para recalibração da estimação posterior. Os *ratings* calculados para cada rodada do segundo turno estão disponíveis no anexo c deste trabalho.

Ao final deste processo de 19 rodadas, obteve-se os valores médios dos parâmetros de ataque e de defesa para estes 20 times da primeira divisão do campeonato Brasileiro de 2017, como é mostrado na tabela 4.4:



Tabela 4.4 – Estimativa dos *ratings* relativos a ataque para os times do segundo turno do Brasileirão de 2017.

Ordem	Clube	$\hat{\gamma}_i$	$s_{\hat{\gamma}_i}$	$z_c$	$P(z < z_c)(\%)$
1	COR	0,1279	0,3303	0,18	85,72
2	PAL	-0,1896	0,3721	-0,69	49,02
3	SAN	0,2870	0,3262	0,67	50,29
4	GRE	-0,043	0,3537	-0,31	75,66
5	CRU	0,1241	0,3363	0,17	86,5
6	FLA	0,1241	0,3331	0,17	86,5
7	VAS	0,2572	0,3316	0,57	56,87
8	CHA	-0,0205	0,3459	-0,25	80,26
9	CAM	0,0565	0,3558	-0,03	97,61
10	BOT	0,0015	0,384	-0,17	86,5
11	CAP	0,1982	0,392	0,33	74,14
12	BAH	-0,0895	0,353	-0,45	65,27
13	SPA	-0,0895	0,4069	-0,39	69,65
14	FLU	0,4495	0,2941	1,3	19,36
15	SPT	-0,1039	0,3349	-0,51	61,01
16	VIT	0,3716	0,1857	1,64	10,1
17	CTB	-0,1039	0,3324	-0,52	60,31
18	AVA	-0,0895	0,3497	-0,45	65,27
19	PON	-0,0430	0,3466	-0,32	74,9
20	ACG	0,1279	0,3378	0,18	85,72

Como visto na tabela 4.4, Os times de modo geral obtiveram valores de  $\hat{\gamma}$  baixos, demonstrando que os times foram menos efetivos ofensivamente, de modo que times de *rating* relativamente baixo como Vitória e Fluminense conseguiram se destacar nessa parâmetro, o que foi observado na competição pois estes times que não tiveram um bom desempenho na competição, não caíram na zona de desclassificação (Z4).

A instabilidade de desempenho não dependente apenas dos parâmetros considerados no modelo, porém para o processo de estimação é interessante que se faça esse paralelo com a competição, haja vista que no fim do Brasileirão times que foram mais uniformes em ambos parâmetros acima da média estiveram próximos a zona de rebaixamento.

Para o parâmetro de defesa, pode-se verificar na tabela 4.5 como foi o desempenho médio dos times nesta etapa da competição:

Tabela 4.5 – Estimativa dos *ratings* relativos a defesa para os times do segundo turno do Brasileirão de 2017.

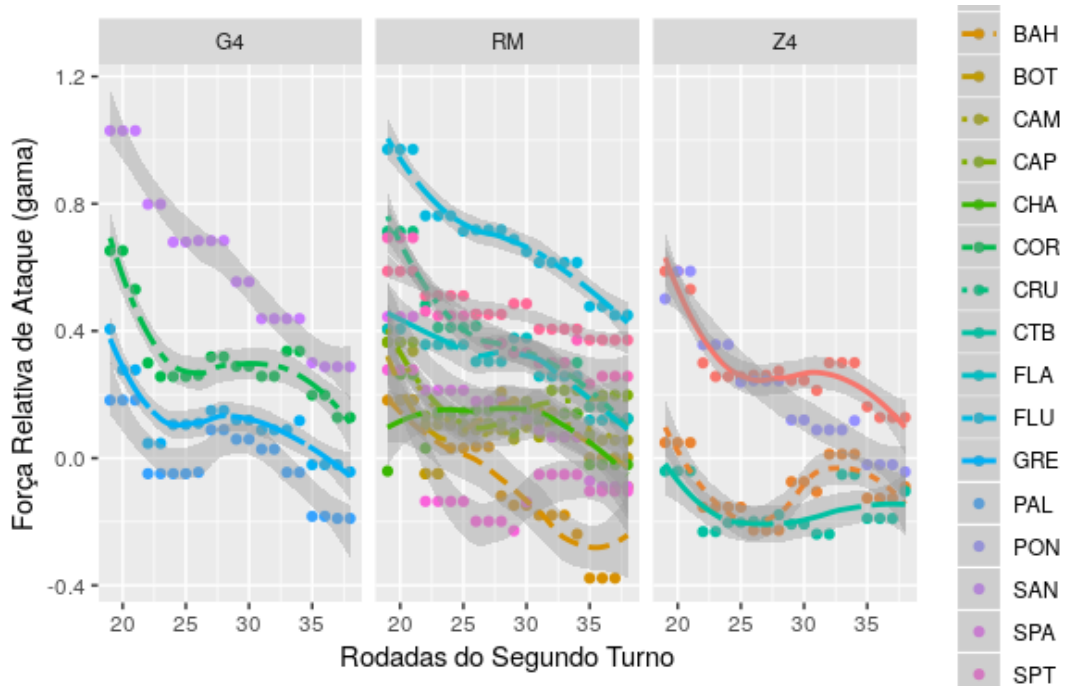
Ordem	Clube	$\hat{\delta}_i$	$s_{\hat{\delta}_i}$	$z_c$	$P(z < z_c)(\%)$
1	COR	0,9380	0,3098	-0,31	75,6
2	PAL	1,3614	0,3378	0,97	33,2
3	SAN	0,8277	0,3015	-0,68	49,6
4	GRE	1,1036	0,3271	0,22	82,5
5	CRU	0,8364	0,3162	-0,62	53,5
6	FLA	0,7899	0,3162	-0,77	44,1
7	VAS	0,9088	0,3030	-0,41	68,1
8	CHA	1,1865	0,3186	0,48	63,1
9	CAM	1,2814	0,3139	0,79	42,9
10	BOT	1,4386	0,3240	1,25	21,1
11	CAP	1,4090	0,3118	1,20	23,0
12	BAH	1,0988	0,3303	0,20	84,1
13	SPA	1,6966	0,3303	2,01	4,44
14	FLU	0,2936	0,2941	-2,51	1,21
15	SPT	1,0362	0,324	0,01	99,2
16	VIT	0,3716	0,1857	-3,56	0,04
17	CTB	0,9954	0,3240	-0,12	90,4
18	AVA	1,0500	0,3303	0,05	96,0
19	PON	1,0035	0,3271	-0,09	92,8
20	ACG	1,0380	0,3098	0,02	98,4

Quando considerou-se os parâmetros de defesa na tabela 4.5 times como Vitória e Fluminense tiveram estimativas altas porém acompanhadas de altos erros, o que deixou mais claro a diferença na força defensiva observada no campeonato pois foram times que tiveram desempenho ruim considerando a pontuação da tabela de classificação do Brasileirão.

Corinthians, Palmeiras e Grêmio que tiveram boas estimativas de *ratings* de defesa, com relação aos demais, e seus resultados são precisos uma vez que a probabilidade no teste informal de hipótese (p.d) foi significativa, enquanto também foi visto que no campeonato de 2017 tiveram boas colocações na tabela. São Paulo obteve o valor mais alto para  $\hat{\delta}$ , o que demonstra o desempenho do time na segunda fase do Brasileirão, pois no fim do primeiro turno estava na zona de rebaixamento, e terminou a competição na região média da tabela (RM), sem almejar título ou classificação porém, longe do risco de rebaixamento para a série B.

A seguir a figura 4.3 que exhibe os valores dos parâmetros de ataque ao longo do segundo turno, divididos em zona de classificação para a libertadores (G4), região média (RM) e zona de rebaixamento (Z4), nesta observa-se a disposição dos valores obtidos para os parâmetros de ataque para cada time ao longo da segunda etapa da competição separados pela cor e tipo de traço que conecta os pontos :

Figura 4.3 – Estimativas para os parâmetros de ataque ( $\hat{\gamma}_i$ ) produzidas a cada rodada do segundo turno para todos os times do Campeonato Brasileiro série A de 2017



Na figura 4.4 é possível identificar os times que estão na zona de classificação (G4) os times que estão na zona de rebaixamento da competição (Z4), e também os times em posições intermediárias (RM) bem como o comportamento do seus *ratings* de ataque ao longo do segundo turno da competição.

As bandas de confiança para cada time forma estimadas pelo pacote *ggplot* pelo método de regressão local (Loess). A partir destas curvas pode-se observar que as estimativas dos times da região média (RM) não se diferenciaram em sua maioria dos times do G4, a exemplo do Sport, Botafogo e Atlético Mineiro, cujas bandas estavam acima do grupo de classificação.

Figura 4.4 – Estimativas para os parâmetros de defesa ( $\hat{\delta}_i$ ) produzidas a cada rodada do segundo turno para todos os times do Campeonato Brasileiro série A de 2017

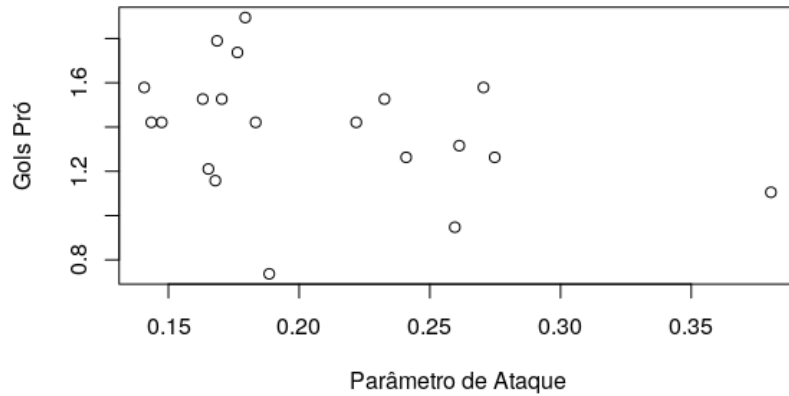


Na figura acima observou-se uma tendência geral de que os valores dos *ratings* de ataque decaírem, ou seja, os times ao final da competição se comportam de maneira menos ofensiva quando comparados ao final da primeira etapa do campeonato.

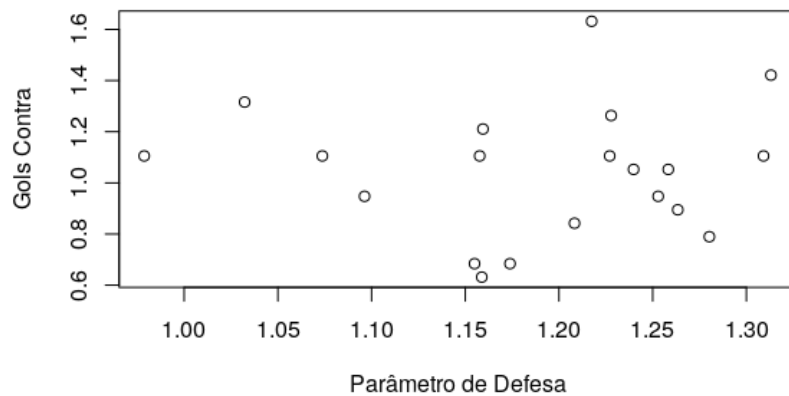
A figura 4.4 apresenta os resultados da estimação dos *ratings* referentes a força defensiva dos times, diferentemente do que é observado na figura 4.3, os pontos mostraram que os times se fecharam mais, fortalecendo a defesa e portanto sofrendo menos gols por jogo, bem como um volume maior de dados em valores positivos do que negativos. Quando comparados ao final do segundo turno os valores para  $\delta$  tenderam a ser mais altos no segundo turno. Os times que figuraram nas primeiras colocações não se distanciaram

Analisando o resultado das regressões locais e as bandas obtidas, viu-se que no critério defensivo os times estão de modo geral semelhantes, com exceção do São paulo que teve os valores mais baixos.

Após a estimação dos parâmetros testou-se as correlações dos parâmetros de defesa e ataque com os gols sofridos e gols marcados em cada partida. Neste momento espera-se que a relação entre os gols pró tenha relação com o parâmetro de ataque, e que o mesmo ocorra na influência do parâmetro de defesa sobre o numero de gols contra. A seguir tem-se os diagramas de dispersão de  $\hat{\gamma}$  e  $\hat{\delta}$  com relação aos gols marcados e gols sofridos:

Figura 4.5 – Diagrama de Dispersão para Gols Pró e  $\hat{\gamma}_i$ 

Na figura 4.6 não se evidencia uma tendência ou concentração dos pontos, esta dispersão da a ideia de que os dados não apresentam associação, porém não exclui a necessidade da aplicação de testes para verificação.

Figura 4.6 – Diagrama de Dispersão para Gols Contra e  $\hat{\delta}_i$ 

A figura 4.7 mostra os pontos referentes as interações entre os valores médios de parâmetros de ataque e os gols marcados, assim como observado na figura 4.6, a nuvem de pontos indica que os valores não se associam, assim procedeu-se com os testes de correlação ao nível de 5% de probabilidade como pode ser observado na tabela 4.6:

Tabela 4.6 – Valores de Correlação entre os Parâmetros  $\hat{\gamma}$  e  $\hat{\delta}$  com gols pró e gols contra, sucessivamente: Segundo Turno

	$\gamma$	$\delta$
<b>Gols Pró</b>	-0.35	0.16
<b>Gols Contra</b>	-0.02	0.01

Os valores não foram significativos no nível de 5%

Como visto na tabela 4.6 os gols marcados nas partidas não se correlacionam significativamente com o parâmetro de ataque, bem como o parâmetro de defesa não foi significativo em relação aos gols contra, porém este resultado, apesar de não esperado, não interfere na precisão do processo de estimação dos *ratings*, uma vez que as técnicas não se basearam no especificamente nos critérios de gols pro e contra e sim no resultado final da partida, considerando o mando de campo neste segundo turno.

Com base nos valores estimados para  $\gamma$  e  $\delta$  no primeiro e segundo turnos da competição, e também os erros de estimação deste parâmetros, foi possível produzir os gráficos de elipses, onde o eixo y recebe os valores das estimativas de defesa e no eixo x tem-se os *ratings* de ataque.

Nas imagens 4.7 e 4.8 foram plotados os valores médios de cada parâmetro para as 38 rodadas competição, portanto serviu de referência para conclusões acerca dos times participantes ao longo do Brasileirão. Os times dispostos nas imagens cuja sigla estão na cor preta, são os times que estão no G4, ou seja, ficaram na liderança da competição ao fim do segundo turno, em vermelho nas posições intermediárias e ao final da tabela nas quatro ultimas colocações as siglas foram separadas na cor verde.

As elipses representam a região de confiança bivariada, porém neste caso não foi uma opção viável representá-las com sua variância real dado o tamanho dos raios, e a quantidade de valores a serem plotados ser elevada (20 times), as circunferências se sobrepuseram, tornando a interpretação difícil e pouco didática. Para tal, procedeu-se com a redução na ordem de 20 para 1 nos valores de raio das elipses, para que estas possam ser apresentadas, como vê-se nas figura a seguir:

Figura 4.7 – Elipses contendo os *ratings* relativos dos times no primeiro turno do Campeonato Brasileiro de 2017

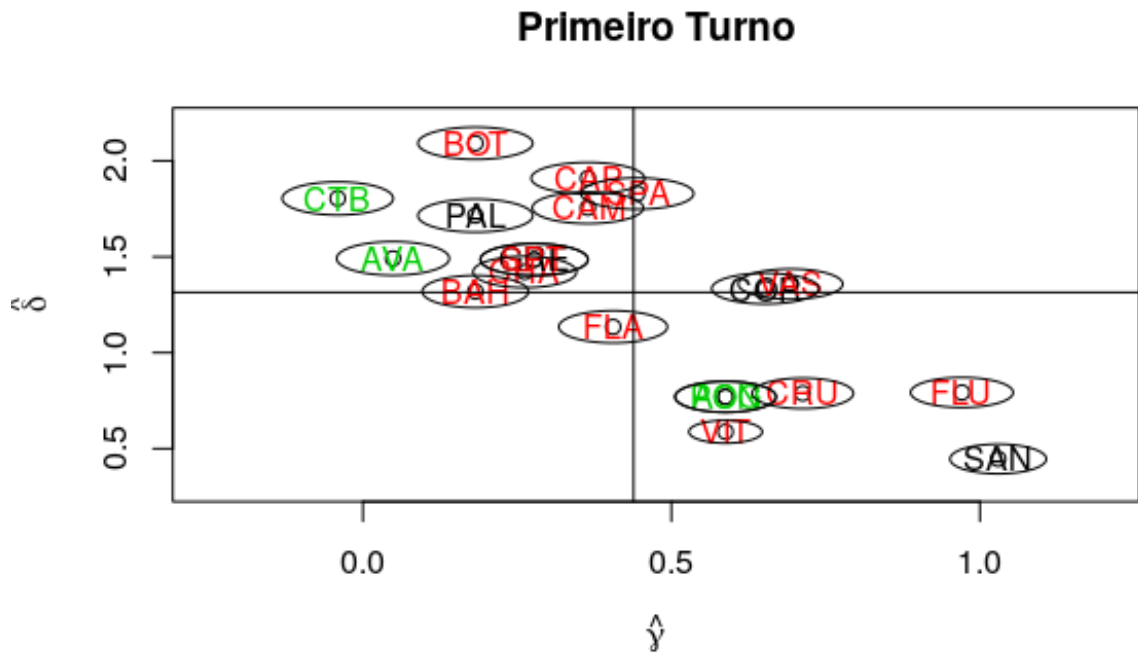
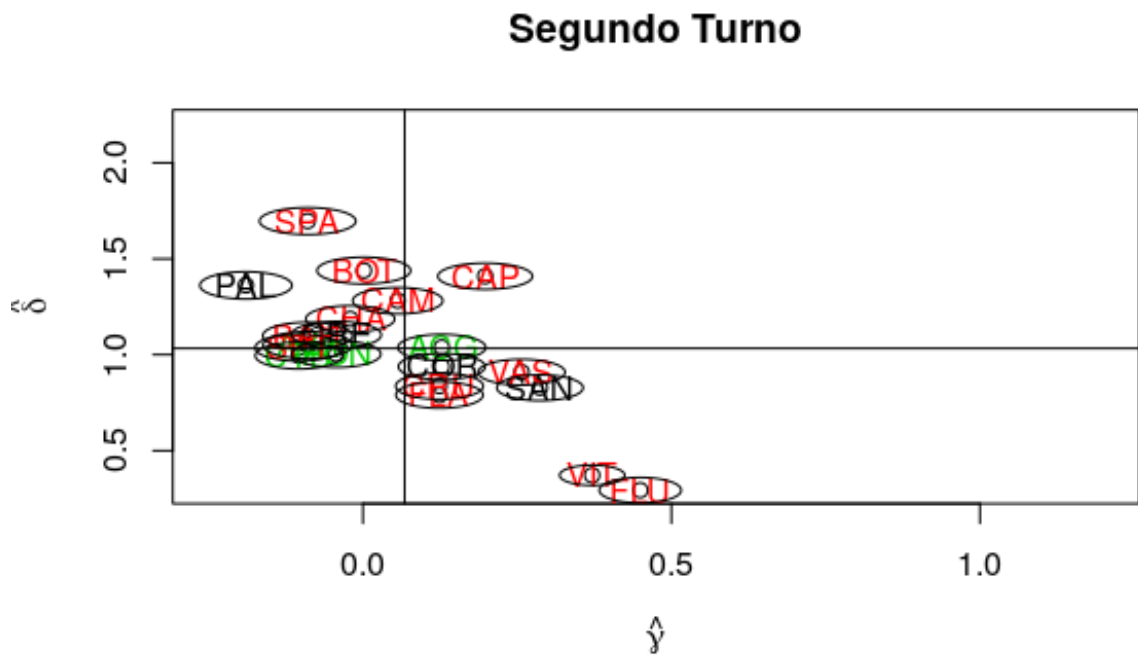


Figura 4.8 – Elipses contendo os *ratings* relativos dos times no segundo turno do Campeonato Brasileiro de 2017



As posições dos times no gráfico mostraram como estes se relacionam com as colocações obtidas na competição pelo sistema de pontos corridos, o time Santos por exemplo tem um alto valor para o parâmetro de defesa ( $\delta$ ), porém um dos mais baixos para o ataque ( $\gamma$ ), mesmo assim manteve em boa posição no primeiro turno, os times Corinthians e Vasco se posicionaram bem com relação as médias de ataque e defesa, sendo que na figura 4.7 o quadrante 1 é o ponto ideal, onde estão os maiores valores para  $\gamma$  e  $\delta$  simultaneamente, similarmente o quadrante 3 concentra os valores mais baixos.

Os times em vermelho, representando os times na região média da tabela de classificação do campeonato permaneceram próximos centro do gráfico, como era de se esperar, com valores medianos para ambos os parâmetros conseguiram manter-se em equilíbrio no primeiro turno, sem risco alto de rebaixamento (Z4) porém sem pontuar o suficiente para acesso a zona de classificação (G4).

Foi observado que os times ficaram mais centralizados quando comparados ao turno anterior, demonstrando menor variação entre os times considerando  $\gamma$  e  $\delta$ , as regiões de confiança bivariada foram reestruturadas na escala de 1 para 20, para efeito didático, e assim pode ser notado que vitória e fluminense se distanciam dos demais, proporcionalmente como valores de defesa bastante inferiores aos demais.

Nenhum time se posicionou exclusivamente no quadrante 3, que configuraria um péssimo desempenho geral, porém tem-se no segundo turno São Paulo isolado no quadrante 4 com *rating* de ataque bem acima da média, o que pode ser um fator que explica o desempenho deste time que deixou a zona de rebaixamento na segunda etapa da competição.

De modo geral os times se movimentaram de maneira uniforme quando comparamos o primeiro e o segundo turno comparando os dois gráficos, percebe-se que os times ficaram mais próximos do centro o que indica que as médias foram semelhantes no segundo turno, com exceção de fluminense e vitória que tiveram variação maior que os demais, tendo obtido valores mais altos de  $\gamma$  e muito inferiores aos demais para  $\delta$ . Em ambos turnos nota-se uma forte tendência diagonal na disposição dos times no gráfico demonstrando uma dispersão com relação às médias de ataque e defesa, ou seja, os times são essencialmente mais efetivos no ataque ou na defesa, principalmente no primeiro turno, no segundo, apesar de ainda presente essa tendência é atenuada, e os times tem a amplitude *rating* entre si reduzida, se concentrando mais ao centro do gráfico.



### 4.3 Probabilidades dos placares e resultados

Após os processos de estimação dos parâmetros, foi possível obter os *ratings* de cada time de acordo com as forças relativas de ataque e defesa, considerando o mando de campo.

Estes resultados podem ser quantificados quanto as chances de ocorrências de vitória, empate e derrota. Assim considerando os *ratings* médios finais do segundo turno como representativos da condição atual do time até que ocorra uma nova competição, foram calculadas regressivamente as probabilidades dos eventos vitória, empate ou derrota para cada jogo ocorrido nos 190 confrontos do segundo turno.

As probabilidades dependem essencialmente de  $\gamma$  e  $\delta$  de cada time analisado no jogo, porém com será mostrado a seguir, o fator mando de campo foi especialmente influente em jogos onde os times possuem *ratings* muito baixos.

Inicialmente havia sido considerado que o mando de campo teria um efeito positivo no modelo, entretanto a estimativa encontrada para as tres primeiras rodadas do segundo turno foi negativa ( $\hat{\mu} = -0.18$ ) como pode ser visto na seção anexos deste trabalho que traz os valores de  $\mu$  para cada rodada do segundo turno. A estimativa, mesmo tendo sido estimada com um valor baixo, já foi indício de que os times de modo geral tiveram um fraco desempenho quando jogaram em casa nesta etapa, influenciando fortemente as estimativas as probabilidades associadas a empate vitória e derrota (vide tabela 4.7).

Ao final da segunda fase da competição o valor da estimativa do parâmetro  $\mu = 0.24$ , esse valor muda expressivamente os valores de *rating* e por consequência as probabilidades associadas, a tabela 4.7 traz exemplos desta variação, nela estipulou-se o efeito do mando de campo sobre dois times ideais, sendo time "A"mandante e time "B"visitante, com *ratings* de ataque e defesa sendo igual a um ( $\gamma_A = \gamma_B = \delta_A = \delta_B = 1$ ) e três valores do mando de campo: o inicialmente definido como ( $\mu = -0.18$ ), a estimativa obtida ao final do campeonato e um terceiro mais expressivo ( $\mu = 0.24$  e  $\mu = 0.5$ ).

Na tabela 4.7 pode-se ver a variação das probabilidades associadas a derrota empate e vitória alterando apenas o mando de campo de modo significativo:

Tabela 4.7 – Probabilidades de vitória, empate e derrota para o time mandante.

$\mu$	Probabilidades (%)		
	Vitória	Empate	Derrota
-0,18	29,3	32,4	38,3
0,24	42,5	28,2	29,3
0,50	52,2	24,5	23,3
0	34,6	30,8	34,6

Outra abordagem seguindo a mesma metodologia consiste em verificar as probabilidades de ocorrência de gols do mandante e gols do visitante, dessa forma apresenta-se a os valores ainda considerando que todos os parâmetros são iguais a 1, e o fator mando de campo será 0.24 conforme é mostrado na tabela 4.8 :

Tabela 4.8 – Probabilidades para cada possível combinação de gols marcados em uma partida, considerando  $\gamma_A = \gamma_B = \delta_A = \delta_B = 1$  e  $\mu = -0.18$ . Os valores na matriz triangular superior representam vitórias do time mandante e na matriz triangular inferior, do visitante.

(%)	0	1	2	3	4	>4
0	10	10	05	02	01	00
1	13	13	07	02	01	00
2	08	08	05	01	01	00
3	02	02	01	00	00	00
4	00	00	00	00	00	00
>4	00	00	00	00	00	00

Pode-se perceber que ao nivelar os parâmetros as probabilidades dos placares com número de gols mais baixos aumentam, tendendo a ser mais altos para o mandante, uma vez que o único parâmetro que causou variação neste caso foi o mando de campo ( $\mu$ ).

Nos exemplos tem-se as probabilidades totais dos placares, estes limitados a 4 gols ou mais para cada time, pois nesta classe as probabilidades já começam a convergir em zeros.

De posse do resultado das estimações, pode-se assim calcular a probabilidade de vitória, empate ou derrota (PV, PE ou PD) para qualquer confronto de times participantes da competição.

Como forma de exemplificação foi montada a tabela 4.9, na qual se faz uso de *ratings* reais de times que tiveram seus parâmetros estimados ao fim do campeonato. Para ilustrar tomou-se o embate entre Corinthians e São Paulo, sendo Corinthians o time mandante :

Tabela 4.9 – Probabilidades para cada possível placar no jogo Corinthians vs São Paulo. Estimativas associadas:  $\hat{\gamma}_{COR} = 0.1279$ ,  $\hat{\gamma}_{SPA} = -0.0895$ ,  $\hat{\delta}_{COR} = 0.9380$ ,  $\hat{\delta}_{SPA} = 1.6966$  e  $\mu = 0.24$ . A matriz triangular superior representa vitórias do mandante e a inferior do visitante. Na diagonal resultados de empate.

(%)	0	1	2	3	4	>4
0	50	21	05	01	00	00
1	13	06	01	00	00	00
2	02	01	00	00	00	00
3	00	00	00	00	00	00
4	00	00	00	00	00	00
>4	00	00	00	00	00	00

Como é mostrado na tabela 4.9 a probabilidade de resultados com muitos gols para qualquer um dos times é baixa, uma vez que ambos parâmetros não foram muito distintos, neste ponto o Corinthians tem *ratings* inferiores ao do São Paulo, mas com o efeito de mando de campo a favor pode se manter em equilíbrio para este confronto.

As probabilidades de Vitória, Empate e Derrota para Corinthians como mandante em um confronto com o São Paulo vê-se na tabela 4.10:

Tabela 4.10 – Probabilidades Vitória, Empate e Derrota para o mandante no jogo Corinthians vs São Paulo. Estimativas associadas:  $\hat{\gamma}_{COR} = 0.9379$ ,  $\hat{\gamma}_{SPA} = 1.6366$ ,  $\hat{\delta}_{COR} = 0.1241$ ,  $\hat{\delta}_{SPA} = -0.0894$  e  $\mu = 0.24$ .

$\mu$	PV	PE	PD
0.24	27.0%	55.6%	16.4%

As probabilidades de vitória, derrota e empate para o mandante Corinthians expostas na tabela 4.9 mostraram que os resultados possíveis da tabela 4.10 estão de acordo com o esperado, pois o resultado mais provável era empate, como de fato foi observado no segundo turno da competição, onde como mandante o São paulo conseguiu um empate contra o Corinthians que até então era tido como o time favorito.

A seguir na tabela 4.11 serão exibidos os resultados referentes ao jogo Palmeiras e Fluminense, a escolha deste confronto visa verificar como as probabilidades se distribuem quando *ratings* muito diferentes são confrontados. O time Palmeiras obteve no segundo turno da competição um dos melhores desempenhos, enquanto Fluminense tem baixos valores para ambos parâmetros, nesse exemplo pôde se visto a influencia prática do mando de campo quando times de *ratings* muito distintos se confrontam:

Tabela 4.11 – Probabilidades de gols marcados numa partida entre Palmeiras e Fluminense :  $\hat{\gamma}_{PAL} = -0,1896$ ,  $\hat{\gamma}_{FLU} = 0,4495$ ,  $\hat{\delta}_{PAL} = 1,3614$ ,  $\hat{\delta}_{FLU} = 0,2936$  e  $\mu = 0,24$ ; Triângulo superior para visitante e inferior pra mandante

(%)	0	1	2	3	4	>4
0	31	12	02	00	00	00
1	24	10	02	00	00	00
2	09	04	01	00	00	00
3	02	01	00	00	00	00
4	00	00	00	00	00	00
>4	00	00	00	00	00	00

Como é visto na diagonal da tabela, a tendência de empates é alta quando comparadas com as demais. O poder de ataque do Fluminense apesar de inferior ao do Palmeiras não foi suficiente para deslocar completamente a probabilidade para vitória do mandante porém na tabela 4.12 fica mais evidente essa tendencia. As probabilidades de vitória, empate e derrota para PAL como mandante num confronto com o FLU sendo o visitante :

Tabela 4.12 – Probabilidades de vitória, empate e derrota para time Mandante :  $PAL_m \times FLU_v$

$\mu$	PV	PE	PD
0.24	43.0%	40.0%	17.0%

Apesar da probabilidade de empate ainda ser alta, Fluminense não foi vitorioso jogando em casa ou fora contra o Palmeiras, sendo derrotado por 3 a 1 no segundo turno, e 1 a 0 no primeiro turno.

Encontra-se no anexo D neste trabalho, a tabela 6.6, nesta se dispõe juntamente com os confrontos para o segundo turno, as probabilidades associadas a cada um dos jogos e o placar que efetivamente foi observado. Nesta Tabela fica evidente os valores aberrantes e suas probabilidades de ocorrência, quanto mais próximo do valor de referência de ( $\chi^2_{v=2} = 5,99$ ) mais improvável o resultado observado. Quando comparamos o ajuste do modelo, os dados concordam sobre o resultado insperado, A tabela 4.13 tem exemplos deste resultados improváveis, conhecidos como "zebras"na linguagem do futebol, ou seja, resultados inesperados devido a fatores de força do time, desempenho médio no campeonato, ou no caso deste estudo, times que possuem *ratings* superiores sendo vencidos por times com *ratings* menores.

A tabela traz as informações de gols do time mandante(G.M), gols do time visitante(G.V), a identificação de cada time na posição mandante(S.M) e visitantes (S.M), as chances de empate, vitória e derrota para cada confronto (PV, PE, PD) e os valores de significância calculadas em cada confronto ( $\chi^2$ ), assim os valores são dispostos na tabela abaixo:

Tabela 4.13 – alguns resultados surpreendentes do segundo turno e probabilidades associadas. O valor de referência para testar a hipótese de que os dados se adequam ao modelo é  $\chi^2 = 5,99$  para  $v = 2g.l$

Jogo	Rodada	Mandante	Placar	Visitante	Probabilidades(%)			$\chi^2_c$
					Vitória	Empate	Derrota	
225	23	SAN	2 - 0	COR	0,48	0,34	0,18	3,19
342	35	CRU	2 - 2	AVA	0,59	0,25	0,16	4,52
286	29	AVA	1 - 1	BOT	0,63	0,22	0,15	5,09
246	25	ACG	1 - 2	CRU	0,72	0,17	0,10	6,03

Como visto na tabela 4.13 estão dispostas a informações que permitem algumas considerações acerca do processo de estimação, os valores aberrantes para a probabilidade de resultado (E.S) não são refletidas na chance de resultado (E.P), ou seja, os *ratings* de fato representam parâmetros que são importantes para explicar o placar, resultado que era esperado uma vez que os modelos foram ajustados para explicar estes eventos. O que talvez seja inesperado é a ausência de grandes zebras (houve poucos resultados realmente surpreendentes). Os jogos 214, 220 e 375 foram as maiores surpresas, pois apesar de se esperar a vitória do mandante pela alta probabilidade do evento citado (P.V= 42%, 39%, 52% sucessivamente), não se esperava um placar tão elástico. Enquanto para os jogos 307 e 378, a derrota do time mandante era considerada um evento raro, de baixa chance de ocorrência, porém ainda assim, foi observado.

### 4.3.1 Considerações gerais

Os resultados encontrados refletiram de modo satisfatório os dados observados na competição em estudo, pois os times que se destacaram nos parâmetros de defesa e de ataque simultaneamente, como foi o caso do Corinthians no primeiro turno, obtiveram um excelente desempenho no campeonato, sendo este inclusive campeão do Brasileirão de 2017, devido em grande parte estes altos valores observados no parâmetro de defesa no primeiro turno. No segundo turno, os *ratings* não houve alterações suficientes para alterar os destinos do campeonato.

A observação dos resultados obtidos mostrou que o Campeonato Brasileiro apresenta uma grande variação entre primeiro e segundo turno, dado que os valores dos *ratings* foram distintos para os mesmos times quando comparados, o que pode auxiliar e alertar as comissões técnicas no preparo das equipes para que tenham constância em seu desempenho durante a competição.

O processo de estimação via método de máxima verossimilhança gerou bons resultados para a obtenção dos *ratings*, sendo que poucos resultados puderam ser considerados surpreendentes pela discrepância entre estimativas e valores observados.

## 5 CONCLUSÃO

Os parâmetros estimados por máxima verossimilhança para o primeiro turno e após cada rodada do segundo turno permitiram calcular predições para probabilidades de placares e resultados de partidas no segundo turno do Campeonato Brasileiro de Futebol, série A de 2017.

Não foram identificados resultados muito surpreendentes com relação às predições do modelo, o que indica sua boa adequação como instrumento preditivo.

Pretende-se estudar formas de implementação bayesiana dinâmica para recomendar o modelo para utilização prática.

## 5.1 Referências

- <https://esporte.uol.com.br/futebol/campeonatos/brasileirao/jogos/>  
<https://oglobo.globo.com/brasil/brasil-movimenta- apenas-2-do-mercado-da-bola-13026765>  
 <acesso em 06 fev de 2018>
- ARTUSO, A. R. **Análise do aproveitamento dos times no campeonato brasileiro a partir de uma distribuição normal**. Revista- Brasileira de Biometria, São Paulo, v. 25, n. 4, p. 49-63, out./dez. 2007.
- Da Silva, C. & Gomes Moreira, D. **A vantagem em casa no futebol: comparação entre o campeonato Brasileiro e as principais ligas nacionais do Mundo**. Revista Brasileira de Cineantropometria e Desempenho Humano, 2008.
- TR dos Santos. **Um modelo de espaço de estados poisson para a modelagem dos confrontos de futebol entre Brasil e Argentina**. Sigmae, Alfenas, v.2, n.1, p. 42-47. 2013.
- Gamerman, D., Rezende dos Santos, T. and Franco, G. C. **A non-Gaussian family of state-space models with exact marginal likelihood**. J. Time Series Anal. 34 625–645.,2013.
- Gamerman, D., de Souza Junior, O. G. **Previsão de partidas de futebol usando modelos dinâmicos**. XXXVI-SBPO. 2004
- P. HOLGATE. **Estimation for the bivariate Poisson distribution**. Oxford University Press on behalf of Biometrika Trust. Biometrika, Vol. 51, No. 1/2, pp. 241-245. 1964.
- Emonet, B & Kuonen, D. **Revisiting Statistical Applications in Soccer**. Technical report. Department of Mathematics, Swiss Federal Institute of Technology, Lausanne, 2000.
- De Araujo CTP, Tavares L, Alvares LG, Neto FL, Suzuki AK. **Modelagem estatística para a previsão de jogos de futebol: Uma aplicação no campeonato brasileiro de futebol** 2014. Revista da Estatística UFOP, 2015
- Karlis, D. and I. Ntzoufras, **Analysis of sports data by using bivariate poisson models**. Journal of the Royal Statistical Society: Series D (The Statistician) 52(3), 2003.
- Knorr-Held L **Dynamic Rating of Sports Teams**. Journal of the Royal Statistical Society: Series D (The Statistician). July 1, 2000
- ARRUDA, M. L. **Poisson, Bayes, Futebol e DeFinetti**. Dissertação (Mestrado em Estatística). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2000.
- Crowder, M., M. Dixon, A. Ledford, and M. Robinson. **Dynamic Modelling and Prediction of English Football League Matches for Betting**. Journal of the Royal Statistical Society: Series D (The Statistician) 51:157–168. 2002.
- OLIVEIRA C.A. FILHO, AK SUZUKI, F. LOUZADA **Uma abordagem bayesiana para previsão de resultados de jogos de futebol: Uma aplicação ao campeonato Inglês**. Revista Brasileira de Biometria, Lavras, v.35, n.1, p.76-97, 2017.
- DeFINETTI, B. (1972), **Probability, Induction and Statistics**. London: John Wiley.
- Reep, C. and Benjamin, B. **Skill an chance in association football** Journal of Royal Statistical Society A, 191, 581-585, 1968.



Maher, M.J. **Modelling association football scores**. Statistica Neerlandica, 36, 109-118, 1982.

Ridder, G., Cramer, J.S. and Hopstaken, P. **Down to ten: estimating the effect of red card in soccer**. Journal of American Statistical Association, 89, 1124-112, 1994.

## 6 ANEXO

Tabela 6.1 – Tabela de Classificação dos Times no Final do Primeiro Turno do Brasileirão de 2017

Posição	Time	Pontuação
1	Corinthians	47
2	Grêmio	39
3	Santos	35
4	Palmeiras	32
5	Flamengo	29
6	Sport	28
7	Cruzeiro	27
8	Atlético-PR	26
9	Coritiba	25
10	Fluminense	25
11	Botafogo	25
12	Vasco	24
13	Bahia	23
14	Atlético-MG	23
15	Ponte Preta	22
16	Chapecoense	22
17	São Paulo	19
18	Vitória	19
19	Avaí	18
20	Atlético Goianiense	12

Tabela 6.2 – Tabela de Classificação dos Times no Final do Segundo Turno do Brasileirão de 2017

Posição	Time	Pontuação
1	Corinthians	72
2	Palmeiras	63
3	Santos	63
4	Grêmio	62
5	Cruzeiro	57
6	Flamengo	56
7	Vasco	56
8	Chapecoense	54
9	Atlético-MG	54
10	Botafogo	53
11	Atlético-PR	51
12	Bahia	50
13	São Paulo	50
14	Fluminense	47
15	Sport	45
16	Vitória	43
17	Coritiba	43
18	Avaí	43
19	Ponte Preta	39
20	Atlético Goianiense	36