



Trabalho de Conclusão de Curso

**Sorte e Habilidade: a Influência Desses Dois  
Aspectos no Futebol Europeu**

Juliana Sena de Souza

27 de julho de 2020

**Juliana Sena de Souza**

**Sorte e Habilidade: a Influência Desses Dois Aspectos no  
Futebol Europeu**

Trabalho de Conclusão apresentado à comissão de Graduação do Departamento de Estatística da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Bacharel em Estatística.

Orientadora: Profa. Dr<sup>a</sup>. Márcia Helena Barbian

Porto Alegre  
Janeiro de 2020

**Juliana Sena de Souza**

**Sorte e Habilidade: a Influência Desses Dois Aspectos no  
Futebol Europeu**

Este Trabalho foi julgado adequado para obtenção dos créditos da disciplina Trabalho de Conclusão de Curso em Estatística e aprovado em sua forma final pela Orientadora e pela Banca Examinadora.

Orientadora: \_\_\_\_\_  
Profa. Dr<sup>a</sup>. Márcia Helena Barbian, UFRGS  
Doutora pela Universidade Federal de Minas Gerais, Belo Horizonte, MG

Banca Examinadora:

Prof. Dr. Rodrigo Citton Padilha dos Reis, UFRGS  
Doutor pela Universidade Federal de Minas Gerais, Belo Horizonte, MG

Porto Alegre  
Janeiro de 2020

*"The great thing about this life of ours is that you can be someone different to everybody". - Jennifer Niven, All the Bright Places*

# Agradecimentos

À minha mãe, mulher guerreira e meu orgulho. Por ser uma super mãe e por ter me criado da melhor maneira possível.

Aos meus irmãos Valdir e Elton, por sempre falarem de forma animada que eu faço UFRGS.

À toda família Sena, por serem a melhor família do mundo e por sempre me proporcionarem as melhores férias de fim de ano. Em especial aos meus tios Eloete e Anailton, por todo amor e apoio que recebo de vocês; e à Karoliny, pelas madrugadas em claro em que incomodamos todo mundo, pelos filmes ruins vistos, pelas horas no twitter, pelo ifood escondido, por ser minha companhia para olhar as estrelas e principalmente por todas as risadas que damos quando estamos juntas.

À minha orientadora, professora Márcia Barbian, por ter feito parte da minha jornada na graduação desde 2017, sendo minha professora em quatro disciplinas, três monitorias e na orientação deste trabalho. Muito obrigada pelas reuniões semanais, por todo o suporte, motivação, conselhos e ensinamentos.

Ao professor Rodrigo Citton Padilha dos Reis, por não somente ter aceitado fazer parte da banca desse TCC e ter contribuído de forma positiva com correções e sugestões, mas também por ter aceitado fazer parte da próxima etapa da minha vida, o mestrado, como meu orientador.

À todos os professores do departamento de estatística que contribuíram de forma direta ou indireta para a minha formação.

Ao professor Markus Stein, por ter me aguentado todas as semanas incomodando na sala enquanto procurava a Márcia.

À professora Luciana Nunes, por ser uma pessoa incrível e uma inspiração a ser seguida.

À professora Vanessa Leotti, a melhor paraninfa de todas.

Ao João Vitor, que é mais que um melhor amigo, é a minha pessoa. Obrigada por estar presente desde o começo de tudo isso, por sempre me apoiar, me ouvir e por me conhecer como ninguém.

Ao Matheus, por termos nos reencontrado e por termos ficado um na vida do outro. Obrigada principalmente por ter me aguentado nessa reta final de faculdade, por ouvir meus desabaços e reclamações, me acalmar nos meus surtos, por sempre me ajudar e principalmente por estar presente sempre que eu preciso.

À Gabriela e Bruna, que conheci no primeiro ano de graduação e estão comigo até hoje. À Gabriela, por ter me mostrado onde ficava a escada da informática, por ter enfrentado aquela escada infinita e ter ido até lá em cima comigo, foi aí que nossa amizade começou e ainda segue firme e forte. À Bruna, por ser sempre gentil e por indiretamente ter feito eu voltar a gostar de futebol, sem dúvidas esse trabalho de

conclusão de curso se deve muito à você. Obrigada por me aguentarem durante esses cinco anos e por não terem desistido de mim (ainda). Ter o apoio e companheirismo de vocês durante toda a graduação sem dúvidas fez toda a diferença.

À Franciele, por ser parceira pra todas as horas e por sempre aceitar qualquer tipo de rolê. Por ser minha voz nos momentos que minha timidez fala mais alto. Por ser paciente o suficiente para aguentar minhas grosserias desnecessárias. Obrigada por existir e por ter entrado na minha vida, mesmo que tenha demorado. Tenho certeza que daqui pra frente iremos recuperar o tempo desperdiçado.

Ao Renan, por ter 84% de compatibilidade comigo e mesmo assim ser uma das pessoas que eu mais discuto. Por ser o tipo de pessoa que sabe as coisas sobre o mundo pop, sabe as gírias da internet e tem o bom gosto de preferir o telegram. Quando eu digo que é fácil gostar de você, porque você é tão gostável e simpático, é verdade. Obrigada por tudo, mas principalmente, por pegar no pé da Franciele comigo.

À Maitê, por ser a mãezona do grupo. Por ser tão fofa e centrada, mas ao mesmo tempo tão sem paciência, ainda mais com os meus dramas. Quem diria que um simples grupo de estudos para sobreviver à probabilidade iria me proporcionar uma amizade com você, a pessoa que pegava o mesmo ônibus que eu, mas que não me dava oi. Obrigada por sempre rir das minhas reações desproporcionais e por ser uma grande amiga.

Ao Maicon, pelo companheirismo nos trabalhos em dupla, sei que os trabalhos comigo nunca são fáceis, mas até que funcionamos bem juntos. Obrigada pelos dias estudando juntos, pelas zoações, alguns comentários ácidos de ambas as partes, por aguentar meus puxões de orelha e pela troca de conhecimento durante a graduação. Obrigada por ter sido uma parte tão importante na minha formação.

À Luana, pelos incontáveis vídeos de gatinhos fofos e pelas conversas sobre o Bob e a Ísis.

À Raquel, pela animação constante e os elogios durante as monitorias.

À Aline, Giulia, Tainá, Lincon e à todo o restante do pessoal da estatística.

À todos os acontecimentos proporcionados por estes cinco anos na graduação. Da menina que começou tímida, sem falar com ninguém, para a menina que sai cheia de momentos e memórias com tantas pessoas incríveis.

À COMGRAD, pela oportunidade de ser representante discente e poder representar meus colegas. Obrigada à Cleo e aos professores Márcio Valk, Márcia Barbian, e Márcia Echeveste por fazerem com que eu me sentisse à vontade para expressar minhas opiniões e sugestões.

Por fim, um grande agradecimento à UFRGS pelo ensino de qualidade. Nove anos atrás quando decidi o curso que iria fazer, a ideia de cursar uma universidade federal parecia um sonho distante, porém aqui estou eu hoje, entregando o trabalho de conclusão de curso para obtenção do título de bacharel em Estatística e seguindo para o próximo capítulo que é o mestrado.

# Resumo

Este trabalho tem como objetivo aplicar duas abordagens diferentes em dados de partidas de futebol das ligas nacionais dos cinco principais campeonatos europeus. Para a definição destes, foi utilizado o *ranking* baseado no “coeficiente de clubes por país” desenvolvido pela UEFA. A primeira abordagem consiste na utilização de um coeficiente que mede a sorte e habilidade (Aoki, 2017), que foi aplicado nas últimas 56 edições dos campeonatos. A segunda abordagem é a de empregar o modelo de Bradley-Terry dinâmico (Cattelan et al., 2013) que estima a habilidade das equipes em termos de partidas dentro e fora de casa através de um modelo de médias móveis exponencialmente ponderadas (EWMA). Esse modelo é aplicado somente na última temporada dos campeonatos. Os resultados encontrados neste trabalho são, de forma geral, evidências que corroboram com o senso comum. Os times mais ricos do mundo dominam suas ligas, além de deixar o campeonato menos aleatório. Na estimação das habilidades, como o modelo só considera o resultado das partidas, a classificação fica semelhante à classificação final do campeonato, com algumas exceções.

**Palavras-Chave:** Estatística Esportiva, Futebol Europeu, Coeficiente de Sorte e Habilidade, Bradley-Terry, Bradley-Terry Dinâmico.

# Abstract

This work aims to apply two different approaches in the national league football data from the top five European championships. To define the leagues, it was used the ranking based on the “association club coefficients” developed by UEFA. The first approach consists in using a coefficient that measures luck and skill ([Aoki, 2017](#)), which was applied in the last 56 editions of the championships. The second approach was to employ the dynamic Bradley-Terry model ([Cattelan et al., 2013](#)), which estimates the teams’ ability in terms of home and away matches using an exponentially weighted moving average (EWMA) model. This model was only applied in the last season of the leagues. The results found in this work are, in general, evidence that corroborates with common sense. The richest teams in the world dominate their leagues and make them less random. In estimating skills, as the model considers the outcome of the matches, the rating is similar to the final league rating, with some exceptions.

**Keywords:** Sports Statistics, European Football, Coefficient to Measure Skill and Luck, Bradley-Terry, Dynamic Bradley-Terry.



# Sumário

<b>1</b>	<b>Introdução</b>	<b>15</b>
<b>2</b>	<b>O Futebol</b>	<b>17</b>
2.1	A Popularização do Futebol Europeu no Brasil . . . . .	18
2.2	O Coeficiente de Clubes por País da UEFA . . . . .	19
2.3	Os Campeonatos Escolhidos . . . . .	20
<b>3</b>	<b>Sorte e Habilidade</b>	<b>21</b>
3.1	Coeficiente de Sorte e Habilidade . . . . .	22
3.1.1	Vantagem da Equipe Mandante . . . . .	23
3.1.2	Sistema de Pontos do Resultado da Partida . . . . .	24
3.1.3	Interpretação do Coeficiente . . . . .	24
3.2	Avaliação da Significância do Coeficiente . . . . .	25
3.3	Metodologia . . . . .	25
3.3.1	Banco de Dados . . . . .	25
3.3.2	Cálculo do Coeficiente . . . . .	27
3.3.3	Teste de Monte Carlo . . . . .	27
3.3.4	Identificação das Equipes Mais ou Menos Habilidosas . . . . .	28
3.4	Resultados . . . . .	28
3.4.1	Base de Dados . . . . .	28
3.4.2	Coeficiente de Sorte e Habilidade . . . . .	31
3.4.3	Equipes Destoantes . . . . .	36
<b>4</b>	<b>Habilidade das Equipes</b>	<b>45</b>
4.1	O Modelo de Bradley-Terry Para Predição de Resultados de Partidas de Futebol . . . . .	46
4.2	O Modelo de Bradley-Terry Dinâmico . . . . .	47
4.2.1	A função de Verossimilhança . . . . .	48
4.2.2	Avaliação do Modelo . . . . .	48
4.3	Metodologia . . . . .	49
4.3.1	Banco de Dados . . . . .	49
4.3.2	Condição Inicial . . . . .	50
4.4	Resultados . . . . .	50
4.4.1	<i>Bundesliga</i> . . . . .	50
	4.4.1.1 Modelo de Bradley-Terry . . . . .	51
	4.4.1.2 Modelo de Bradley-Terry Dinâmico . . . . .	51
	4.4.1.3 Modelo Estático x Modelo Dinâmico . . . . .	54

4.4.1.4	Avaliação do modelo . . . . .	54
4.4.2	Brasileirão . . . . .	55
4.4.2.1	Modelo de Bradley-Terry . . . . .	55
4.4.2.2	Modelo de Bradley-Terry Dinâmico . . . . .	56
4.4.2.3	Modelo Estático x Modelo Dinâmico . . . . .	59
4.4.2.4	Avaliação do modelo . . . . .	59
<b>5</b>	<b>Trabalhos Futuros</b>	<b>61</b>
<b>6</b>	<b>Considerações Finais</b>	<b>62</b>
	<b>Referências Bibliográficas</b>	<b>64</b>
	<b>Apêndice</b>	<b>70</b>
<b>A</b>	<b>Distribuições do Teste de Monte Carlo para <math>\phi</math></b> . . . . .	<b>71</b>
<b>B</b>	<b>Estatísticas dos jogadores por partida</b> . . . . .	<b>75</b>
<b>C</b>	<b>Código em R do Processo de <i>Webscraping</i></b> . . . . .	<b>77</b>
<b>D</b>	<b>Habilidade das Equipes</b> . . . . .	<b>81</b>
D.1	<i>Premier League</i> . . . . .	81
	D.1.1 Modelo de Bradley-Terry . . . . .	81
	D.1.2 Modelo de Bradley-Terry Dinâmico . . . . .	81
	D.1.3 Modelo Estático x Modelo Dinâmico . . . . .	84
	D.1.4 Avaliação do modelo . . . . .	85
D.2	<i>La Liga</i> . . . . .	85
	D.2.1 Modelo de Bradley-Terry . . . . .	86
	D.2.2 Modelo de Bradley-Terry Dinâmico . . . . .	86
	D.2.3 Modelo Estático x Modelo Dinâmico . . . . .	89
	D.2.4 Avaliação do modelo . . . . .	89
D.3	<i>Ligue 1</i> . . . . .	90
	D.3.1 Modelo de Bradley-Terry . . . . .	91
	D.3.2 Modelo de Bradley-Terry Dinâmico . . . . .	91
	D.3.3 Modelo Estático x Modelo Dinâmico . . . . .	94
	D.3.4 Avaliação do modelo . . . . .	94
D.4	<i>Serie A</i> . . . . .	95
	D.4.1 Modelo de Bradley-Terry . . . . .	96
	D.4.2 Modelo de Bradley-Terry Dinâmico . . . . .	96
	D.4.3 Modelo Estático x Modelo Dinâmico . . . . .	99
	D.4.4 Avaliação do modelo . . . . .	100

## Lista de Figuras

Figura 3.1:	Probabilidades estimadas para os resultados das partidas nas ligas	29
Figura 3.2:	Probabilidades estimadas para os resultados das partidas . . . . .	30
Figura 3.3:	Boxplot dos valores de $\phi$ . . . . .	31
Figura 3.4:	Valores de $\phi$ ao longo das temporadas . . . . .	32
Figura 3.5:	Histograma para $\phi$ entre os campeonatos . . . . .	33
Figura 3.6:	Representação gráfica das pontuações finais do campeonato alemão de 1968 e campeonato espanhol de 2014 . . . . .	33
Figura 3.7:	Exemplos de distribuições do Teste de Monte Carlo para $\phi$ . . . . .	35
Figura 3.8:	Percentual geral de todos os times que foram retirados para as temporadas se tornarem aleatórias . . . . .	36
Figura 3.9:	Percentual de times que foram retirados para a temporada se tornar aleatória dentro de cada campeonato . . . . .	37
Figura 3.10:	Clubes mais destoantes dentro de cada campeonato . . . . .	38
Figura 3.11:	Quantidade de clubes retirados por posição . . . . .	39
Figura 3.12:	Quantidade de retiradas dentro de cada campeonato . . . . .	40
Figura 3.13:	Porcentagem de saídas de cada equipe da Bundesliga . . . . .	42
Figura 3.14:	Porcentagem de saídas de cada equipe da Premier League . . . . .	42
Figura 3.15:	Porcentagem de saídas de cada equipe da La Liga . . . . .	43
Figura 3.16:	Porcentagem de saídas de cada equipe da Ligue 1 . . . . .	43
Figura 3.17:	Porcentagem de saídas de cada equipe da Serie A . . . . .	44
Figura 4.1:	Habilidades e QSE estimados das 18 equipes da edição de 2018 da <i>Bundesliga</i> pelo modelo de Bradley-Terry . . . . .	51
Figura 4.2:	Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da <i>Bundesliga</i> - 1 <sup>o</sup> ao 9 <sup>o</sup> colocado . . . . .	52
Figura 4.3:	Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da <i>Bundesliga</i> - 10 <sup>o</sup> ao 18 <sup>o</sup> colocado . . . . .	53
Figura 4.4:	Boxplot do RPS para a <i>Bundesliga</i> de 2018 computado para o modelo não-estruturado, BT-Dinâmico e previsões baseadas em proporções empíricas . . . . .	55
Figura 4.5:	Habilidades e QSE estimados das 20 equipes da edição de 2019 do Brasileirão pelo modelo de Bradley-Terry . . . . .	56
Figura 4.6:	Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2019 do Brasileirão - 1 <sup>o</sup> ao 12 <sup>o</sup> colocado . . . . .	57

Figura 4.7: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2019 do Brasileirão - 13 <sup>o</sup> ao 20 <sup>o</sup> colocado . . . . .	58
Figura 4.8: Boxplot do RPS para o Brasileirão de 2019 computado para o modelo não-estruturado, BT-Dinâmico e previsões baseadas em proporções empíricas . . . . .	60
Figura A.1: A densidade do Teste de Monte Carlo para os cinco campeonatos (Década de 1960) . . . . .	71
Figura A.2: A densidade do Teste de Monte Carlo para os cinco campeonatos (Década de 1970) . . . . .	72
Figura A.3: Densidade do Teste de MC (Década de 1980) . . . . .	72
Figura A.4: A densidade do Teste de Monte Carlo para os cinco campeonatos (Década de 1990) . . . . .	73
Figura A.5: A densidade do Teste de Monte Carlo para os cinco campeonatos (Década de 2000) . . . . .	73
Figura A.6: A densidade do Teste de Monte Carlo para os cinco campeonatos (Década de 2010) . . . . .	74
Figura D.1: Habilidades e QSE estimados das 20 equipes da edição de 2018 da <i>Premier League</i> pelo modelo de Bradley-Terry . . . . .	81
Figura D.2: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da <i>Premier League</i> - 1 <sup>o</sup> ao 12 <sup>o</sup> colocado . . . . .	82
Figura D.3: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da <i>Premier League</i> - 13 <sup>o</sup> ao 20 <sup>o</sup> colocado . . . . .	83
Figura D.4: Boxplot do RPS para a <i>Premier League</i> de 2018 computado para o modelo não-estruturado, BT-Dinâmico e previsões baseadas em proporções empíricas . . . . .	85
Figura D.5: Habilidades e QSE estimados das 20 equipes da edição de 2018 da <i>La Liga</i> pelo modelo de Bradley-Terry . . . . .	86
Figura D.6: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da <i>Premier League</i> - 1 <sup>o</sup> ao 12 <sup>o</sup> colocado . . . . .	87
Figura D.7: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da <i>La Liga</i> - 13 <sup>o</sup> ao 20 <sup>o</sup> colocado . . . . .	88
Figura D.8: Boxplot do RPS para a <i>La Liga</i> de 2018 computado para o modelo não-estruturado, BT-Dinâmico e previsões baseadas em proporções empíricas . . . . .	90
Figura D.9: Habilidades e QSE estimados das 20 equipes da edição de 2018 da <i>Ligue 1</i> pelo modelo de Bradley-Terry . . . . .	91
Figura D.10: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da <i>Ligue 1</i> - 1 <sup>o</sup> ao 12 <sup>o</sup> colocado . . . . .	92
Figura D.11: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da <i>Ligue 1</i> - 13 <sup>o</sup> ao 20 <sup>o</sup> colocado . . . . .	93

Figura D.12:Boxplot do RPS para a <i>Ligue 1</i> de 2018 computado para o modelo não-estruturado, BT-Dinâmico e previsões baseadas em proporções empíricas . . . . .	95
Figura D.13:Habilidades e QSE estimados das 20 equipes da edição de 2018 da <i>Serie A</i> pelo modelo de Bradley-Terry . . . . .	96
Figura D.14:Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da <i>Serie A</i> - 1 <sup>o</sup> ao 12 <sup>o</sup> colocado . . . . .	97
Figura D.15:Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da <i>Serie A</i> - 13 <sup>o</sup> ao 20 <sup>o</sup> colocado . . . . .	98
Figura D.16:Boxplot do RPS para a <i>Serie A</i> de 2018 computado para o modelo não-estruturado, BT-Dinâmico e previsões baseadas em proporções empíricas . . . . .	100

## Lista de Tabelas

Tabela 2.1: Primeiras cinco posições do Ranking de coeficiente de clubes por país . . . . .	19
Tabela 3.1: Probabilidade dos resultados na primeira e última temporada . . . . .	30
Tabela 3.2: Valor médio e desvio padrão das estimativas das probabilidades dos resultados das partidas das ligas . . . . .	31
Tabela 3.3: Classificação final da <i>Bundesliga</i> de 1968 . . . . .	34
Tabela 3.4: Classificação final da <i>La Liga</i> de 2014 . . . . .	34
Tabela 3.5: Número de vezes que o time mais destoante foi o campeão ou lanterna . . . . .	37
Tabela 3.6: Classificação final da <i>Bundesliga</i> de 2008 . . . . .	40
Tabela 3.7: Quantidade de diferentes clubes excluídos em cada edição das ligas . . . . .	41
Tabela 4.1: Condições iniciais utilizadas no algoritmo do modelo de Bradley-Terry Dinâmico de cada liga . . . . .	50
Tabela 4.2: Tabela resumo das habilidades e QSE estimados pelo modelo BT-Estático e as habilidades suavizadas médias pelo BT-Dinâmico em termos gerais, como mandante e visitante da <i>Bundesliga</i> de 2018 . . . . .	54
Tabela 4.3: Tabela resumo das habilidades e QSE estimados pelo modelo BT-Estático e as habilidades suavizadas médias pelo BT-Dinâmico em termos gerais, como mandante e visitante do Brasileirão de 2019 . . . . .	59
Tabela 1: Tabela resumo das habilidades e QSE estimados pelo modelo BT-Estático e as habilidades suavizadas médias pelo BT-Dinâmico em termos gerais, como mandante e visitante da <i>Premier League</i> de 2018 . . . . .	84
Tabela 2: Tabela resumo das habilidades e QSE estimados pelo modelo BT-Estático e as habilidades suavizadas médias pelo BT-Dinâmico em termos gerais, como mandante e visitante da <i>La Liga</i> de 2018 . . . . .	89
Tabela 3: Tabela resumo das habilidades e QSE estimados pelo modelo BT-Estático e as habilidades suavizadas médias pelo BT-Dinâmico em termos gerais, como mandante e visitante da <i>Ligue 1</i> de 2018 . . . . .	94
Tabela 4: Tabela resumo das habilidades e QSE estimados pelo modelo BT-Estático e as habilidades suavizadas médias pelo BT-Dinâmico em termos gerais, como mandante e visitante da <i>Serie A</i> de 2018 . . . . .	99

# 1 Introdução

Um dos esportes mais populares do mundo é o futebol, que possui alta quantidade de torcedores, admiradores e simpatizantes. De acordo com dados da cobertura oficial da *Fédération Internationale de Football Association* (FIFA, 2018), a Copa do Mundo na Rússia bateu recordes de audiência, em que mais da metade da população mundial (com idade igual ou superior à quatro anos), cerca de 3.572 bilhões de telespectadores, assistiram a maior competição mundial entre seleções.

Mesmo estando atrás de esportes como o beisebol e o basquete em termos de análises sofisticadas e acesso a dados mais complexos (Brooks et al., 2016), aos poucos o futebol vem criando espaços dentro dos clubes para analistas utilizarem os dados disponíveis para tomadas de decisões. Ian Graham, diretor de pesquisa do *Liverpool*, foi o responsável por sugerir contratações como as de Salah, Van Dijk e Alisson, que influenciaram a grande campanha do clube na última temporada e culminou no título de campeão da *Champions League*. Quando o *Liverpool* estava a procura de um ponta esquerda, foram os dados de Graham que sugeriram a compra de Philippe Coutinho, comprado por €13 milhões e depois vendido ao Barcelona por €170 milhões (Schoenfeld, 2019).

Na temporada de 2014-15 da *Premier League* o *Leicester* conseguiu reagir no campeonato e se salvou do rebaixamento depois de uma sequência de bons resultados (sete vitórias, entre nove partidas disputadas). Na edição seguinte, sob comando de Claudio Ranieri, consagrou-se campeão do campeonato inglês, uma vitória que surpreendeu a todos. Uma das estratégias do clube foi a utilização de estatísticas como forma de entender os pontos fortes e fracos dos seus próprios jogadores e assim construir um time competitivo e encaixado. Mesmo tendo dificuldades na troca de passes e posse de bola, o *Leicester* tinha um dos melhores ataques da competição. Sabendo de suas fraquezas, construíram um esquema tático que permitia ao adversário controlar a partida enquanto esperavam por um erro. Quando este acontecia, utilizavam-se do fato de possuir jogadores velozes no plantel e assim marcar gols de contra-ataque ou de passes longos (Pham, 2016).

O futebol é decidido por chances e nem sempre isso significa a equipe mais habilidosa vencer. Um passe errado pode levar a um contra-ataque adversário que resulta em uma jogada que pode definir a partida. Um chute de fora da área pode desviar no zagueiro e entrar na rede. O melhor jogador da partida pode sofrer uma lesão, desfaltar o time e mudar o curso da jogo. A imprevisibilidade nos resultados dos jogos é o que faz o esporte tão interessante para o público (Brooks et al., 2016; Buraimo e Simmons, 2015).

Em termos de sorte e habilidade, nem sempre existe uma linha que separa onde

termina um e começa outro. [Simon \(2017\)](#) comenta acerca do merecimento da sorte, sob argumento de que jogadores mais habilidosos se colocam em posições em que a sorte irá desempenhar papel fundamental e decisivo no lance.

Existem fatores não relacionados a habilidade, mas que influenciam o resultado de uma partida. Um exemplo é a vantagem da equipe mandante, que apesar de possuir causas conhecidas, elas ainda não foram bem determinadas. Questões como efeito da torcida, efeitos de viagem, familiaridade, viés do árbitro, territorialidade e esquema tático ([Pollard, 2008](#)) ajudam a explicar como jogar dentro do próprio estádio pode trazer maiores chances de vitória para uma equipe.

Porém, mesmo que uma partida possa ser explicada pela quantidade de chutes de um jogador, por um lance certo, por sorte ou até mesmo pelo simples fato de se estar disputando a partida dentro de casa, a habilidade da equipe ainda mantém um papel importante dentro do esporte. Uma abordagem comum é utilizar o modelo de Bradley-Terry ([Bradley e Terry, 1952](#)) para modelar a habilidade de times.

O objetivo deste trabalho é analisar as cinco principais ligas nacionais de futebol da Europa a fim de verificar a relação entre sorte e habilidade nos campeonatos, bem como estimar as habilidades das equipes por meio de modelos estatísticos.

Esta monografia está organizada da seguinte forma: o [Capítulo 2](#) apresenta uma breve introdução sobre a história do futebol e a relação com a estatística dentro desse esporte. Além disso, são expostas as motivações que levam os brasileiros a se interessarem pelos campeonatos europeus, como por exemplo a qualidade superior e o fato dos melhores jogadores do mundo estarem nos clubes destas ligas. Uma breve explicação do *ranking* utilizado para classificar as cinco principais ligas da Europa com base no coeficiente de clubes por país da União das Associações Europeias de Futebol (UEFA) também é abordada, bem como uma breve introdução dos cinco campeonatos escolhidos. Em seguida são apresentadas duas abordagens diferentes aplicadas às ligas selecionadas: no [Capítulo 3](#) é apresentado o coeficiente de sorte e habilidade ([Aoki, 2017](#)), uma medida que permite avaliar se o campeonato é decidido de forma aleatória ou se as habilidades das equipes interferem no desfecho, enquanto que o [Capítulo 4](#) apresenta uma versão dinâmica do modelo de Bradley-Terry ([Cattelan et al., 2013](#)), que modela a evolução temporal das habilidades das equipes. A organização deste trabalho é dada de forma que a metodologia empregada e resultados obtidos sejam expostos conjuntamente aos modelos empregados. Para a habilidade das equipes, são apresentados somente os resultados da *Bundesliga* e Brasileiro. O [Capítulo 5](#) fornece sugestões de trabalhos futuros a serem desenvolvidos e o [Capítulo 6](#) apresenta as discussões finais acerca dos resultados encontrados nas duas abordagens. E, por fim, são apresentados os apêndices, em que tem-se em [A](#) as distribuições do teste de Monte Carlo para o coeficiente de sorte e habilidade, em [B](#) uma descrição das estatísticas dos jogadores por partida, em [C](#) o código disponibilizado em R do processo de *webscraping* e em [D](#) os resultados da aplicação dos modelos de Bradley-Terry e Bradley-Terry Dinâmico na *Premier League*, *La Liga*, *Ligue 1* e *Serie A*.



## 2 O Futebol

O futebol contemporâneo surgiu em 1863 na Inglaterra quando a primeira associação de futebol foi formada no país e regras como tamanho e peso da bola foram estabelecidas. O esporte, que era tido inicialmente como um entretenimento para a classe trabalhadora britânica, não demorou muito para aumentar a popularidade, especialmente na Europa, com várias ligas surgindo: Holanda e Dinamarca em 1889, Suíça e Bélgica em 1895, Itália em 1898, Alemanha em 1900, Hungria em 1901, França em 1903, Finlândia em 1907, entre outras. Na América do Sul, as ligas na Argentina (1893) e Uruguai (1900) surgiram paralelamente com as europeias e o esporte difundiu-se pelo continente durante o século 20, passando a fazer parte da cultura da população sul-americana, com a primeira Copa do Mundo, ocorrida em 1930, sendo disputada em solo uruguaio (Crocombe, 2018). No Brasil, surgiu de forma organizada através de Charles Miller que, ao ir estudar na Inglaterra conheceu o esporte. Ao retornar ao país em 1894, trouxe consigo acessórios utilizados para a prática do jogo e assim colaborou com a divulgação do futebol entre os trabalhadores de empresas inglesas que atuavam em São Paulo (Mills, 2005). Segundo Helal et al. (2000), inicialmente elitista, a segregação com negros e pobres era comum, porém depois de lutas e resistências houve uma democratização, ascensão e afirmação desses grupos marginalizados.

Se antigamente o que contava era o instinto e tradições para saber o que era bom e mau futebol, atualmente se tem recorrido aos analistas futebolísticos para dar algum sentido aos dados e tirar deles algum aprendizado (Anderson e Sally, 2013), porém mesmo atraindo grande quantidade de adeptos, o futebol ainda carece de análises mais sofisticadas se comparado a outros esportes profissionais (Brooks et al., 2016).

O *podcast* FiveThirtyEight (2016) discute sobre as três eras da estatística no futebol. A primeira refere-se aos tempos em que a análise era escassa e corresponde a maior parte da história do esporte. Algumas pessoas anotavam manualmente alguns resultados, como por exemplo Charles Reep, que em 1950 elaborava relatórios por meio de um sistema de anotações de distância, direção, altura e resultados dos passes e finalizações (Freitas, 2017).

A segunda fase veio entre o final da década de 90 e início dos anos 2000 com o surgimento de empresas de dados esportivos como a Opta Sports, que contava os chutes e chutes ao gol, que servem de base para a métrica chamada *expected goals* (xG), que mensura a qualidade de um chute baseando-se em várias variáveis (OptaSports, 2019a) e equivale a uma das estatísticas mais avançadas voltadas para o público em análises de futebol.

A terceira etapa ainda está em desenvolvimento e consiste em rastreamento de jogadores baseado em tecnologias como as câmeras SportVU, usadas já no basquete. Esses dados já estão sendo fornecidos às equipes, mas não são informações disponíveis ao público. A [STATS \(2019\)](#) possui várias ferramentas baseadas em algoritmos estatísticos e inteligência artificial, como por exemplo:

- o sistema de rastreamento que fornece dados por meio da extração e processamento de coordenadas dos jogadores e da bola;
- a tecnologia GPS em tempo real que permite a tomada de decisão ao vivo;
- uma auto codificação e edição de vídeo que permite aos usuários acesso à análises personalizadas e *feedback* de desempenho otimizado das equipes enquanto a partida ainda está em andamento;
- uma plataforma *on-line* projetada para ajudar os treinadores a avaliar os efeitos da carga de treinamento no desempenho e a construção de perfis fisiológicos completos para minimizar o risco de lesões.

## 2.1 A Popularização do Futebol Europeu no Brasil

A globalização trouxe a descaracterização do elemento nacional, com os principais jogadores brasileiros indo atuar nos campeonatos europeus e a seleção brasileira transformando-se, a partir dos anos 1990, em uma seleção “estrangeira” ([Guterman, 2009](#)). Entre os 23 profissionais convocados para as últimas edições da Copa do Mundo, tem-se que, entre todo o elenco, 87% dos jogadores estavam em times da Europa em 2010 ([Dias e Lavinas, 2010](#)), 78,2% em 2014 ([Epoca, 2014](#)) e, na Copa da Rússia de 2018, 82,6% ([Prosperi, 2018](#)), o que evidencia que o “apelido” concedido ao time brasileiro na década de 90 estendeu-se até os dias atuais.

Em um estudo realizado pelo *CIES Football Observatory* (Centro Internacional de Estudos Esportivos), observou-se que o Brasil é o maior exportador de futebolistas do mundo, com mais de 1200 brasileiros jogando fora de seu país de origem, representando 22,7% das exportações totais e 43,5%, levando em conta somente os dez maiores países exportadores. Dentre esses números, 65% dos jogadores estão em times da Europa ([Poli et al., 2017](#)).

Em sua coluna esportiva, [Boas \(2019\)](#) discorre acerca dos motivos do futebol europeu ser tão popular no Brasil. Além da qualidade superior, há uma constante evolução no jogo tático e técnico, que em contrapartida, tem-se um conservadorismo no modo de jogar no Brasil e na América Latina como um todo. Os times mais vencedores e ricos estão na Europa, assim como os melhores jogadores e competições. Um jogo entre um time brasileiro e um time europeu soa quase que injusto, pois a qualidade dos jogadores é desproporcional. Em 2017 a final da Copa do Mundo de Clubes da FIFA foi disputada entre Grêmio e Real Madrid e a [FoxSports \(2017\)](#), com base nos números do Transfermarkt, mostrou que o valor de mercado do time espanhol era quase dez vezes o valor de mercado do time gaúcho (R\$ 2.9 bilhões contra R\$ 307 milhões).

A Europa é o coração do futebol e, além de abrigar grande parte das principais seleções do mundo, também conta com os maiores clubes nacionais, que detém

a força dominante mundial em termos de sucesso esportivo e tamanho financeiro (Gerrard, 2004, p.39).

## 2.2 O Coeficiente de Clubes por País da UEFA

Diante do atual cenário do futebol mundial e da organização dos campeonatos, optou-se por utilizar o *ranking* de coeficientes de clubes por país definido pela *Union of European Football Associations* (UEFA), autoridade desportiva que representa as federações europeias de futebol, agindo sob os interesses destas junto à FIFA, como critério de delimitação para definir os principais torneios da Europa. Foram escolhidas as cinco primeiras ligas nacionais da Europa: os campeonatos da Espanha, Inglaterra, Itália, Alemanha e França (UEFA, 2019b).

Tabela 2.1: Primeiras cinco posições do Ranking de coeficiente de clubes por país

Posição	País	Pontuação por Temporada					Pontuação
		14/15	15/16	16/17	17/18	18/19	
1	Espanha	20,214	23,928	20,142	19,714	19,571	103,569
2	Inglaterra	13,571	14,250	14,928	20,071	22,642	85,462
3	Itália	19,000	11,500	14,250	17,333	12,642	74,725
4	Alemanha	15,857	16,428	14,571	9,857	15,214	71,927
5	França	10,916	11,083	14,416	11,500	10,583	58,498

A Tabela 2.1 fornece as primeiras cinco posições no *ranking* de coeficientes de clubes por país ao final de temporada de 2018-19. A regra utilizada para definir o *ranking* tem por base os resultados nas últimas cinco edições da *UEFA Champions League* (UCL), conhecida como apenas *Champions League*, ou Liga dos Campeões e da *UEFA Europa League* (UEL), conhecida também como *Europe League* ou Liga Europa. O sistema de pontos do coeficiente possui três critérios (UEFA, 2019a):

1. O clube conquista um ponto por empate e dois pontos por vitória, exceto nas pré-eliminatórias e *play-offs*, em que recebem um ponto por vitória e meio ponto por empate.
2. A partir das oitavas-de-final da UCL e quartas-de-final da UEL, os clubes recebem um ponto extra por cada fase em que marquem presença.
3. A presença na fase de grupos da *Champions* ou classificação para as oitavas-de-final acarreta mais quatro pontos extras ao clube.

O cálculo do coeficiente (UEFA, 2019a) é feito através do número total de pontos dividido pelo número total de clubes que representam a federação nas duas competições na temporada atual. O resultado é somado aos valores das quatro temporadas anteriores, obtendo-se assim um valor final, que representa o coeficiente. Em caso de empate, o desempate se dá em favor da federação que possuir valor mais elevado na pontuação da temporada mais atual. O *ranking* é atualizado após cada rodada das competições.

Como as finais da UCL e UEL ocorreram em 1 de junho e 29 de maio de 2019, respectivamente, as pontuações por temporada da Tabela 2.1 só irão sofrer alterações quando os dois campeonatos que dizem respeito à temporada de 2019/2020 começarem e então os pontos obtidos em 2014/2015 serão desconsiderados do cálculo final do *ranking*.

## 2.3 Os Campeonatos Escolhidos

A seguir, são apresentadas brevemente as cinco ligas escolhidas, que além de serem as que melhor se saem nas competições UCL e UEL, também são as mais assistidas da Europa ([Kenmare, 2018](#)), portanto as mais populares.

### 1. Campeonato Espanhol:

A *Primera División*, popularmente conhecida por *La Liga*, diz respeito à maior competição de futebol masculino da Espanha. Criada em 1929, tinha como maior objetivo atrair maiores públicos e partidas mais frequentes entre os principais clubes do país. A primeira edição contou com dez clubes jogando sob o sistema de pontos corridos em dois turnos ([Stein, 2019](#)).

### 2. Campeonato Inglês:

Entre os anos de 1888 à 1992 o torneio de maior importância na Inglaterra era o *Football League First Division*, atualmente conhecido como *EFL Championship*, cujo formato original consistia em um torneio simples composto por 12 equipes. Após uma série de divergências entre dirigentes, jogadores e emissoras de televisão, em 1992 foi anunciado a criação de uma nova liga, a *English Premier League* (EPL), que passou a configurar como a principal competição de futebol do país ([EPL, 2019](#)).

### 3. Campeonato Italiano:

Na Itália, a principal competição de liga profissional de clubes de futebol é a *Lega Nazionale Professionisti Serie A*, conhecida simplesmente como *Serie A*. O campeonato existe desde 1898, mas a divisão era feita através de grupos regionais, até que em 1929 foi estruturado aos moldes atuais.

### 4. Campeonato Alemão:

A *Fußball-Bundesliga*, conhecida somente por *Bundesliga*, fundada em 1962, é uma liga profissional da Alemanha, considerada a principal competição de futebol do país.

### 5. Campeonato Francês:

Fundada em 1932 com a implementação do futebol profissional, a *Division 1*, popularmente conhecida como *Ligue 1* (ou *Division 1* até 2002), consiste na primeira divisão da liga francesa, sendo a principal competição de futebol no país. É organizada pela *Ligue de Football Professionnel* desde sua fundação. É a que menos mudou o número total de competidores, já que a quantidade oscilou entre 18 e 20 clubes ao longo dos anos.

Todos os cinco campeonatos utilizam o sistema de pontos corridos, onde cada time enfrenta seus adversários em duas ocasiões: uma como mandante e outra como visitante. O time mandante (*home team*) realiza a partida “dentro de casa”, ou seja, no seu próprio estádio, enquanto que o time visitante (*away team*) joga “fora de casa”, no estádio do oponente. Atualmente 20 clubes disputam cada um dos torneios, com exceção da *Bundesliga*, que possui 18 times na disputa.

### 3 Sorte e Habilidade

Sem um certo nível de competição, um esporte torna-se pouco atraente aos olhos do telespectador, o que faz com que busquem outras opções de entretenimento (Wills, 2017). A concorrência acirrada no futebol cria uma imprevisibilidade mais acentuada nos resultados das partidas, o que gera um interesse maior do público (Brooks et al., 2016). Outros fatores que influenciam esse interesse é a mistura entre sorte e habilidade (Chan et al., 2008).

A sorte e a habilidade sempre estiveram presentes no futebol, onde um lance pode definir toda a partida. Simon (2017) argumenta que a linha entre esses dois elementos é tênue e nem sempre facilmente identificável, tendo variações e papéis diferentes, dependendo do esporte. Outra questão levantada pelo autor é o merecimento da sorte: alguns jogadores são mais habilidosos do que outros e essa aptidão é a responsável por colocar o atleta em uma posição em que a sorte irá desempenhar o papel determinante no lance e por consequência no resultado final. Por exemplo, um chute de fora da área feito pelo praticante teve sua trajetória deslocada e assim resultando no gol. Pode-se concluir que o êxito só foi possível por causa do vento, porém se não fosse a habilidade do jogador ao dar o chute, a sorte não teria entrado em ação de forma tão certa.

Em Wood et al. (2015) é discutido acerca do papel da sorte nas cobranças de pênaltis. Se o sucesso na disputa é predominantemente por causa da sorte, então as taxas de sucesso das equipes seriam aproximadamente iguais, porém não é o que acontece. Desde 1982, a seleção da Alemanha ganhou 85% das decisões de pênaltis em que participou, já a seleção da Inglaterra obteve uma acurácia de apenas 14%, o que sugere que a sorte não é um fator preponderantemente importante, dado a disparidade entre as taxas. Já em Bell et al. (2013) foi ajustado um modelo que avalia até que ponto o desempenho dos treinadores nas temporadas entre 2004 e 2009 da *Premier League* pode ser atribuído à sorte ou habilidade, quando mensurados separadamente às qualidades dos clubes, chegando a conclusão que o sucesso de alguns treinadores da liga não pode ser atribuído exclusivamente à sorte ou qualidade dos jogadores, mas sim também às habilidades do próprio treinador.

Este capítulo tem como objetivo analisar as cinco principais ligas de futebol da Europa em termos de sorte e habilidade em um campeonato, por meio da medida proposta por Aoki (2017), de forma a avaliar se há diferentes habilidades entre os clubes ou se o campeonato é determinado aleatoriamente, bem como também identificar, quando houver, quais equipes que, se retiradas, deixariam a liga aleatória.

### 3.1 Coeficiente de Sorte e Habilidade

Spiegelhalter (2007) propôs uma medida que mede o quanto da variação dos pontos de um campeonato é devida à aleatoriedade, em que seu valor deveria estar sempre entre 0% e 100%, porém Aoki et al. (2016) mostraram que quando essa medida extrapola os 100%, não necessariamente indica que o resultado do campeonato foi totalmente devido ao acaso, como indica o autor. Isto acontece porque a fórmula é dada pela razão entre a variância teórica considerando igual habilidade dos times e a variância empírica dos pontos obtidos pelos times ao final do campeonato, portanto quando essa razão for maior que um, tem-se que a variância teórica é superior à variância empírica, indicando que os pontos obtidos pelos clubes ao final da liga são muito similares e então a interpretação da métrica torna-se errônea. Aoki et al. (2016) propuseram uma correção, utilizando a medida da variabilidade teórica como um valor intermediário quando os times não possuem habilidades diferentes, ao contrário de Spiegelhalter (2007), que utilizou essa medida como se fosse um mínimo de variabilidade.

O coeficiente que mede a sorte e a habilidade, proposto por Aoki (2017), permite avaliar se há diferentes habilidades entre as equipes de uma temporada ou se o campeonato é determinado aleatoriamente. A única limitação imposta no uso do coeficiente é que os campeonatos precisam seguir o *Round-Robin Tournament*, conhecido no Brasil como sistema de pontos corridos, em que há  $k+1$  equipes, com cada equipe jogando  $k$  vezes como mandante e  $k$  vezes como visitante, totalizando  $2k$  partidas.

Seja  $X_m$  uma variável aleatória com distribuição multinomial e que representa os pontos ganhos de uma equipe quando esta joga em casa. Os valores que  $X_m$  assume em caso de empate e derrota são 1 e 0, respectivamente, com suas probabilidades intrínsecas  $P_e$  e  $P_v$ . Em caso de vitória,  $X_m$  pode assumir 2 ou 3 pontos, com probabilidade  $P_m$ , de acordo com o campeonato e ano em questão.

As probabilidades  $P_m, P_v$  e  $P_e$  são estimadas com base em suas ocorrências nos torneios. Por exemplo, seja  $N$  o número total de jogos no campeonato e  $W_m$  a quantidade de vezes que uma equipe mandante venceu um jogo, então  $\hat{P}_m = W_m/N$ . Similarmente, utilizando  $W_v$  e  $W_e$  como a quantidade de vitórias de uma equipe visitante e a quantidade de empates, respectivamente, no campeonato, tem-se as estimativas das probabilidades de  $\hat{P}_m, \hat{P}_v$  e  $\hat{P}_e$ .

Seja  $\mu_{X_m}$  e  $\sigma_{X_m}^2$  o valor esperado e variância associados à  $X_m$ , com expressões matemáticas dada por:

$$\mu_{X_m} = \begin{cases} (2 \times P_m) + (1 \times P_e), & \text{se vitória vale 2 pontos} \\ (3 \times P_m) + (1 \times P_e), & \text{caso contrário} \end{cases} \quad (3.1)$$

e

$$\sigma_{X_m}^2 = \begin{cases} (2^2 \times P_m) + (1^2 \times P_e) - \mu_{X_m}^2, & \text{se vitória vale 2 pontos} \\ (3^2 \times P_m) + (1^2 \times P_e) - \mu_{X_m}^2, & \text{caso contrário} \end{cases} \quad (3.2)$$

em que ambas as fórmulas dependem da quantidade de pontos distribuídos ao time vencedor.

As definições seguem similarmente para  $X_v$ , que representa os pontos ganhos de uma equipe quando esta joga fora de casa. O valor esperado e sua variância,  $\mu_{X_v}$  e  $\sigma_{X_v}^2$ , são obtidos equivalentemente às expressões em (3.1) e (3.2).

Seja  $Y_{2k}$  uma variável aleatória que representa o total de pontos obtidos pela equipe ao final do campeonato e  $X_{m_i}$  e  $X_{v_i}$  os pontos obtidos no  $i$ -ésimo jogo dentro e fora de casa, respectivamente. Então:

$$Y_{2k} = \sum_{i=1}^{2k} (X_{m_i} + X_{v_i}).$$

Para um número  $k$  grande de jogos no campeonato e supondo independência estocástica entre as partidas, pelo Teorema Central do Limite, se as equipes possuírem as mesmas habilidades, então a distribuição final dos dados  $Y_{2k}$  é aproximadamente dada por:

$$Y_{2k} \sim N(\mu_{2k}, \sigma_{2k}^2),$$

com os parâmetros especificados em  $\mu_{2k} = k(\mu_{X_m} + \mu_{X_v})$  e  $\sigma_{2k}^2 = k(\sigma_{X_m}^2 + \sigma_{X_v}^2)$ .

O coeficiente definido por [Aoki \(2017\)](#), denotado  $\phi$ , pode ser usado na comparação da influência da sorte em diferentes temporadas, ligas e esportes e é dado por:

$$\phi = \frac{S^2 - \sigma_{2k}^2}{S^2}, \quad (3.3)$$

em que  $S^2$  representa a variância observada na distribuição dos pontos finais do campeonato e  $\sigma_{2k}^2$  a variância teórica, em que os resultados dos  $2k$  jogos são aleatórios.

### 3.1.1 Vantagem da Equipe Mandante

Vários contextos influenciam o resultado do jogo, porém só é considerado no cálculo do coeficiente o fator mandante/visitante. A medida pode ser estendida levando em consideração outros contextos que interferem nos resultados dos jogos, se estas informações estiverem disponíveis.

A existência de uma vantagem do time mandante sobre o time visitante é conhecida, porém as causas do porquê isso ocorre ainda não são bem determinadas. Em seu artigo, [Pollard \(2008\)](#) pondera acerca desses motivos, que tendem a ser maiores no futebol do que em outros esportes. O fator mais óbvio é o “efeito torcida”, porém é desconhecido se esse resultado é devido a uma vantagem para a equipe que joga em casa, pelo apoio da torcida com os cantos e vibrações, que estimulam os jogadores, ou uma desvantagem para o time que joga fora de casa, com os xingamentos e vaias direcionadas quando o time oponente possui a posse da bola que podem criar pressão extra aos profissionais.

Outros dois pontos são “efeitos de viagem” e “familiaridade”: enquanto para o primeiro tem-se resultados contraditórios acerca da desvantagem sobre o time visitante, para o segundo fator observa-se a familiaridade com o estádio, gramado, clima e altitude como aliados à vantagem do time mandante. Porém esses pontos não parecem afetar em caso de *derbies* (os clássicos), isso porque normalmente as rivalidades são locais, portanto os efeitos de viagem e familiaridade quase não existem. No caso do campeonato italiano, por exemplo, a maioria dos estádios não são dos clubes: o estádio Olímpico de Roma é de propriedade do CONI (Comitê Olímpico Nacional Italiano) e utilizado pela *Roma* e *Lazio*. Já o Giuseppe Meazza (conhecido como San Siro) é da prefeitura de Milão, mas tanto a *Internazionale* quanto o *Milan* utilizam-o nos jogos em que atuam como mandante.

Outras questões que influenciam são: “viés do árbitro”, que tendem a serem mais rígidos com os visitantes; “territorialidade”, principalmente em partidas entre rivais, que a competitividade entre os jogadores e torcida são mais acentuadas; e “táticas especiais”, já que as equipes tendem a jogar sob um esquema mais defensivo quando não estão no próprio estádio, o que influencia diretamente na posse de bola e chutes a gol, já que é comum os times visitantes jogarem por um empate.

### 3.1.2 Sistema de Pontos do Resultado da Partida

Inicialmente o sistema de pontos para o resultado da partida era definido por 2 pontos a vitória, 1 o empate e 0 a derrota. Em 1980, como uma solução para vencer a crise econômica no futebol inglês e queda de presença nos estádios, foi proposto pela liga de futebol inglesa um novo sistema, em que era concedido ao time vitorioso 3 pontos pela vitória. O objetivo era deixar a partida mais ofensiva, já que os clubes não se contentariam com o empate e correriam mais riscos para marcar gols, o que deixaria o esporte mais atrativo para os torcedores e consequentemente despertaria o interesse de assistir aos jogos no estádio (Aylott e Aylott, 2007).

A *Premier League* optou por utilizar o novo sistema de pontos a partir da temporada de 1981-82, porém somente na década de 90 que as outras principais ligas da Europa resolveram seguir o mesmo caminho: a *Ligue 1* e a *Serie A* implementaram o sistema em 1994. Como a FIFA adotou formalmente o sistema em 1995, outros campeonatos como a *La Liga* e a *Bundesliga* seguiram a federação ao aderirem “três pontos pela vitória”, tornando posteriormente o padrão em torneios internacionais e ligas nacionais. Dilger e Geyer (2009) observaram que houve influência da troca do sistema de três pontos nos resultados das partidas. Ao analisarem a copa e o campeonato alemão entre as temporadas de 1985-86 à 2004-05, chegaram a conclusão de que o número de empates diminuiu significativamente, assim como a diferença de gols, indicando que o time que está na liderança tende a jogar para segurar o resultado da partida, ao invés de tentar marcar mais gols.

### 3.1.3 Interpretação do Coeficiente

O coeficiente em (3.3) assume valores entre  $(-\infty, 1]$ , com interpretação feita como:

- $\phi > 0 \iff \sigma_{2k}^2/S^2 < 1$  então  $\sigma_{2k}^2 < S^2$ : o coeficiente será positivo se, e somente se, a variância empírica for maior que a variância teórica, indicando uma variabilidade adicional induzida pelas diferentes habilidades das equipes. Quanto mais próximo de 1 for  $\phi$ , mais a habilidade é determinante nos resultados dos jogos da liga.
- $\phi = 0 \iff \sigma_{2k}^2/S^2 = 1$  então  $\sigma_{2k}^2 = S^2$ : como o coeficiente é uma medida empírica, esse valor não será exatamente zero porque  $\sigma_{2k}^2 = S^2$  não irá ocorrer, porém quanto mais próximo desse valor  $\phi$  for, menos a habilidade interfere nos resultados dos jogos, indicando que o *rank* final do campeonato é decidido aleatoriamente.
- $\phi < 0 \iff \sigma_{2k}^2/S^2 > 1$  então  $\sigma_{2k}^2 > S^2$ : valores negativos para  $\phi$  são surpreendentes, mas possíveis. Isso acontece quando a variância empírica é



menor do que a variância teórica, indicando que, ao final do campeonato, os times ficaram com pontuações muito parecidas. Aoki (2017) indica que esse fenômeno pode ocorrer quando há algum mecanismo compensatório ou conluio entre as equipes ao longo do torneio.

## 3.2 Avaliação da Significância do Coeficiente

Depois do cálculo de  $\phi$ , é necessário a construção de um intervalo de confiança utilizando simulações de Monte Carlo (MC). O procedimento, introduzido por [Barnard \(1963\)](#) baseia-se na comparação entre os dados observados e os dados simulados sob a hipótese nula. Amostras aleatórias a partir de pseudo-populações são extraídas e, em seguida, calcula-se o valor da estatística para cada replicação. Com base nestes valores, é estimado a distribuição da estatística de teste.

Suponha  $U$  como a estatística de teste com uma distribuição  $f(x)$  associada à hipótese nula  $H_0$  e  $u_0$  o valor observado de  $\phi$  na população de interesse. O vetor de valores  $u = \{u_1, u_2, \dots, u_m\}$  representa as estimativas da distribuição de  $U$  geradas sob a hipótese nula ser verdade, a partir de  $m$  replicações independentes. O valor crítico para o nível de significância  $\alpha$  no teste bilateral é definido através dos quantis amostrais  $\hat{q}_{\alpha/2}$  e  $\hat{q}_{1-\alpha/2}$ . Se  $u_0$  estiver na região crítica então rejeita-se a hipótese nula.

[Hope \(1968\)](#) cita que é preferível usar um teste conhecido de boa eficiência ao invés de um teste de Monte Carlo, porém, em certas situações, a distribuição subjacente pode ser desconhecida ou a aplicação do teste conhecido pode não ser possível devido à suposições não atendidas. Então nestes casos, o teste de MC torna-se a alternativa mais apropriada, visto que os resultados do teste são não-viesados a partir de algum valor de  $m$ .

No futebol, [Dobson e Goddard \(2003\)](#) utilizaram o teste de Monte Carlo na investigação do fenômeno “persistência” na sequência de resultados consecutivos da EPL. [Schmicker \(2013\)](#) verificou se os gols marcados de escanteio na *Major Soccer League* eram aleatoriamente distribuídos espacialmente e [Montes et al. \(2014\)](#) averiguaram a falta de competitividade no campeonato Espanhol.

## 3.3 Metodologia

Para as análises e cálculos foi utilizado o *software* livre R ([R Core Team, 2019](#)) sob a interface do [RStudio \(2018\)](#). Os pacotes utilizados foram o “*ggplot2*” ([Wickham et al., 2019](#)), para a visualização dos dados, juntamente com as extensões “*ggsci*” ([Xiao, 2018](#)) e “*lemon*” ([Edwards, 2018](#)), para as paletas de cores e configurações das legendas dos gráficos, respectivamente. Além disso, os pacotes “*plotly*” ([Sievert et al., 2019](#)) e “*tidyr*” ([Wickham e Henry, 2019](#)) também foram usados. Para todos os testes feitos, foi utilizado um nível de significância  $\alpha$  de 5%.

### 3.3.1 Banco de Dados

O calendário de jogos europeu, ao contrário do calendário brasileiro, não é todo dentro de um ano, pois normalmente os campeonatos se iniciam em agosto e se encerram em maio do outro ano, por isso as temporadas são nomeadas como, por

exemplo, “*Serie A* de 2018–19”, que equivale ao campeonato italiano disputado entre agosto de 2018 e maio de 2019. Por questões de facilidade de notação, a terminologia utilizada na apresentação das informações e resultados irá considerar somente o primeiro ano em questão. No caso do exemplo anterior, a nomenclatura ficará “*Serie A* de 2018”.

Serão consideradas as últimas 56 edições (1963-2018) dos cinco principais campeonatos da Europa. O banco de dados utilizado foi retirado do site [Football-Data \(2019\)](#) e do pacote “*engsoccerdata*”, desenvolvido por [Curley \(2016\)](#). Os dados analisados são a partir da temporada de 1963, pois foi quando o campeonato alemão iniciou-se, portanto para que todas as ligas tivessem o mesmo período investigado não foram utilizadas as temporadas anteriores dos campeonatos italiano (1929-1962), espanhol (1929-1962) inglês (1888-1962) e francês (1932-1962). Apesar de atualmente a quantidade de clubes que disputam os campeonatos serem 18 (na Bundesliga) e 20 no restante, ao longo dos anos esse número variou:

- **Campeonato Espanhol:**

Nas primeiras temporadas (1963-1970) da *La Liga*, o campeonato possuía 16 equipes entretanto, nos próximos anos (1971-1986), expandiu-se para 18. Um novo acréscimo ocorreu e entre 1987-1994 houveram uma adição de mais dois clubes, totalizando 20. Em 1995, por problemas de pagamentos à Real Federação Espanhola de Futebol, dois times (*Sevilla FC* e *Celta de Vigo*) foram rebaixados no lugar dos últimos colocados da liga, porém como eles foram readmitidos posteriormente, o número de competidores nessa temporada e na próxima (1995-1996) ficou em 22 equipes. A partir de 1997 a quantidade de clubes voltou a ser 20.

- **Campeonato Inglês:**

Entre 1963-1986 a EPL era composta por 22 clubes. Em 1987, como houveram três rebaixamentos para segunda divisão, duas promoções para a primeira e uma vaga decidida através dos *play-offs*, o número de times ficou ímpar (21). Nas próximas temporadas (1988-1990) a quantidade de competidores caiu para 20, posteriormente sendo aumentado aos já conhecidos 22 clubes (1991-1994). Em 1995 houve uma redução no total de equipes para 20, o que segue até o campeonato atual.

- **Campeonato Italiano:**

No período das 56 temporadas analisadas, a *Serie A* mudou a quantidade de equipes no campeonato somente quatro vezes. Primeiramente (1963-1966), a liga era composta por 18 clubes, sendo posteriormente diminuída, após 24 anos, para 16 (1967-1987) e voltando para os 18 times (1988-2003). O torneio é composto de 20 participantes desde a reforma do campeonato, ocorrido em 2004.

- **Campeonato Alemão:**

Inicialmente a *Bundesliga* foi composta de 16 clubes disputando o campeonato (1963-1964), porém no ano seguinte a associação de futebol alemã decidiu expandir o seu tamanho para 18 times (1965-1990). Em 1991, na sua 29<sup>a</sup> edição,

com a adição de duas equipes da antiga Alemanha Oriental, a quantidade passou a ser de 20 clubes no campeonato, o que posteriormente foi modificado aos tradicionais 18 times da liga, que estende-se até os dias atuais (1991-).

- **Campeonato Francês:**

Nas primeiras temporadas (1963-1964) eram 18 equipes, mudando posteriormente para 20 nos próximos três anos (1965-1967) e voltando para 18 nas temporadas seguintes (1968-1969). Entre 1970-1996 o número de times voltou ao padrão mais conhecido de 20 e depois (1997-2001) reduziu novamente para 18. Desde 2002 a *Ligue 1* conta com 20 clubes disputando o campeonato.

Ao longo dos anos, muitos times sofreram punições por questões de irregularidades, como por exemplo o *Chievo Verona*, que foi punido por fraudar contabilidade, ocasionando a perda de 3 pontos na *Serie A* de 2018 ou o *Arminia Bielefeld*, que teve seus pontos zerados na *Bundesliga* de 1971 por participar do escândalo de manipulação de resultados da temporada anterior. Esses pontos subtraídos dos clubes por questões extra-campo não serão considerados porque o interesse nos cálculos é o resultado da partida, então a tabela final dos campeonatos não contará com essas penalidades de irregularidades.

### 3.3.2 Cálculo do Coeficiente

Os dados utilizados para o cálculo do coeficiente de habilidade foram: time mandante, time visitante, quantidade de gols marcado pelo time mandante e quantidade de gols marcado pelo time visitante.

Com relação ao resultado final da partida, foi criada uma quinta variável, indicando se o vencedor foi o time mandante, visitante, ou se ocorreu empate. Essa variável é utilizada no cálculo das probabilidades estimadas  $\hat{P}_m$ ,  $\hat{P}_v$  e  $\hat{P}_e$  e consequentemente em  $\mu_{X_m}$ ,  $\mu_{X_v}$ ,  $\sigma_{X_m}^2$  e  $\sigma_{X_v}^2$ , necessários para o cálculo da variância teórica  $\sigma_{2k}^2$  do coeficiente  $\phi$ .

As variáveis de pontos ganhos pelo time mandante,  $X_{m_i}$ , e visitante,  $X_{v_i}$ , derivam do resultado final da partida e são necessários para o cálculo da variância empírica  $S^2$  observada ao final do campeonato. Os pontos obtidos em caso de vitória variam de acordo com a liga e ano da edição, conforme [Subseção 3.1.2](#).

Depois de todas as variáveis criadas, é feito o cálculo dos termos necessários para obtenção do valor final do coeficiente para as cinco ligas nas 56 temporadas observadas, totalizando 280 valores de  $\phi$ , conforme equação (3.3).

### 3.3.3 Teste de Monte Carlo

As simulações obtidas através de Monte Carlo são geradas supondo a hipótese nula de que a habilidade não desempenha papel nos resultados das partidas, ou seja:

$$\begin{cases} H_0 : \phi = 0 \\ H_1 : \phi \neq 0, \end{cases} \quad (3.4)$$

o que significa gerar simulações em que o resultado final do jogo é definido pela sorte. Como o coeficiente considera a vantagem do mandante no cálculo, as simulações de Monte Carlo levam em consideração  $\hat{P}_m$ ,  $\hat{P}_e$  e  $\hat{P}_v$  encontradas por temporada de cada

campeonato. Essas probabilidades são utilizadas para gerar resultados provenientes de uma distribuição multinomial para as partidas, seguindo a ordem original dos jogos. Esse procedimento é feito para cada uma das 280 edições analisadas, as 56 temporadas de cada uma das cinco ligas.

Foram utilizadas  $m = 10.000$  replicações, obtendo-se um vetor de estimativas do valor da estatística  $u = \{u_1, u_2, \dots, u_m\}$  e com o observado  $u_0 = \phi$ , o próprio valor do coeficiente encontrado na edição. Considerou-se um nível de significância  $\alpha = 0,05$ , com os quantis amostrais definidos por  $\hat{q}_{0,025}$  e  $\hat{q}_{0,975}$ , estabelecidos como o limite inferior e superior, respectivamente. Se  $\phi$  estiver dentro dos limites  $[\hat{q}_{0,025}, \hat{q}_{0,975}]$  então não rejeita-se a hipótese nula em (3.4), indicando que o coeficiente de habilidade não é significativamente diferente de zero.

### 3.3.4 Identificação das Equipes Mais ou Menos Habilidosas

Um valor de  $\phi$  significativamente diferente de zero indica que a habilidade influencia no resultado das partidas. Uma próxima questão de interesse é verificar quais são as equipes que, se retiradas do campeonato, tornariam a liga aleatória. O algoritmo para identificar esses clubes é feito através dos seguintes passos:

1. Os clubes são ordenados, em ordem decrescente, de acordo com as suas pontuações ao final da temporada. Em caso de empate no score, é considerado a ordem alfabética como desempate na ordenação das equipes;
2. O time mais distante da pontuação média será desconsiderado, seus jogos serão excluídos do campeonato e  $\phi$  será recalculado. Em caso de empate, o critério utilizado será considerar o clube que estiver em melhor posição na tabela.
3. Repete-se os passos 1 e 2 até que o coeficiente de habilidade  $\phi$  seja significativamente diferente de zero.

Apesar do maior interesse ser nas equipes com habilidades superiores aos demais, a alta discrepância entre alguns times que acabam o campeonato nas últimas posições da tabela faz com que a variabilidade na pontuação final seja maior, portanto deve-se considerar retirar os clubes com habilidades inferiores também.

## 3.4 Resultados

Os resultados apresentados nessa seção dizem respeito à análises descritivas da base de dados das ligas de futebol europeu e posteriormente os resultados encontrados associados ao coeficiente  $\phi$ .

### 3.4.1 Base de Dados

A base de dados utilizada nessa primeira parte da monografia consiste em informações de 56 (1963-2018) temporadas das principais ligas da Europa. Como o coeficiente de habilidade  $\phi$  leva em conta as probabilidades associadas ao time mandante e visitante vencer, bem como empate na partida, o cálculo de  $\hat{P}_m$ ,  $\hat{P}_v$  e  $\hat{P}_e$  se faz necessário. Pela Figura 3.1 observa-se o comportamento das probabilidades

para os resultados das partidas ao longo das temporadas nos cinco torneios.

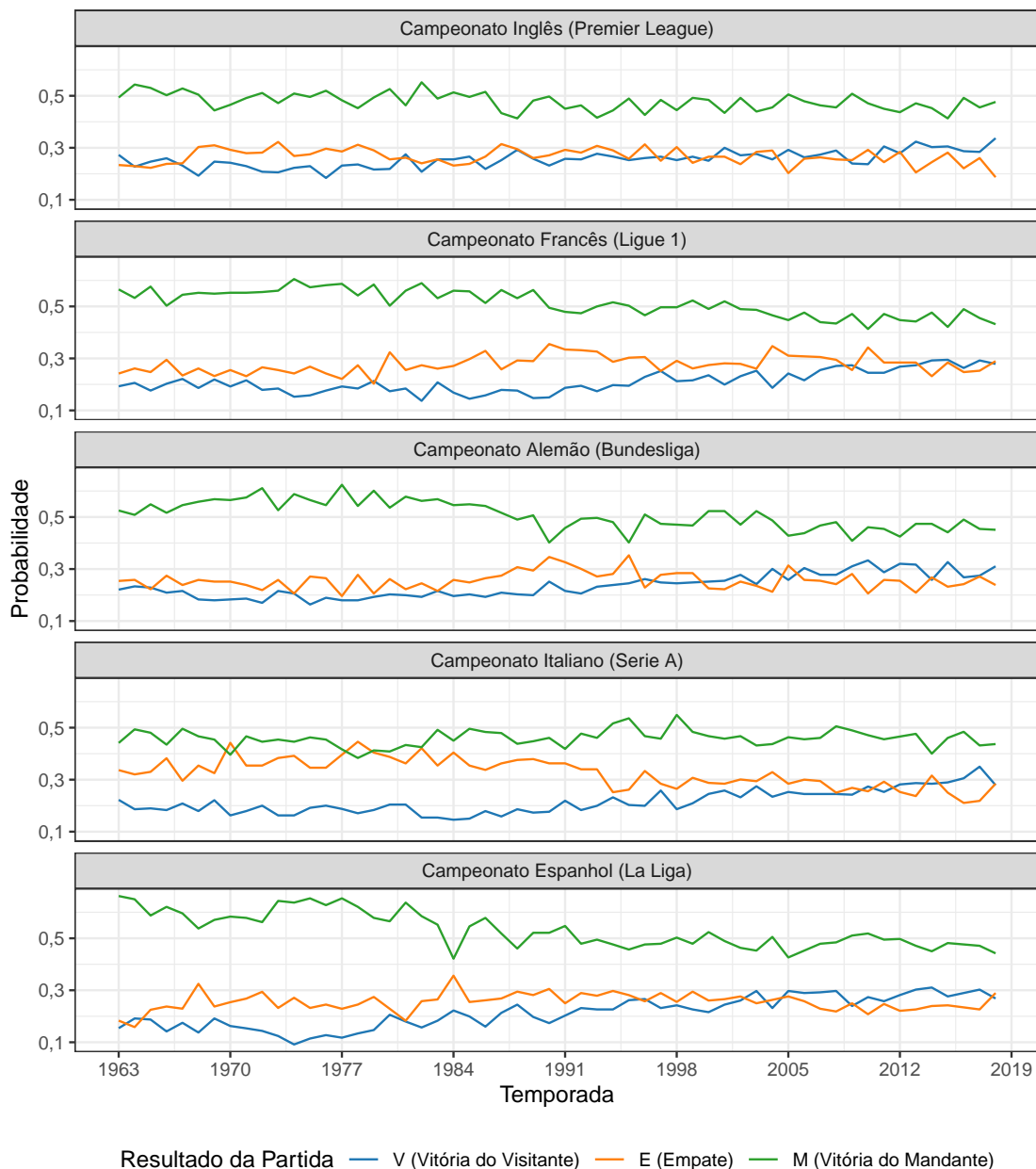


Figura 3.1: Probabilidades estimadas para os resultados das partidas nas ligas

Em geral,  $\hat{P}_m$  é maior do que  $\hat{P}_v$  e  $\hat{P}_e$ , com exceção da temporada de 1970 e 1978 do campeonato italiano, em que a probabilidade de empate (0,442 e 0,445) foi maior do que a de vitória do time mandante (0,396 e 0,383). Os campeonatos francês, alemão e espanhol possuíam grande vantagem do mandante sobre os outros resultados, porém conforme os anos foram passando, a discrepância foi diminuindo.

Se o fator mandante não influenciasse nos resultados das partidas, seria esperado que as três probabilidades fossem aproximadamente iguais ( $\approx 0,33$ ), o que sabe-se que não ocorre devido à diversos fatores intrínsecos a essa vantagem que não diz respeito às habilidades das equipes, conforme discutido na [Seção 3.1](#).

A [Tabela 3.1](#) apresenta as probabilidades de vitória do visitante, empate e vitória do mandante na primeira e última temporada analisada de cada uma das cinco ligas.

Tabela 3.1: Probabilidade dos resultados na primeira e última temporada

Resultado	Campeonato				
	Alemão	Inglês	Espanhol	Francês	Italiano
	1963-2018	1963-2018	1963-2018	1963-2018	1963-2018
Vitória do Visitante	0,221-0,310	0,273-0,337	0,154-0,268	0,193-0,279	0,222-0,279
Empate	0,254-0,239	0,234-0,187	0,183-0,290	0,242-0,289	0,337-0,284
Vitória do Mandante	0,525-0,451	0,494-0,476	0,663-0,442	0,565-0,432	0,441-0,437

Observa-se uma tendência de aumento na vitória do time visitante ao longo do tempo, assim como decréscimo da probabilidade de vitória do time mandante nos cinco campeonatos. Já em relação a quantidade de empates, a *La Liga* e a *Ligue 1* obtiveram aumento nas probabilidades, enquanto que para a *Bundesliga*, *Premier League* e *Serie A* o comportamento foi de declínio. A discordância entre as menores e maiores probabilidades para as diferentes ligas para a primeira e última temporada diminuiu: de 0,119 para 0,069 em  $\hat{P}_v$ ; de 0,154 para 0,103 em  $\hat{P}_e$  e de 0,222 em 1963 para 0,044 em 2018 para  $\hat{P}_m$ . Porém, olhar só os extremos não fornece informações sobre o comportamento ao longo de todo o período. A [Figura 3.2](#) mostra o desempenho dos três resultados possíveis comparados separadamente entre os campeonatos, para se analisar as diferenças nas probabilidades entre as cinco ligas.

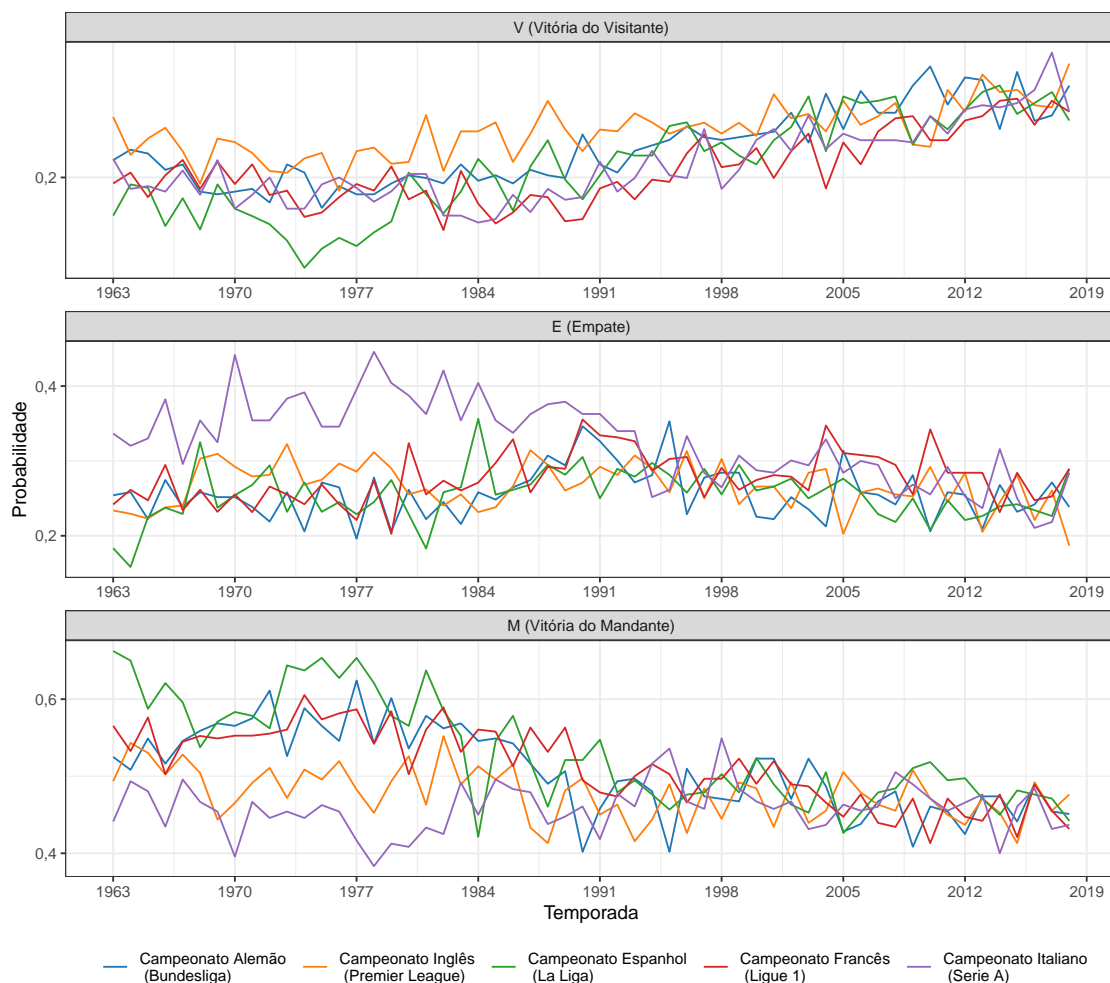


Figura 3.2: Probabilidades estimadas para os resultados das partidas

Entre as décadas de 60 e 70, a probabilidade do time visitante vencer uma partida

da *La Liga* era a menor entre as ligas. A *Serie A* possuía valores destoantes dos demais com relação às vitórias do mandante (que são menores) e empates (que são maiores e se estende até o final da década de 80). A partir dos anos 90, a diferença nas probabilidades dos torneios tornaram-se menores.

A estimativa média para as probabilidades dos possíveis resultados de uma partida estão dispostos na [Tabela 3.2](#), juntamente com o desvio padrão entre parênteses.

Tabela 3.2: Valor médio e desvio padrão das estimativas das probabilidades dos resultados das partidas das ligas

Resultado	Campeonato				
	Alemão	Inglês	Espanhol	Francês	Italiano
Vitória do Visitante	0,236 (0,045)	0,256 (0,032)	0,214 (0,059)	0,210 (0,042)	0,215 (0,046)
Empate	0,257 (0,034)	0,266 (0,031)	0,255 (0,034)	0,277 (0,034)	0,326 (0,056)
Vitória do Mandante	0,508 (0,054)	0,478 (0,033)	0,531 (0,068)	0,513 (0,050)	0,459 (0,032)

A probabilidade de vitória média das equipes que jogam no próprio estádio é menor no campeonato italiano e maior no espanhol, o que já era esperado, visto que possuem variabilidade maior quando há empate e vitória do mandante, respectivamente. A variação no campeonato inglês é pequena, independentemente do resultado da partida. Para os campeonatos francês e alemão não houve uma grande variabilidade, com ambas as ligas possuindo probabilidades similares nos três possíveis resultados, com uma diferença mais acentuada para as vitórias do time visitante.

### 3.4.2 Coeficiente de Sorte e Habilidade

O coeficiente de habilidade proposto na [Seção 3.1](#) pela [Equação 3.3](#) foi calculado para cada uma das 56 temporadas das cinco ligas europeias. Os resultados observados e diferenças entre os torneios são mostrados na [Figura 3.3](#).

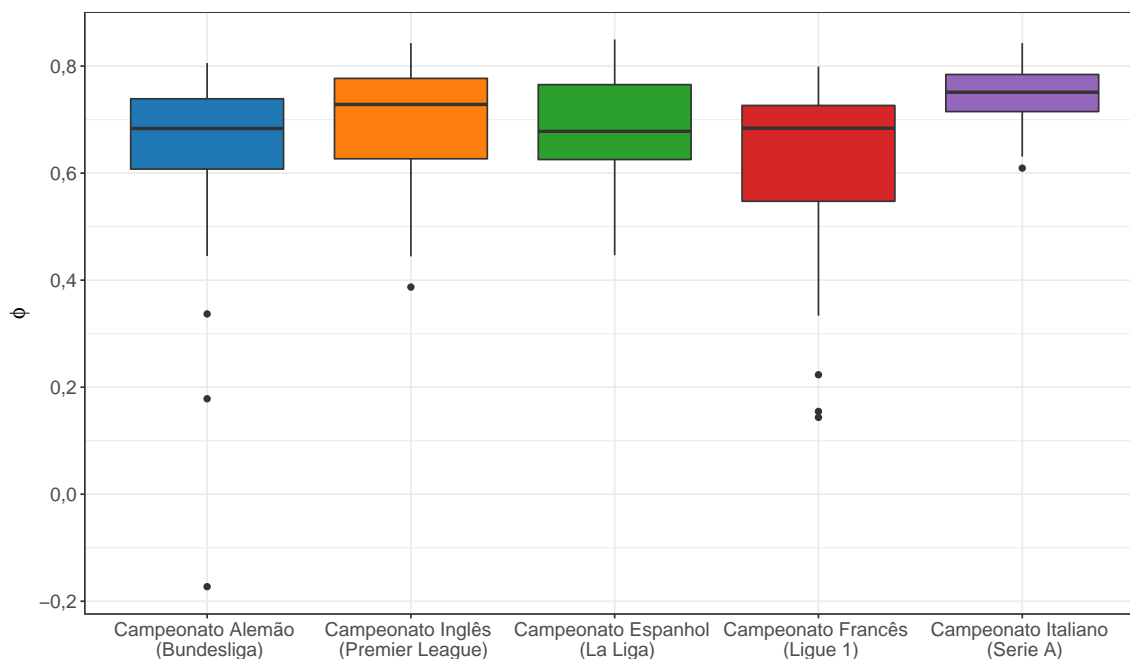


Figura 3.3: Boxplot dos valores de  $\phi$

A [Figura 3.3](#) mostra o comportamento de  $\phi$  em relação às ligas. A *Ligue 1* possui maior variabilidade para os valores do coeficiente, indicando que houveram anos em que as habilidades dos times foram mais decisivas para o resultado das partidas. A *Serie A* possui menor variabilidade entre os valores do coeficiente, com seu mínimo e máximo em  $[0,61; 0,84]$ , sendo que este primeiro, referente ao ano de 1979, é considerado um ponto *outlier* por ser um valor discrepante do conjunto de dados desse campeonato. O restante dos *outliers* encontrado nos valores de  $\phi$  são:

- *Bundesliga*: as edições de 1968, 1966 e 1964, com  $\phi$  de -0,1728, 0,1784 e 0,3369, respectivamente;
- *Premier League*: em 1992 com  $\phi = 0,3871$ ;
- *Ligue 1*: nos anos de 1963, 1964 e 1999, tendo coeficiente de 0,1435, 0,1547 e 0,2232, na ordem.

A [Figura 3.4](#) fornece os valores de  $\phi$  ao longo do tempo, para melhor visualização do coeficiente de sorte e habilidade no decorrer das temporadas.

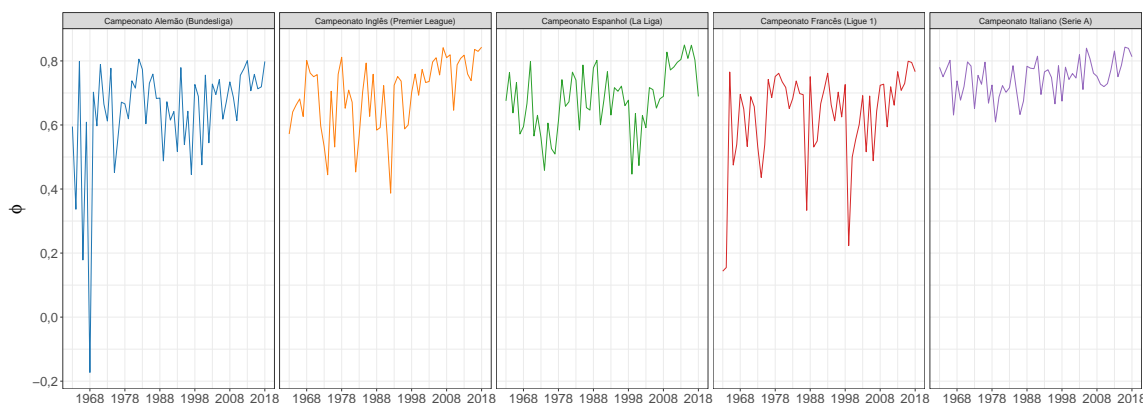


Figura 3.4: Valores de  $\phi$  ao longo das temporadas

Visualizando o coeficiente ao longo das edições através da [Figura 3.4](#), consegue-se notar mais nitidamente a variação dessa métrica dentro dos torneios nas 56 edições analisadas. Até o início da década de 1970, o valor de  $\phi$  para o campeonato alemão variou bastante, o que pode ser explicado pelo fato de que foi quando deu-se início à *Bundesliga*, então os times ainda estavam em fase de adaptação. Tanto foi que após esse período não houveram mais variações tão acentuadas quanto nos primeiros anos. A partir dos anos 2000 parece haver uma tendência crescente no coeficiente da *Premier League*, assim como no da *Ligue 1*, que ficou mais marcante na década atual, o que pode ser explicado pelo fato das equipes possuírem cada vez mais poder financeiro. No caso do campeonato francês, em 2011 o PSG (*Paris Saint-Germain*) foi comprado por um fundo de investimentos do Catar, passando a ter poder de mercado maior e podendo contratar jogadores mais caros, como no caso de Neymar Jr., foi comprado por €222 milhões em 2017 e o Mbappé, por €180 milhões em 2018. O PSG possui oito títulos nacionais, sendo que seis foram obtidos desde 2012.

Os histogramas apresentados na [Figura 3.5](#) servem para visualização de possível comportamento Normal nos dados. Utilizando o teste de normalidade de Shapiro-Wilk, observou-se que os campeonatos italiano (p-valor: 0,3017) e espanhol (p-valor:



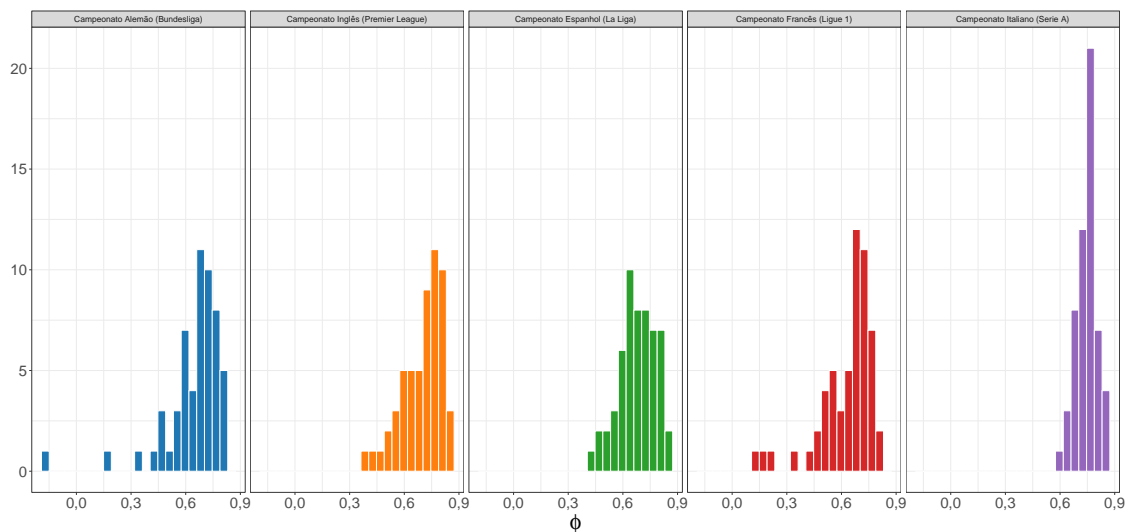


Figura 3.5: Histograma para  $\phi$  entre os campeonatos

0,2334) não rejeitam a hipótese nula de os dados serem normalmente distribuídos, enquanto que os outros rejeitaram  $H_0$ .

Todas as competições possuem um valor de coeficiente positivo, com exceção da *Bundesliga* de 1968 que obteve  $\phi \approx -0,17$ , indicando que a variância empírica observada ao final do campeonato foi menor do que a teórica. Em contrapartida, a *La Liga* de 2014 obteve o maior coeficiente,  $\phi \approx 0,85$ , onde a habilidade das equipas desempenhou papel fundamental nos resultados das partidas. A Figura 3.6 fornece a pontuação das equipas ao final dos respectivos campeonatos.

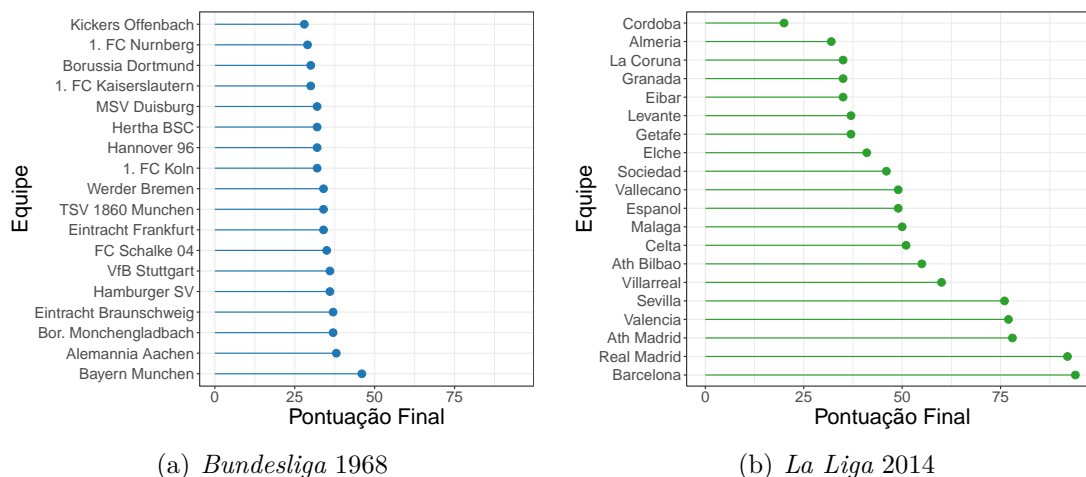


Figura 3.6: Representação gráfica das pontuações finais do campeonato alemão de 1968 e campeonato espanhol de 2014

Nota-se, tanto pela Tabela 3.3 e Tabela 3.4, quanto pela representação gráfica na Figura 3.6 que a diferença entre os pontos finais obtidos pelas equipas ao final dos campeonatos são muito diferentes. Na *Bundesliga* a variação foi pequena, já que a diferença entre o primeiro e último colocado foi de apenas 18 pontos, enquanto que na *La Liga* essa diferença foi de 74 pontos.

Tabela 3.3: Classificação final da *Bundesliga* de 1968

	<b>Equipe</b>	<b>Pontuação</b>
1	Bayern München	46
2	Alemannia Aachen	38
3	Bor. Monchengladbach	37
4	Eintracht Braunschweig	37
5	VfB Stuttgart	36
6	Hamburger SV	36
7	FC Schalke 04	35
8	Eintracht Frankfurt	34
9	Werder Bremen	34
10	TSV 1860 Munchen	34
11	Hannover 96	32
12	MSV Duisburg	32
13	1. FC Köln	32
14	Hertha BSC	32
15	1. FC Kaiserslautern	30
16	Borussia Dortmund	30
17	1. FC Nürnberg	29
18	Kickers Offenbach	28

Tabela 3.4: Classificação final da *La Liga* de 2014

	<b>Equipe</b>	<b>Pontuação</b>
1	Barcelona	94
2	Real Madrid	92
3	Athletico Madrid	78
4	Valencia	77
5	Sevilla	76
6	Villarreal	60
7	Athletico Bilbao	55
8	Celta	51
9	Málaga	50
10	Espanyol	49
11	Vallecano	49
12	Sociedad	46
13	Elche	41
14	Levante	37
15	Getafe	37
16	La Coruña	35
17	Granada	35
18	Eibar	35
19	Almería	32
20	Córdoba	20

Obviamente que a diferença entre os pontos obtidos pela vitória nos dois casos influencia no fato de que os pontos no campeonato espanhol são relativamente mai-

ores do que o do alemão, mas essa variabilidade entre todas as equipes não é afetada e por consequência o coeficiente também não é afetado, tanto que na série temporal da [Figura 3.4](#) os valores de  $\phi$  não parecem ter nenhuma modificação quando o sistema passa de 2 para 3 pontos a vitória.

Com respeito a significância dos coeficientes, observou-se que:

- A *Bundesliga* obteve 91,07% dos valores de  $\phi$  significativamente diferente de zero. As edições em que o coeficiente não rejeitou a hipótese nula são: 1964, 1966, 1968, 1975 e 1997.
- A *Premier League* obteve 98,2% dos valores de  $\phi$  significativamente diferente de zero. A edição em que o coeficiente não rejeitou a hipótese nula são foi somente a de 1992.
- A *La Liga* obteve 100% dos valores de  $\phi$  significativamente diferente de zero.
- A *Ligue 1* obteve 91,07% dos valores de  $\phi$  significativamente diferente de zero. As edições em que o coeficiente não rejeitou a hipótese nula são: 1963, 1964, 1974, 1987 e 1999.
- A *Serie A* obteve 100% dos valores de  $\phi$  significativamente diferente de zero.

Exemplos da distribuição do teste de Monte Carlo para  $\phi$  em diferentes temporadas e ligas são expressas na [Figura 3.7](#).

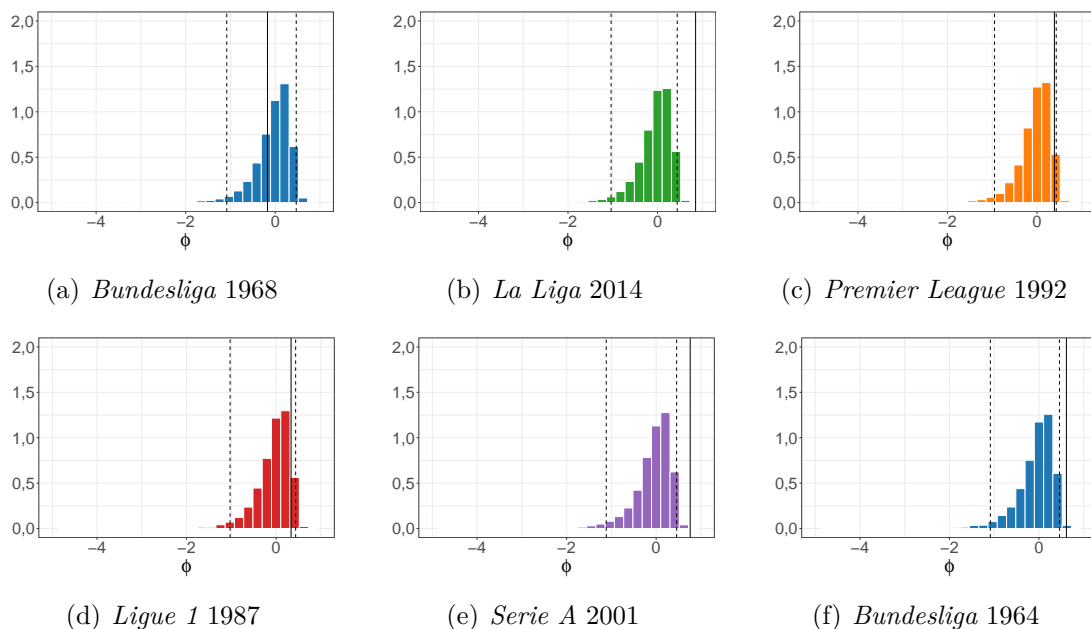


Figura 3.7: Exemplos de distribuições do Teste de Monte Carlo para  $\phi$

Como se pode observar, a [Figura 3.7\(a\)](#) e [Figura 3.7\(b\)](#) são referentes às edições de menor e maior coeficiente. Enquanto que para o primeiro caso o valor pontual de  $\phi$  (linha sólida) encontra-se dentro dos limites (traços pontilhados), para o segundo exemplo o valor da métrica está bem afastado do limite superior. Os gráficos da densidade para cada temporada de cada liga (ver Apêndice A) foram gerados para a distribuição do teste de Monte Carlo para  $\phi$ . O comportamento mostrou-se

similar, já que independentemente do campeonato e do ano das edições os gráficos mostraram-se semelhantes.

### 3.4.3 Equipes Destoantes

Na [Subseção 3.3.4](#) é apresentado um algoritmo para identificação de quais equipes destoam o campeonato. No total foram 5341 times disputando as 56 temporadas das cinco ligas. Dentre estes, foi preciso retirar 941 (17,62%) equipes para todas as edições ficarem aleatórias. Como entre uma temporada e outra muda-se poucos clubes (sai os rebaixados e entram os que foram “promovidos”), então nesse valor total considera-se que alguns times estão sendo representados mais de uma vez.

A [Figura 3.8](#) apresenta o percentual de equipes que foram retiradas para que a temporada se torne aleatória. Como critério para distinguir se o clube terminou o campeonato nas primeiras ou últimas colocações, a tabela foi dividida pela metade: se o clube está na metade de cima da tabela então está entre os primeiros colocados, caso contrário entre os últimos.

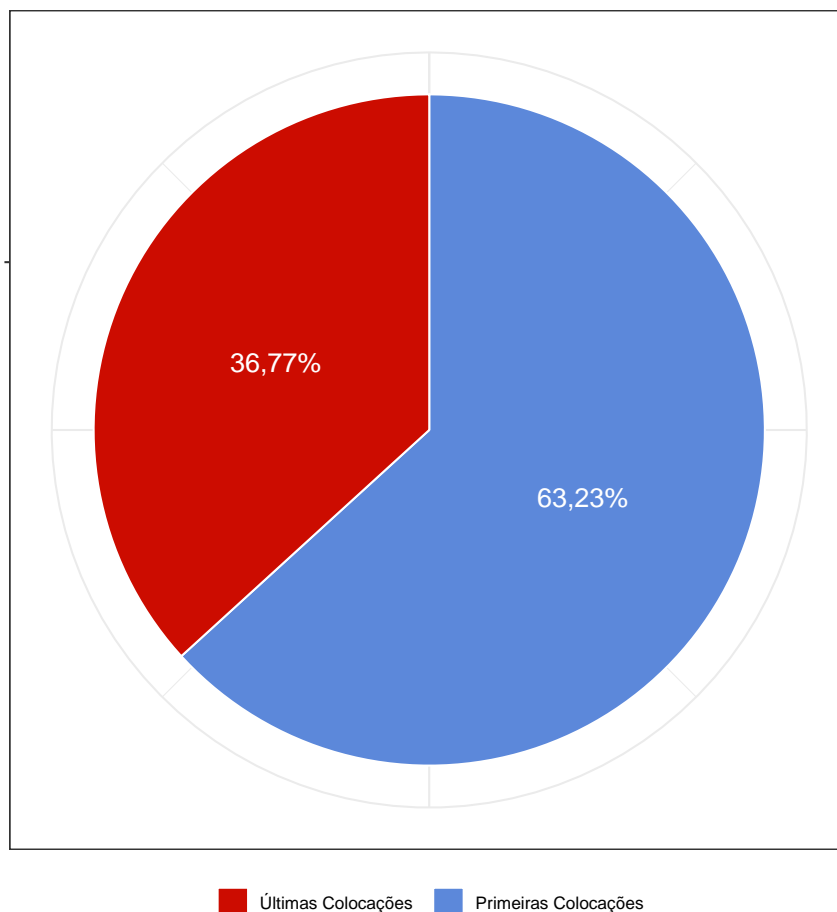


Figura 3.8: Percentual geral de todos os times que foram retirados para as temporadas se tornarem aleatórias

A [Figura 3.8](#) mostra que 595 (63,23%) eram clubes que se encontravam na parte de cima da tabela e 346 (36,77%) na parte de baixo, indicando que a maioria dos times que foram retirados do campeonato para este não depender do fator habilidade nos resultados dos jogos são devido a uma habilidade superior em relação aos demais,

já que equipes que se encontram na parte de cima da classificação são as que pontuam mais, portanto melhores que as demais. Ao levar em consideração as diferentes ligas, tem-se a [Figura 3.9](#), que apresenta o percentual de retiradas de equipes que estavam nas primeiras ou últimas colocações dos campeonatos.

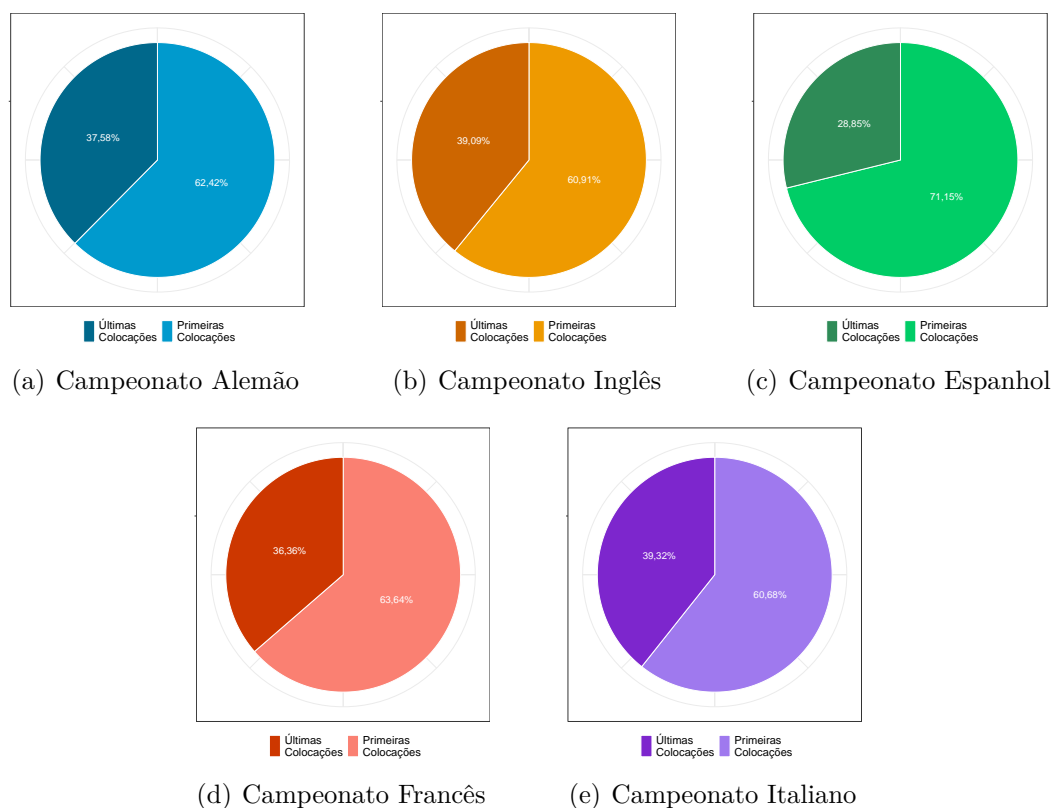


Figura 3.9: Percentual de times que foram retirados para a temporada se tornar aleatória dentro de cada campeonato

Pela [Figura 3.9](#) torna-se evidente que a *La Liga* é o campeonato em que há uma porcentagem maior de retirada de clubes em posições superiores na tabela (71,15%), o que condiz com o fato de que os times mais famosos do país possuem poder de compra muito grande, portanto investem no plantel, por consequência são vencedores e suas habilidades superiores são mais evidentes do que a falta de qualidade de equipes menores que disputam a liga.

A [Tabela 3.5](#) traz informações referente a quantidade de vezes que a equipe com pontuação mais afastada da média foi a campeã ou lanterna (último colocado) de determinada edição.

Tabela 3.5: Número de vezes que o time mais destoante foi o campeão ou lanterna

Posição	Campeonato				
	Alemão	Inglês	Espanhol	Francês	Italiano
	N = 51 n (%)	N = 55 n (%)	N = 56 n (%)	N = 51 n (%)	N = 56 n (%)
Campeão	38 (74,51)	41 (74,55)	46 (82,14)	40 (78,43)	41 (73,21)
Lanterna	13 (23,59)	14 (25,45)	10 (17,86)	11 (21,57)	15 (26,79)

Quando retira-se de cada edição somente o primeiro clube que possui a maior

distância para a média da distribuição final dos pontos ao final do campeonato, então o time foi o campeão, ou o lanterna da competição no seu respectivo ano. Claramente isso não significa que estes foram os únicos retirados para que a temporada ficasse aleatória, porém serve para analisar quais equipes foram as mais destoantes dentro de determinada liga e ano das competições.

Novamente o campeonato espanhol é a liga com maior porcentagem (82,14%), o que significa que os clubes mais vencedores possuem habilidades superiores aos demais. No caso da Itália e Inglaterra, ambas as competições possuíam inicialmente clubes muito fracos que pontuavam pouco, pois suas habilidades eram inferiores as das demais. Na *Serie A* entre as décadas de 60 e 90 foram 14 equipes que tiveram o menor rendimento do campeonato e foram as primeiras a serem removidas. A partir dos anos 2000, somente o rebaixado *Venezia* foi retirado por ser o pior.

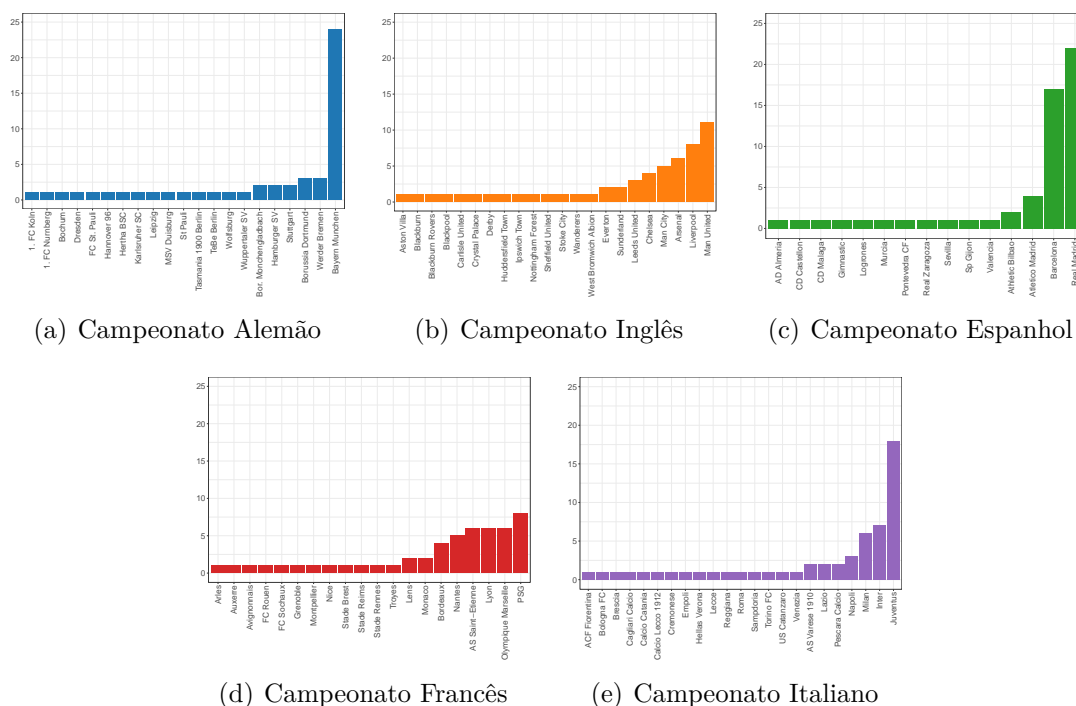


Figura 3.10: Clubes mais destoantes dentro de cada campeonato

Na *Bundesliga*, observa-se uma forte presença do *Bayern München* (conhecido como Bayern de Munique no Brasil), que possui quantidade de aparições seis vezes maior que a segunda e terceira equipe que mais aparece, *Borussia Dortmund* e *Werder Bremen*. De acordo com dados da *Deloitte Football Money League* (DFML, 2019), o *Bayern* possui a 4<sup>a</sup> maior receita do futebol mundial (€629,2 milhões), quase o dobro do *Dortmund*, segundo clube mais rico da Alemanha, que possui receita de €317.2 milhões.

Na *Premier League* não há hegemonia de um clube específico, o que já era esperado visto que os times da Inglaterra alternam a liderança de forma mais rotativa que outras ligas. Normalmente os times que possuem uma contagem superior à 1 são aqueles que foram campeões mais de uma vez, com exceção do *Sunderland* que aparece como o time mais conflitante, nas temporadas de 2002 e 2005, como lanterna nas edições. Ou seja, ele ficou em último lugar em ambas as vezes e divergia quanto as habilidades por ser o mais fraco entre todos, possuindo uma pontuação ao final

do campeonato muito baixa.

O resultado da *La Liga* indica que o *Barcelona* e *Real Madrid* juntos, interferem como sendo as equipes mais destoantes em 69,6% das edições. Isso mostra a grande discrepância dessas duas equipes, se comparado com o restante do campeonato espanhol.

Já na *Ligue 1*, assim como na EPL, não há uma presença forte de um time específico. Conforme levantado na [Subseção 3.4.2](#), o PSG possui oito títulos nacionais, sendo que em todas as ocasiões em que foi campeão, também foi a primeira equipe a ser retirada dos campeonatos, pois era o clube mais distante da média de pontos corridos.

Pelos dados da *Serie A*, não é nenhuma surpresa a *Juventus* ser a equipe com maior número de retiradas da [Figura 3.10\(e\)](#). Atual octacampeã, foi o clube com maior divergência entre a pontuação final e a média de pontuação do campeonato nessas últimas temporadas em que se consagrou campeã. Na edição de 2013, por exemplo, obteve 102 pontos ao final da edição, cuja média foi de 52,2, uma diferença de 49,5 pontos.

A [Figura 3.11](#) apresenta a quantidade de equipes que foram retiradas de acordo com as posições da tabela.

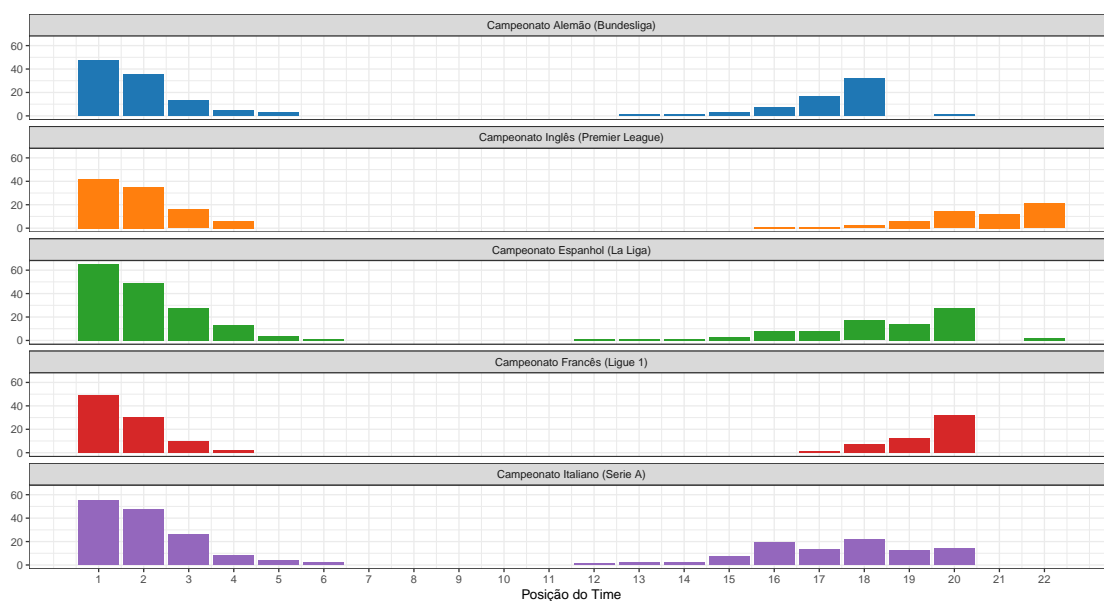


Figura 3.11: Quantidade de clubes retirados por posição

Ao analisar todas as retiradas, primeiro observa-se que os clubes que acabam o campeonato no meio da tabela não são retirados para a temporada ficar aleatória, o que já era esperado, visto que se há discrepâncias entre as habilidades das equipes de um campeonato, espera-se que sejam por clubes que são mais ou menos habilidosos que os demais, o que por consequência acabariam terminando a temporada nos extremos da tabela. Como a *Ligue 1* possuía, em média, o valor de  $\phi$  mais baixo, então em comparação com as outras ligas é o que precisa da retirada de menos times. A diferença na frequência das últimas posições a partir da posição 18 deve-se ao fato de haver edições com números diferentes de equipes, já que ao longo dos anos o número total de clubes variou.

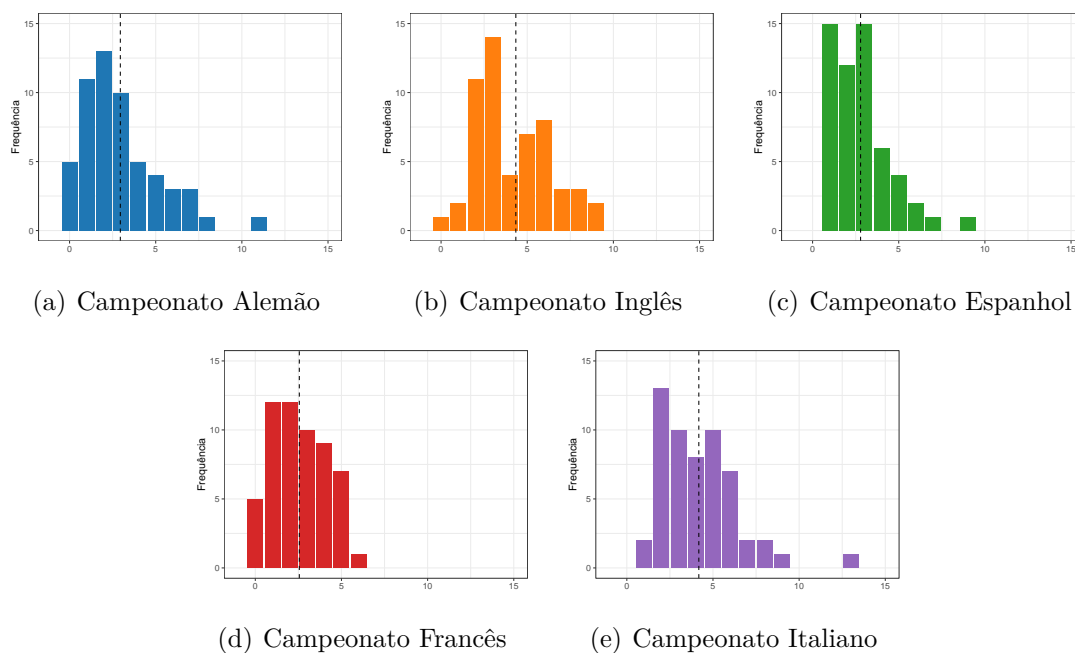


Figura 3.12: Quantidade de retiradas dentro de cada campeonato

Na [Figura 3.12](#), observa-se que apesar de uma média de 2,95 (linha vertical pontilhada) retiradas, na temporada de 2008 da *Bundesliga* foi-se necessário a retirada de onze clubes para a edição ficar aleatória: os cinco primeiros colocados e os seis últimos, o que equivale à 61,1% da quantidade total de equipes. A [Tabela 3.6](#) apresenta a classificação final dessa temporada.

Tabela 3.6: Classificação final da *Bundesliga* de 2008

	<b>Equipe</b>	<b>Pontuação</b>
1	Wolfsburg	69
2	Bayern München	67
3	Stuttgart	64
4	Hertha	63
5	Hamburger	61
6	Dortmund	59
7	Hoffenheim	55
8	Schalke 04	50
9	Leverkusen	49
10	Werder Bremen	45
11	Hannover	40
12	FC Koln	39
13	Ein Frankfurt	33
14	Bochum	32
15	M'gladbach	31
16	Cottbus	30
17	Karlsruhe	29
18	Bielefeld	28

Na [Tabela 3.6](#), os times pintados de cinza são os que foram retirados pelo al-



goritmo para o coeficiente de habilidade tornar-se não significativo. A média do campeonato foi de 47 pontos e, apesar dessa edição ter sido acirrada, muitas equipes foram retiradas para a aleatorização da temporada, pois como os pontos entre os clubes foram parecidos entre as posições, a remoção de um clube não modificava muito o valor do coeficiente.

Na *Premier League* e *Serie A* precisa-se retirar, em média 4,34 e 4,18, clubes em cada edição para o torneio tornar-se aleatório, respectivamente. Na temporada de 1966 da liga italiana foram retirados 72,2% das equipes e em 2013 na EPL foram 75%, sendo que os primeiros quatro times estavam na parte de cima da tabela: *Manchester City*, *Liverpool*, *Chelsea* e *Arsenal*.

Em ambas as temporadas ocorreu o mesmo fenômeno da [Tabela 3.6](#): as pontuações ao longo da classificação final não possuem discrepâncias altas entre os clubes, mas os primeiros e últimos dessas edições foram muito bem e muito mal, respectivamente. No caso da liga italiana por exemplo, o campeão venceu 18 partidas e perdeu apenas 3, enquanto que o lanterna obteve 3 vitórias e 20 derrotas, das 34. Na temporada de 2015 da *Premier League*, o *Aston Villa* obteve 17 pontos em 38 rodadas (3 vitórias, 8 empates e 27 derrotas), com aproveitamento de 15%, ficando 34,65 pontos abaixo da média (51,65) do total de pontos obtidos pela equipe ao final do campeonato.

O campeonato francês, em geral, precisa da exclusão de poucos clubes porque já possui um coeficiente menor, em comparação às outras ligas, conforme a [Figura 3.3](#). Então são poucos os times que destoam o campeonato, sejam por suas habilidades superiores ou inferiores aos demais.

Em termos de proporção da quantidade total de clubes que participaram das 56 edições e que em algum momento foram retirados do campeonato para aleatorização da temporada, no campeonato espanhol 62,96% das equipes participantes tiveram que sair em pelo menos uma edição, a menor proporção dentre as ligas. Em seguida vem o campeonato alemão com 66,13%, o francês com 72,88%, inglês com 84,21% e o italiano com 91,37%, conforme [Tabela 3.7](#).

Tabela 3.7: Quantidade de diferentes clubes excluídos em cada edição das ligas

	Campeonato				
	Alemão	Inglês	Espanhol	Francês	Italiano
<b>Excluídos</b>	41	48	34	43	52
<b>Total</b>	62	57	54	59	58

No caso da *Serie A*, esperava-se que a porcentagem fosse a maior. Somente em 10 edições do italiano o lanterna do campeonato não foi retirado para aleatorização, sendo que dentre esses 10 clubes, todos foram excluídos em alguma outra edição. Ou seja, em algum momento, com exceção de seis equipes, todo o restante participou, de algum modo (seja pela habilidade superior ou inferior) para o desequilíbrio da principal divisão de futebol da Itália.

Espera-se que, ao observar todas as equipes retiradas para aleatoriedade do campeonato, que o comportamento observado na [Figura 3.10](#) se repita. De fato, é isso que acontece. Em todos os cinco campeonatos, os clubes com maiores porcentagens de saídas são os que terminaram como campeões das edições e são os mais conhecidos nacionalmente e mundialmente.

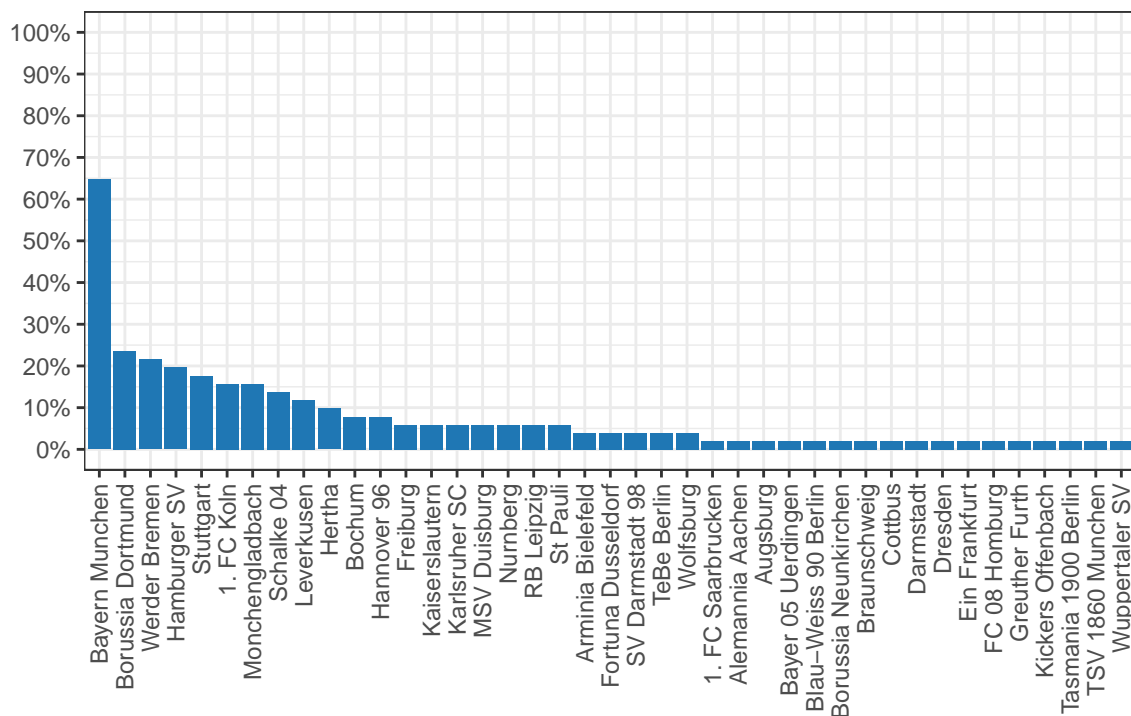


Figura 3.13: Porcentagem de saídas de cada equipe da Bundesliga

No campeonato alemão, na [Figura 3.13](#), nota-se que o *Bayern München* saiu em 64,7% das edições. Em seguida o *Dortmund* com 23,5% de saídas e o *Werder Bremen* com 21,6%, porcentagem quase três vezes menor que o do *Bayern*, o que indica a hegemonia do mesmo perante o restante dos competidores da *Bundesliga*.

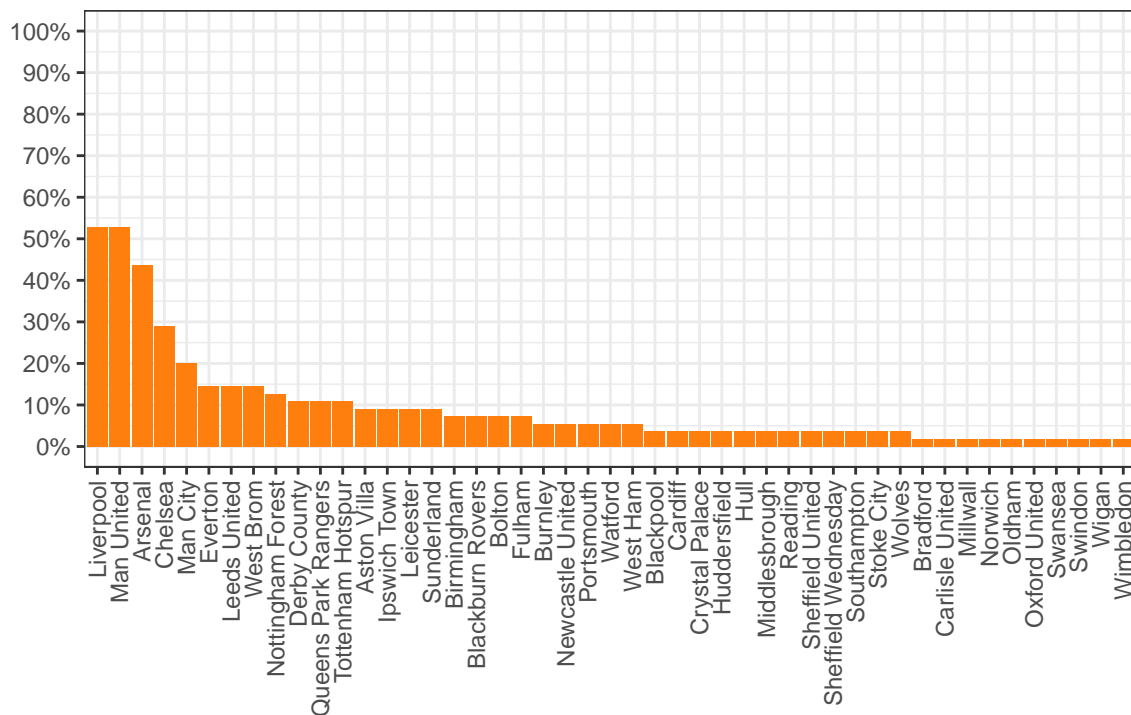


Figura 3.14: Porcentagem de saídas de cada equipe da Premier League

No caso da *English Premier League*, a porcentagem de exclusão não está em

discrepância para um único clube, porém evidentemente as equipes mais valiosas possuem a maior retirada, já que geralmente dinheiro equivale à títulos ou melhores posições na tabela. O *Liverpool* e o *Manchester United*, os maiores campeões do torneio, saíram em 56,9% das temporadas. Em seguida está o *Arsenal*, com 47,1%.

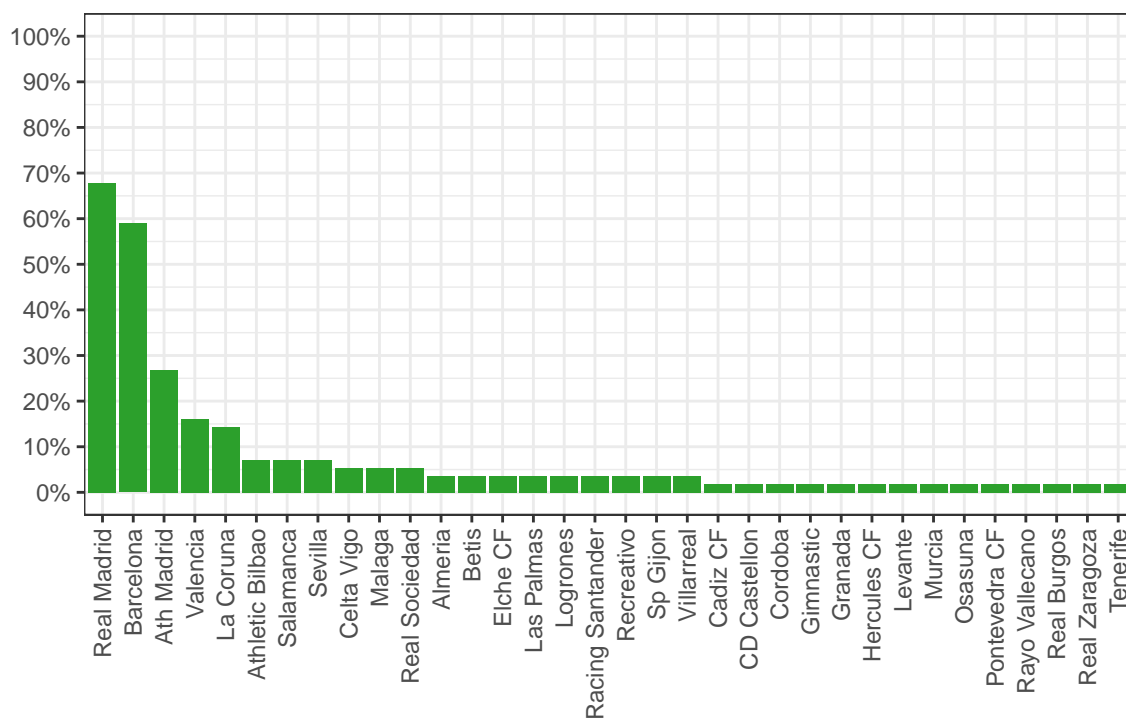


Figura 3.15: Porcentagem de saídas de cada equipe da La Liga

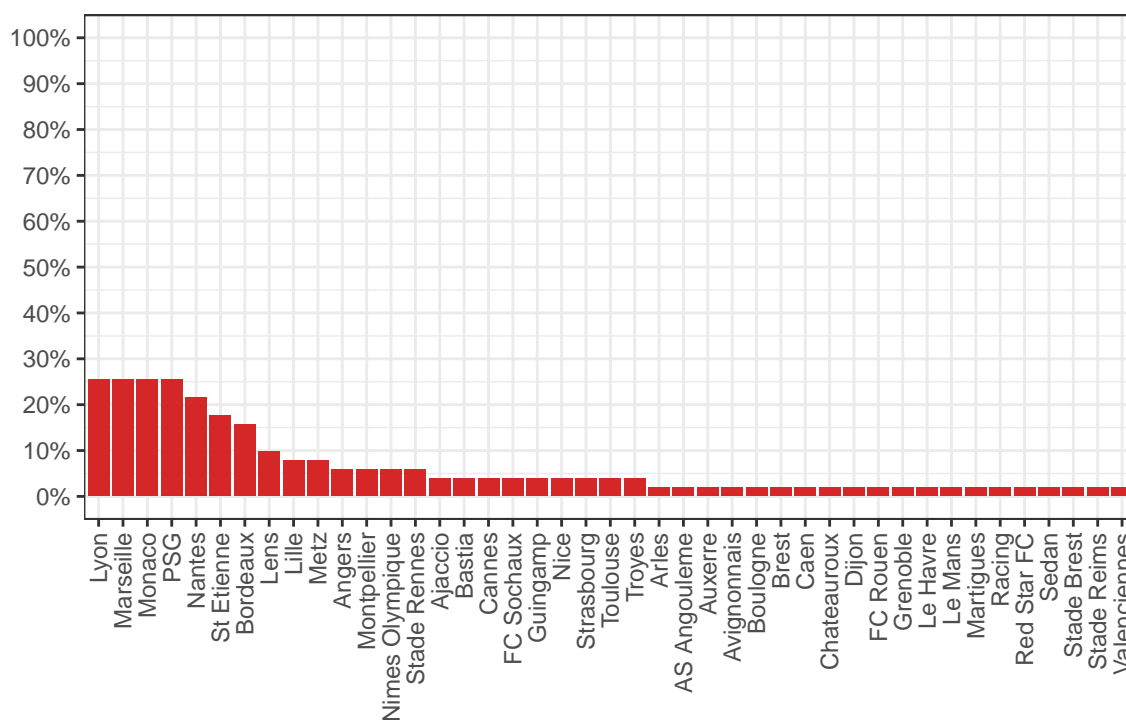


Figura 3.16: Porcentagem de saídas de cada equipe da Ligue 1

Não é surpresa o que é visto na [Figura 3.15](#). O *Real Madrid* e *Barcelona* sempre foram grandes clubes mundialmente e, obviamente isso reflete no desempenho doméstico ao longo das competições. Enquanto o time merengue saiu em 74,5% e o Barça em 64,7%, o Atlético de Madrid, terceira equipe com maior número de torcedores da Espanha, precisou ser excluído em apenas 29,4% das edições.

Na *Ligue 1*, houve um empate entre quatro clubes: *Lyon*, *Olympique de Marseille*, *Monaco* e *PSG*, cuja porcentagem foi de 25,5% então, ao contrário das outras ligas, não parece haver uma hegemonia entre uma ou duas equipes no campeonato francês, o que já era esperado visto os resultados anteriores.

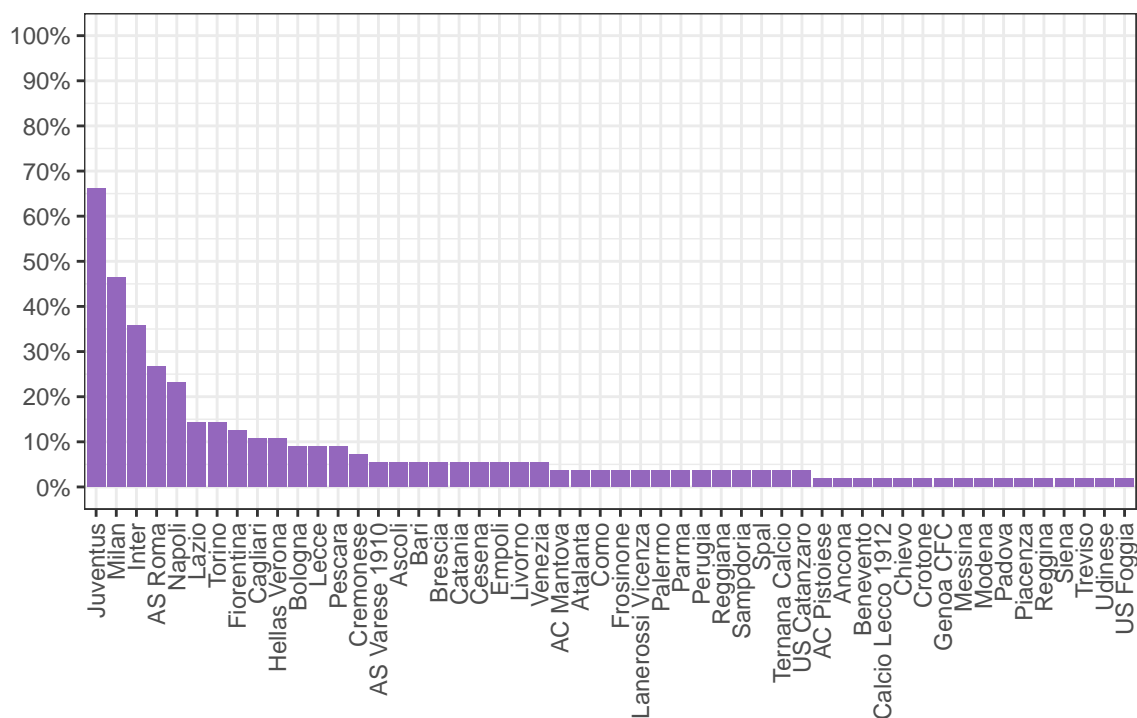


Figura 3.17: Porcentagem de saídas de cada equipe da Serie A

A maior campeã do *scudetto* teve que ser excluída de 72,55% das edições, porcentagem menor somente que a do *Real Madrid* na *La Liga*. O *Milan* e a *Inter*, dois grandes clubes, ficaram em 2º e 3º lugar na ordem de maiores saídas com 50,98% e 39,22%.

## 4 Habilidade das Equipes

Em análises esportivas é comum o interesse em predições acerca dos resultados das partidas. Na literatura há uma vasta lista de diferentes técnicas para prever o desfecho do jogo. [Goddard \(2005\)](#) utilizou duas abordagens diferentes: uma regressão de Poisson bivariada para estimar modelos de previsão para gols marcados e sofridos; e uma regressão ordinal, baseando-se em vitória, empate e derrota do time nos modelos de previsão. O que [Knorr-Held \(2000\)](#) também considerou no seu método, utilizando parâmetros latentes para representar a força de cada equipe. [Schumaker et al. \(2016\)](#) analisaram *tweets* relacionados às vinte equipes que disputavam a EPL utilizando análise de sentimentos para tentar prever os jogos.

Apesar da incerteza acerca do resultado da partida ser um atrativo em esportes como o futebol, [Buraimo e Simmons \(2015\)](#) discorrem acerca do fato de que essa importância não é mais relevante à audiência do público, que evoluiu para preferências que dizem respeito ao talento dos jogadores em campo. Apesar disso, com o aumento na competitividade nos mercados de apostas, modelos preditivos de classificações para previsão de resultado da partida estão sendo bastante usados, como por exemplo o método Elo ([Hvattum e Arntzen, 2010](#)). Em [Constantinou e Fenton \(2013\)](#) foi proposto uma abordagem bayesiana para predição chamada de classificação- $\pi$ , uma forma dinâmica de classificar as equipes com base nas discrepâncias relativas nas pontuações das partidas.

O modelo de Bradley-Terry ([Bradley e Terry, 1952](#)) é um modelo para prever o resultado de uma comparação pareada. No caso dos esportes, é bastante usado para modelar a probabilidade dos possíveis resultados de uma partida entre duas equipes, estimando as habilidades dos times, por exemplo ([Hunter, 2004](#)). No futebol, a aplicação desse método já foi utilizada, também de forma adaptada, por diversos autores. [Tutz e Schauburger \(2015\)](#) utilizaram de técnicas de penalização e regularização para identificar grupos de times que possuem a mesma habilidade e também contribuições das covariáveis, como por exemplo despesas, no sucesso das equipes. Em [Ley et al. \(2019\)](#) é feito uma comparação de vários modelos, entre eles o de Bradley-Terry, para criação de um novo método de ranqueamento. [Menke e Martinez \(2008\)](#) reparametrizaram o modelo de Bradley-Terry como uma única camada de uma rede neural artificial, com extensões sendo apresentadas que podem ser usadas para prever os resultados de competições complexas e desiguais.

Diferentes aplicações do modelo em outros esportes são encontrados na literatura, como no caso do futebol americano ([Yan et al., 2012](#); [Pelechrinis e Papalexakis, 2016](#)), tênis ([McHale e Morton, 2011](#)), beisebol ([Sire e Redner, 2009](#); [Phelan e Whelan, 2017](#)), entre outros.

Para este trabalho, a estimação da habilidade das equipes é feita através de um ajuste com base nos resultados de todas as partidas de uma temporada, utilizando o modelo de [Bradley e Terry \(1952\)](#), bem como a versão dinâmica desenvolvida por [Cattelan et al. \(2013\)](#), que modela a evolução temporal das habilidades das equipes em termos de jogos dentro e fora de casa utilizando médias móveis exponencialmente ponderadas (EWMA – *Exponentially Weighted Moving Average*).

O objetivo deste capítulo é estimar as habilidades das equipes que compõem as cinco principais ligas nacionais da Europa por meio dos dois modelos estatísticos.

## 4.1 O Modelo de Bradley-Terry Para Predição de Resultados de Partidas de Futebol

Suponha um campeonato com  $k + 1$  equipes e  $N = k \times (k + 1)$  jogos. Seja  $Y_i$  uma variável aleatória binária que denota o resultado da  $i$ -ésima partida que é disputada entre o time mandante  $m_i$  contra o time visitante  $v_i$ , com  $i = 1, \dots, N$ ,  $m_i, v_i = 1, \dots, k + 1$  e  $m_i \neq v_i$ , então:

$$Y_i = \begin{cases} 1, & \text{se time mandante vence a partida } i \\ 0, & \text{se time visitante vence a partida } i. \end{cases}$$

Especifique a probabilidade do resultado como  $\mathbb{P}(Y_i = 1) = F(\alpha_{m_i} - \alpha_{v_i})$ , definindo  $F$  como uma função de distribuição acumulada logística.

Como discutido anteriormente na [Subseção 3.1.1](#), existe uma vantagem relacionada à equipe que joga em casa, o que não deve ser ignorado. Esse efeito é levado em consideração ao ser incluído um parâmetro  $\eta$  para todos os times ([Fahrmeir e Tutz, 1994](#)):

$$\mathbb{P}(Y_i = 1) = \frac{\exp(\eta + \alpha_{m_i} - \alpha_{v_i})}{1 + \exp(\eta + \alpha_{m_i} - \alpha_{v_i})}, \quad (4.1)$$

com uma restrição no conjunto de habilidades para identificabilidade de modo que

$$\sum_{j=1}^{k+1} \alpha_j = 0.$$

No caso da equação (4.1) o modelo não considera como possível resultado da partida o empate. A definição da variável  $Y_i$  é binária somente em termos de vitória da equipe mandante ou vitória do visitante. É então adicionada outra categoria em  $Y_i$ , com a seguinte codificação:

$$Y_i = \begin{cases} 2, & \text{se time mandante vence a partida } i \\ 1, & \text{se ocorre um empate entre os times} \\ 0, & \text{se time visitante vence a partida } i \end{cases}$$

Assim, estendendo a equação (4.1) para considerar empates nas partidas ([Agresti, 2002](#)):

$$\mathbb{P}(Y_i \leq y_i) = \frac{\exp(\delta_{y_i} + \eta + \alpha_{m_i} - \alpha_{v_i})}{1 + \exp(\delta_{y_i} + \eta + \alpha_{m_i} - \alpha_{v_i})}, \quad (4.2)$$

com  $y_i \in \{0, 1, 2\}$  e  $-\infty < \delta_0 < \delta_1 < \delta_2 = \infty$  são parâmetros de ponto de corte. Para a identificabilidade é necessário impor uma restrição simétrica  $\delta_0 = -\delta$  e  $\delta_1 = \delta$ , com  $\delta \geq 0$ . Essa restrição garante que quando duas equipes de mesma habilidade se enfrentam em território neutro, ambas tem a mesma probabilidade de vencer a partida (Cattelan et al., 2013).

A limitação desse modelo é que as habilidades estimadas para os clubes são definidas como sendo o mesmo valor ao longo de toda a temporada, o que não reflete à realidade.

## 4.2 O Modelo de Bradley-Terry Dinâmico

Cattelan et al. (2013) propuseram uma versão do modelo de Bradley-Terry que modela os resultados das partidas de forma dinâmica:

$$\mathbb{P}(Y_i \leq y_i | Y_{i-1} = y_{i-1}, \dots, Y_1 = y_1) = \frac{\exp(\delta_{y_i} + \alpha_{m_i}(t_i) - \alpha_{v_i}(t_i))}{1 + \exp(\delta_{y_i} + \alpha_{m_i}(t_i) - \alpha_{v_i}(t_i))}, \quad (4.3)$$

onde  $\alpha_{m_i}(t_i)$  descreve a habilidade do time mandante  $m_i$  na partida  $i$  disputada contra o time visitante  $v_i$ .

Os autores consideram que a evolução na habilidade do time que joga em casa depende somente do desempenho anterior nas partidas da equipe como mandante e o mesmo para o visitante. A habilidade do clube  $m_i$  na partida  $i$  e no tempo  $t_i$  assume evoluir no tempo seguindo um processo de Médias Móveis Exponencialmente Ponderadas (EWMA). O raciocínio segue o mesmo para a equipe  $v_i$ , portanto:

$$\alpha_{m_i}(t_i) = \lambda_1 \mu_{m_i}(t_i) + (1 - \lambda_1) \alpha_{m_i}(t_i^{(-1)}) \quad (4.4)$$

$$\alpha_{v_i}(t_i) = \lambda_2 \mu_{v_i}(t_i) + (1 - \lambda_2) \alpha_{v_i}(t_i^{(-1)}), \quad (4.5)$$

onde  $t_i^{(-1)}$  representa o tempo anterior de  $t_i$  em que o clube jogou dentro (4.4) ou fora (4.5) de casa. Já  $\lambda_1, \lambda_2 \in [0, 1]$  são os parâmetros de suavização específicos do mandante e visitante, respectivamente; que retrata a persistência da dependência em jogos anteriores como mandante,  $\lambda_1$ , e visitante,  $\lambda_2$ .

A habilidade média da equipe considerando a partida  $i$  mais próxima em que a equipe  $m_i$  foi mandante é denotada pelo termo  $\mu_{m_i}(t_i)$ , enquanto que  $\mu_{v_i}(t_i)$  segue a mesma ideia, mas para o time visitante, O cálculo é obtido por:

$$\mu_{m_i}(t_i) = \beta_1 r_{m_i}(t_i^{(-1)}) \quad (4.6)$$

$$\mu_{v_i}(t_i) = \beta_2 r_{v_i}(t_i^{(-1)}), \quad (4.7)$$

com  $\beta_1$  e  $\beta_2$  sendo os parâmetros específicos do mandante e visitante, enquanto que  $r_{m_i}(t_i^{(-1)})$  e  $r_{v_i}(t_i^{(-1)})$  são variáveis medindo o resultado do time  $m_i$  e  $v_i$  na partida disputada no tempo  $t_i^{(-1)}$ . No caso do futebol, é o número de pontos recebidos pela equipe na partida anterior em que jogou em casa, ou fora de casa ao considerar  $v_i$ .

Considere  $Q$  a quantidade de jogos de  $m_i$  anterior a partida  $i$  no tempo  $t_i$  em que o time jogou como mandante e para  $v_i$  as partidas anteriores como visitante. As fórmulas baseadas nos pares de equações [(4.4), (4.6)] e [(4.5), (4.7)] são:

$$\alpha_{m_i}(t_i) = \beta_1 \left\{ \lambda_1 \sum_{q=0}^{Q-1} (1 - \lambda_1)^q r_{m_i}(t_i^{(-q-1)}) + (1 - \lambda_1)^Q \bar{r}_m \right\} = \beta_1 x_{m_i}(t_i; \lambda_1) \quad (4.8)$$

$$\alpha_{v_i}(t_i) = \beta_2 \left\{ \lambda_2 \sum_{q=0}^{Q-1} (1 - \lambda_2)^q r_{v_i}(t_i^{(-q-1)}) + (1 - \lambda_2)^Q \bar{r}_v \right\} = \beta_2 x_{v_i}(t_i; \lambda_2) \quad (4.9)$$

As habilidades são funções de todo o passado do time, de acordo com os resultados das partidas em casa para  $m_i$  e fora de casa para  $v_i$ . As covariáveis  $x_{m_i}(t_i; \lambda_1)$  e  $x_{v_i}(t_i; \lambda_2)$  são médias ponderadas desses resultados anteriores com pesos  $\lambda_s, \lambda_s(1 - \lambda_s), \dots, \lambda_s(1 - \lambda_s)^Q$  e  $s = \{1, 2\}$ , que decresce geometricamente para zero. Para os valores extremos dos parâmetros, se  $\lambda_s = 1$ , então a habilidade do time mandante/visitante depende apenas da partida anterior em casa/fora de casa ((4.6) e (4.7)). Se  $\lambda_s = 0$ , a habilidade é constante no tempo e igual para todos os times (parâmetro  $\eta$  em (4.1)).

O modelo precisa de uma condição inicial. Assumindo que todas as equipes começam o campeonato com as mesmas habilidades como mandante e visitante igual a  $\beta_1 \bar{r}_m$  e  $\beta_2 \bar{r}_v$ , respectivamente, onde  $\bar{r}_m$  é a quantidade média de pontos obtidos pelos times que jogaram em casa durante a temporada anterior. O raciocínio segue o mesmo para  $\bar{r}_v$ . Portanto, no modelo de Bradley-Terry Dinâmico, a especificação EWMA é usada para acomodar a dependência temporal entre resultados das partidas do mesmo time com diferenças adequadas entre jogos dentro e fora de casa (Cattelan et al., 2013).

### 4.2.1 A função de Verossimilhança

Seja  $\boldsymbol{\gamma} = (\beta_1, \beta_2, \delta)^T$  o vetor de parâmetros de interesse e  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$  o vetor dos parâmetros de suavização. A função de verossimilhança para  $\boldsymbol{\theta} = (\boldsymbol{\gamma}^T, \boldsymbol{\lambda}^T)^T$  é escrita como:

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{y}) = \mathbb{P}(Y_i = y_i; \boldsymbol{\theta}) = \prod_{i=2}^N \mathbb{P}(Y_i = y_i | Y_{i-1} = y_{i-1}, \dots, Y_1 = y_1; \boldsymbol{\theta}) \quad (4.10)$$

A probabilidade condicional para o resultado da partida  $i$  é expresso pelo modelo:

$$\mathbb{P}(Y_i \leq y_i | Y_{i-1} = y_{i-1}, \dots, Y_1 = y_1; \boldsymbol{\theta}) = \frac{\exp\{\delta_{y_i} + \beta_1 x_{m_i}(t_i; \lambda_1) - \beta_2 x_{v_i}(t_i; \lambda_2)\}}{1 + \exp\{\delta_{y_i} + \beta_1 x_{m_i}(t_i; \lambda_1) - \beta_2 x_{v_i}(t_i; \lambda_2)\}}.$$

A maximização da verossimilhança é feita em dois passos. Primeiramente o vetor de parâmetros  $\boldsymbol{\lambda}$  é estimado maximizando  $\mathcal{L}(\boldsymbol{\gamma}_{\hat{\lambda}}, \boldsymbol{\lambda}; \mathbf{y})$  e, em seguida  $\boldsymbol{\gamma}$  é estimado dado  $\boldsymbol{\lambda}$ , ou seja  $\boldsymbol{\gamma}_{\hat{\lambda}}$ .

### 4.2.2 Avaliação do Modelo

A avaliação do modelo é baseado na comparação das probabilidades ajustadas do modelo dinâmico com as probabilidades ajustadas do modelo não-estruturado, ou seja, o modelo estático.



O Bradley-Terry dinâmico tem como objetivo capturar a evolução no tempo das habilidades de todos os times com a estimação de apenas cinco parâmetros ( $\beta_1, \beta_2, \lambda_1, \lambda_2$  e  $\delta$ ), enquanto que o modelo não estruturado estima  $k+2$  parâmetros ( $\alpha_1, \dots, \alpha_{k+1}$  e  $\delta$ ). É esperado que o modelo não-estruturado se ajuste melhor aos dados, por isso ele pode ser visto como um *benchmark*.

Como o futebol leva em conta empates, é melhor empregar um índice que leva em consideração toda a distribuição de probabilidade, como o *ranked probability score* (RPS), que foi originalmente introduzido para *rankear* dados categóricos (Epstein, 1969) e que foi adaptado para dados de contagem (Czado et al., 2009) definido por:

$$\text{RPS}_i = \sum_{l=0}^{L-1} \left\{ \mathbb{P}(Y_i \leq l | Y_{i-1} = y_{i-1}, \dots, Y_1 = y_1; \hat{\theta}) - \mathbf{1}(y_i \leq l) \right\}^2, \quad (4.11)$$

onde  $\mathbf{1}(y_i \leq l)$  é a função indicadora do evento  $\{y_i \leq l\}$ ,  $L = 3$  quando empates são permitidos em um campeonato e  $\hat{\theta}$  é a estimativa de máxima verossimilhança de  $\theta$  baseado nos resultados de todas as partidas jogadas, isto é  $y_1, \dots, y_N$ .

O RPS generaliza o erro absoluto e fornece uma forma direta de comparar previsões pontuais e distribuições preditivas (Czado et al., 2009). O modelo ajustado será avaliado por um ponto de vista preditivo, portanto o RPS será quantificado usando a estimativa de máxima verossimilhança  $\hat{\theta}_{(i-1)}$  computada apenas com partidas disputadas antes da partida prevista  $i$ , isto é, apenas com resultados  $y_1, \dots, y_{i-1}$ .

## 4.3 Metodologia

O conjunto de dados utilizado neste capítulo da monografia é o mesmo do [Capítulo 3](#). Os pacotes do R também foram os mesmos, mas com a adição do pacote “*qvcalc*” (Firth, 2019) para calcular as quasi-variâncias (Firth e De Menezes, 2004) das habilidades estimadas através do Bradley-Terry e conseqüentemente obter os quasi-erros padrão (QSE).

Todas as implementações dos modelos deste capítulo foram fornecidos por [Cattelan et al. \(2013\)](#). Além disso, como tópico especial será considerado o campeonato brasileiro de 2019, o Brasileirão. A motivação dá-se pelo fato de ter havido uma alta discrepância entre o Flamengo e outros clubes do campeonato, o que se parece com o que é visto atualmente nas ligas europeias.

### 4.3.1 Banco de Dados

Diferentemente do [Capítulo 3](#) que foram consideradas 56 edições, neste capítulo serão considerados apenas a última edição (2018-19) de todos os campeonatos para o cálculo das habilidades estimadas dos times. Serão ajustados os dois modelos de Bradley-Terry: o estático (não-estruturado), que apresenta a estimativa da habilidade de cada equipe no campeonato; e o dinâmico, que estima uma habilidade por partida para cada equipe, de forma separada em termos de mandante e visitante.

Utiliza-se somente a identificação das equipes que jogaram como mandante e visitante, a quantidade de gols feitos por cada uma e a data do jogo. Os dados devem estar organizados em ordem cronológica. Em caso de partidas no mesmo dia, ordena-se os jogos por ordem alfabética em relação às equipes que disputam a partida em casa.

Foi considerado o uso das estatísticas dos jogadores em cada partida (ver Apêndice B), que foram retiradas do site [WhoScored \(2019\)](#) através de *webscraping* (ver Apêndice C) como potenciais variáveis explicativas do modelo, porém como não havia discriminação das covariáveis em termos de resultado da partida, optou-se por utilizar o modelo sem essas variáveis.

### 4.3.2 Condição Inicial

O modelo necessita de uma condição inicial (CI) que é o número médio de pontos obtidos pelos mandantes e visitantes na temporada anterior, denotado por  $\bar{r}_m$  e  $\bar{r}_v$ , respectivamente. Por exemplo, na [Tabela 4.1](#), observa-se que o número médio de pontos obtido pelos times que jogaram em casa na temporada de 2017-18 da *Bundesliga* foi de 1,634 pontos. Já para os clubes que disputavam a partida no campo adversário, a pontuação média foi de 1,095 e, portanto, esses são as condições iniciais para os modelos de habilidade do campeonato alemão de 2018-19.

Tabela 4.1: Condições iniciais utilizadas no algoritmo do modelo de Bradley-Terry Dinâmico de cada liga

Condição Inicial	Campeonato					
	<i>Bundesliga</i>	<i>P. League</i>	<i>La Liga</i>	<i>Ligue 1</i>	<i>Serie A</i>	Brasileirão
$\bar{r}_m$	1,634	1,626	1,639	1,618	1,513	1,884
$\bar{r}_v$	1,095	1,113	1,134	1,129	1,268	0,826

## 4.4 Resultados

Os resultados dispostos nesta seção dizem respeito à modelagem por meio do Bradley-Terry e Bradley-Terry Dinâmico aplicado a edição de 2018-19 da *Bundesliga* e a edição de 2019 do Brasileirão. O restante dos resultados referente à EPL, *La Liga*, *Ligue 1* e *Serie A* estão apresentados no Apêndice D.

### 4.4.1 *Bundesliga*

Na temporada de 2018-19 da *Bundesliga* houveram  $k+1 = 18$  equipes, com cada time disputando 34 jogos, sendo metade disputado como mandante e a outra metade como visitante. O número total de partidas da edição foi de  $N = 306$  e ocorreu entre 24 de agosto 2018 e 18 de maio 2019, em 101 dias distintos. O campeão foi o *Bayern München* com 78 pontos (24V, 6E e 4D) e os rebaixados foram o *Hannover* (5V, 6E e 23D) e *Nürnberg* (3V, 10E e 21D). O *Stuttgart* (7V, 7E e 20D), como ficou em 16<sup>o</sup> na classificação, jogou os *play-offs*, mas acabou perdendo e foi rebaixado também.

A temporada foi decidida na última rodada e a diferença entre o primeiro e o segundo colocado foi de apenas dois pontos. Ao todo, o *Borussia Dortmund* passou vinte rodadas no topo da tabela, porém por causa de tropeços, na vigésima oitava rodada o *Bayern* assumiu o topo e manteve-se lá até que consagrou-se campeão novamente.

#### 4.4.1.1 Modelo de Bradley-Terry

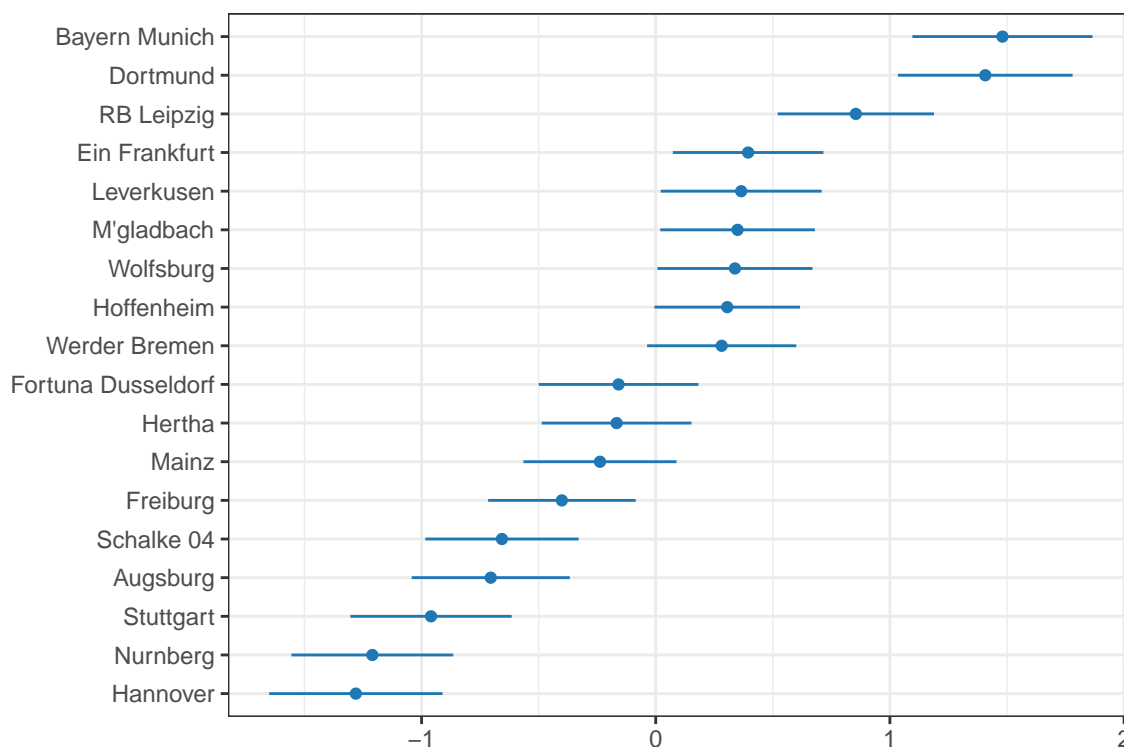


Figura 4.1: Habilidades e QSE estimados das 18 equipes da edição de 2018 da *Bundesliga* pelo modelo de Bradley-Terry

Na [Figura 4.1](#) estão dispostas as habilidades e QSE estimados pelo modelo de Bradley-Terry. A ordenação está em ordem crescente da estimativa da habilidade de cada um das 18 equipes. Observa-se que as habilidades estimadas do *Bayern* (1,481) e do *Dortmund* (1,407) ao fim da edição de 2018 ficaram muito próximas. Em termos de aproveitamento durante o campeonato ambos obtiveram campanhas parecidas, porém o BVB teve uma vitória a menos e um empate a mais e por isso perdeu o título para o rival. Em termos gerais, alguns clubes obtiveram estimativas bem parecidas, o que já era esperado visto que na classificação final do torneio a diferença de pontos entre o *Bayer Leverkusen* (4º) e o *Hoffenheim* (9º) foram de apenas 7 pontos. Algumas posições entre o *ranking* do modelo e do campeonato mudam, mas nada muito extremo.

#### 4.4.1.2 Modelo de Bradley-Terry Dinâmico

Os parâmetros estimados de suavização específicos do mandante e visitante,  $\hat{\lambda}_1 = 0,083$  e  $\hat{\lambda}_2 = 0,069$ , foram valores bem próximos de zero. O vetor de parâmetros de interesse estimado foi de  $\hat{\gamma} = (\hat{\beta}_1; \hat{\beta}_2; \hat{\delta})^T = (1,553; 1,926; 0,610)^T$ , com erros padrão de 0,2670, 0,368 e 0,107, respectivamente.

Diferentemente do modelo estático, o modelo de Bradley-Terry dinâmico estima uma habilidade para a equipe em cada partida e como na especificação do modelo considera-se só as partidas anteriores em casa para estimar a habilidade do jogo atual como mandante, então os gráficos suavizados representando a aptidão dos clubes ao

longo das partidas são dispostos em termos de mandante e visitante. Os primeiros nove clubes, em ordem de classificação final, são apresentados na [Figura 4.2](#).

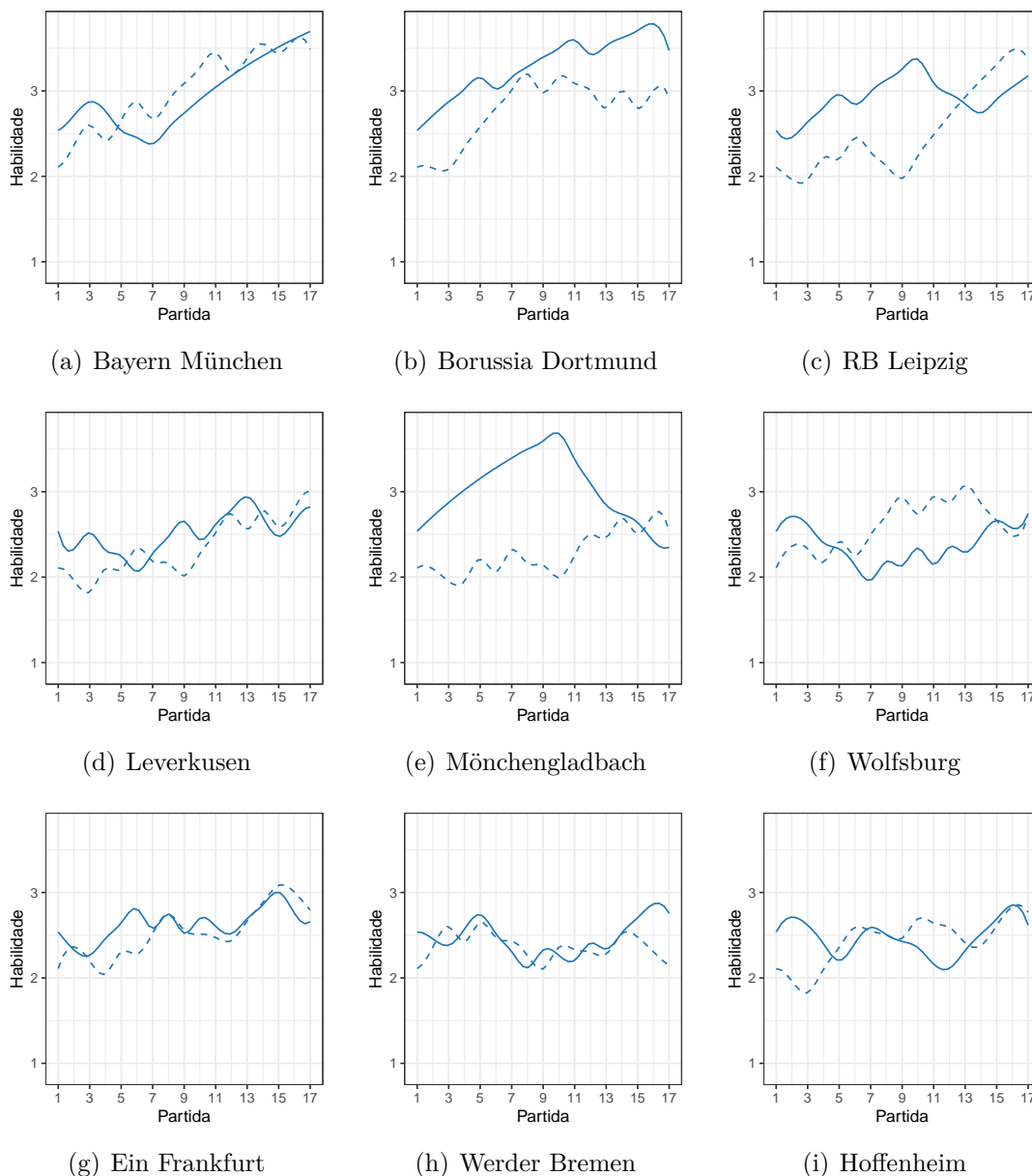


Figura 4.2: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da *Bundesliga* - 1<sup>o</sup> ao 9<sup>o</sup> colocado

Pode-se notar que a diferença entre o *Bayern* e o *Dortmund* é que a ascensão na habilidade estimada do primeiro foi mais acentuada, já que o BVB teve um aumento mais contido entre um jogo e outro, principalmente a partir da nona rodada. O *M'gladbach* teve uma queda de rendimento dentro de casa a partir do meio até o final da temporada, enquanto que o *Leipzig* obteve uma tendência positiva nos jogos fora de casa praticamente no mesmo período.

No caso do *Wolfsburg*, em alguns períodos a habilidade estimada quando jogava fora de casa era maior do que quando jogava dentro de casa. Apesar disso, o clube terminou o campeonato em 6<sup>o</sup> colocado, com 16 vitórias, 7 empates e 11 derrotas, sendo que tanto em termos de mandante (8V, 5E e 4D) quanto visitante (8V, 2E e

7D), continuou sendo o sexto na classificação.

As equipes que ficaram na segunda parte da tabela de classificação estão apresentadas na [Figura 4.3](#).

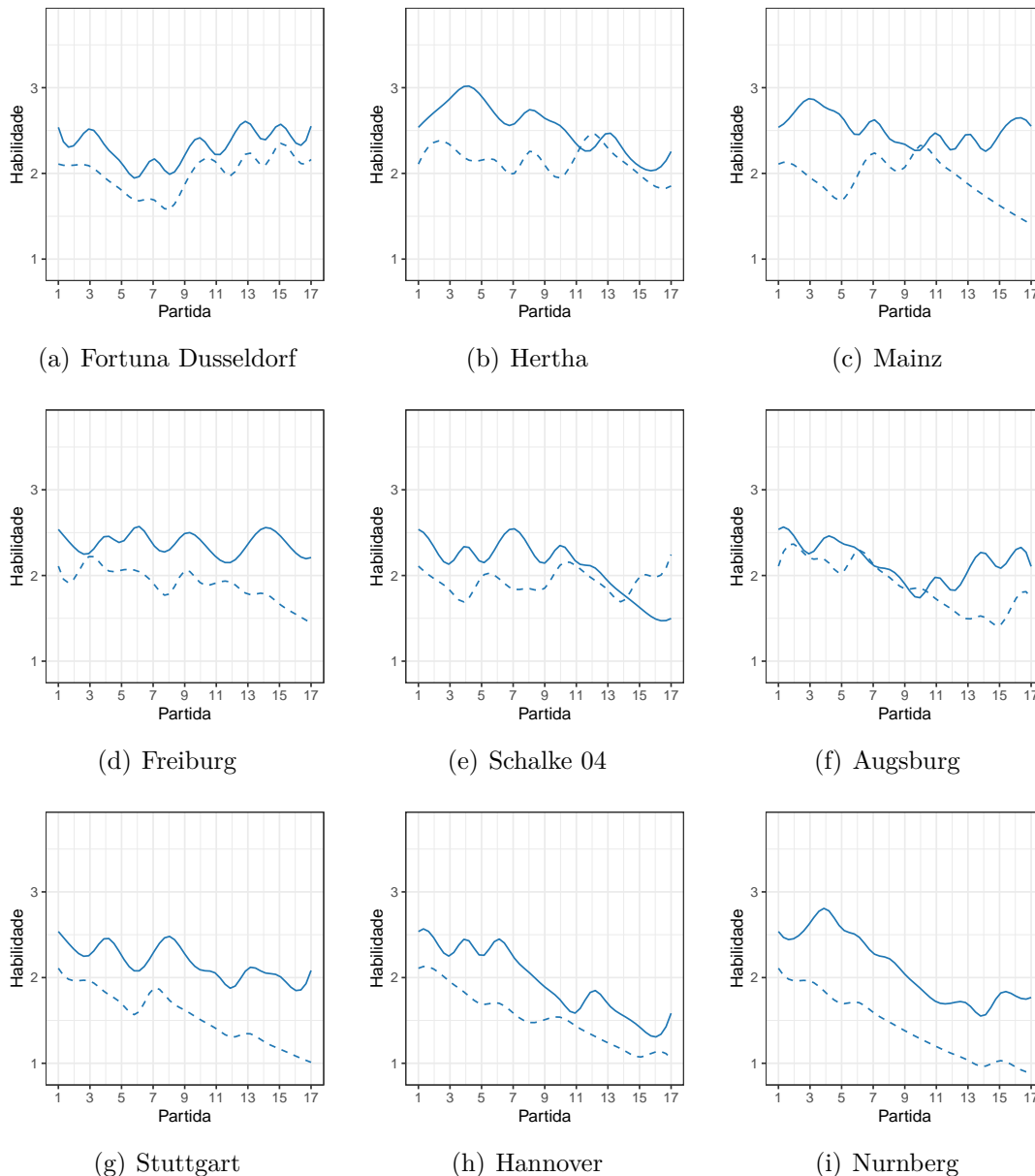


Figura 4.3: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da *Bundesliga* - 10<sup>o</sup> ao 18<sup>o</sup> colocado

Não é surpresa o declínio nas habilidades estimadas das três piores equipes dessa edição. Pelo gráfico do *Hannover*, nota-se o motivo que o fez ocupar a última posição geral da [Tabela 4.2](#) do modelo dinâmico. A curva suavizada para as partidas dentro de casa possui estimações menores, se comparadas às do *Nurnberg*, que por sua vez obteve um decréscimo forte na curva de visitante.

#### 4.4.1.3 Modelo Estático x Modelo Dinâmico

Pela Tabela 4.2 é possível analisar as estimativas pontuais para o modelo estático e dinâmico. Os dados estão dispostos por ordem de classificação final do campeonato e observa-se que o *rank* considerando a habilidade estimada no modelo não-estruturado foi parecida com a ordem de pontos da edição. Isso se dá pelo fato de que o resultado obtido pela equipe na partida é utilizado para estimar as habilidades dos clubes, que também é o que define a quantidade de pontos obtidos pelo time na partida.

Tabela 4.2: Tabela resumo das habilidades e QSE estimados pelo modelo BT-Estático e as habilidades suavizadas médias pelo BT-Dinâmico em termos gerais, como mandante e visitante da *Bundesliga* de 2018

Equipe	Modelo Estático			Modelo Dinâmico					
	Habilidade	QSE	Rank	Habilidade	Rank	Habilidade (Mandante)	Rank	Habilidade (Visitante)	Rank
1. Bayern Munich	1,481	0,385	1	0,643	2	0,458	3	0,829	1
2. Dortmund	1,407	0,373	2	0,687	1	0,786	1	0,589	2
3. RB Leipzig	0,855	0,334	3	0,391	3	0,431	4	0,351	5
4. Leverkusen	0,365	0,344	5	0,103	8	0,025	7	0,182	8
5. M'gladbach	0,349	0,330	6	0,308	4	0,516	2	0,100	9
6. Wolfsburg	0,338	0,331	7	0,157	6	-0,108	11	0,422	3
7. Ein Frankfurt	0,395	0,322	4	0,243	5	0,132	5	0,354	4
8. Werder Bremen	0,282	0,318	9	0,081	9	-0,025	9	0,186	7
9. Hoffenheim	0,305	0,311	8	0,118	7	-0,025	9	0,260	6
10. Fortuna Dusseldorf	-0,159	0,341	10	-0,166	12	-0,164	13	-0,168	11
11. Hertha	-0,167	0,320	11	-0,003	10	0,030	6	-0,037	10
12. Mainz	-0,238	0,327	12	-0,116	11	0,019	8	-0,251	13
13. Freiburg	-0,401	0,315	13	-0,202	13	-0,123	12	-0,280	14
14. Schalke 04	-0,657	0,327	14	-0,310	15	-0,389	17	-0,232	12
15. Augsburg	-0,705	0,338	15	-0,306	14	-0,330	15	-0,282	15
16. Stuttgart	-0,960	0,344	16	-0,467	16	-0,315	14	-0,619	16
17. Hannover	-1,281	0,370	18	-0,594	18	-0,536	18	-0,652	17
18. Nurnberg	-1,210	0,346	17	-0,567	17	-0,382	16	-0,752	18

O *ranking* do modelo estático é parecido com a classificação geral do campeonato, com pequenas modificações. O *Ein Frankfurt* foi o 4º no *ranking*, mas 7º na *Bundesliga*. É curioso observar como o *Wolfsburg* equilibra a 6ª posição da habilidade média geral considerando o modelo dinâmico, sendo o 11º entre os mandantes e 3º entre os visitantes.

A campanha do *Schalke 04* como mandante teve um aproveitamento de apenas 29,4%. No *ranking*, pelo modelo dinâmico, a habilidade média estimada como mandante foi de -0,389, 17º colocado.

#### 4.4.1.4 Avaliação do modelo

Para a avaliação foi utilizado o RPS. O modelo foi ajustado para os dados até o 50º dia de campeonato, que equivale ao primeiro turno do campeonato alemão dessa edição. Os resultados das partidas a partir do 51º dia são preditos e comparados com os resultados observados pelo RPS. O modelo então é reajustado incluindo também as partidas do dia 51 e é utilizado pra prever os resultados do dia 52 e assim por diante até o último dia. As predições são comparadas também com as proporções empíricas de vitórias e derrotas do time mandante, computada pelas partidas dos primeiros 50 dias.

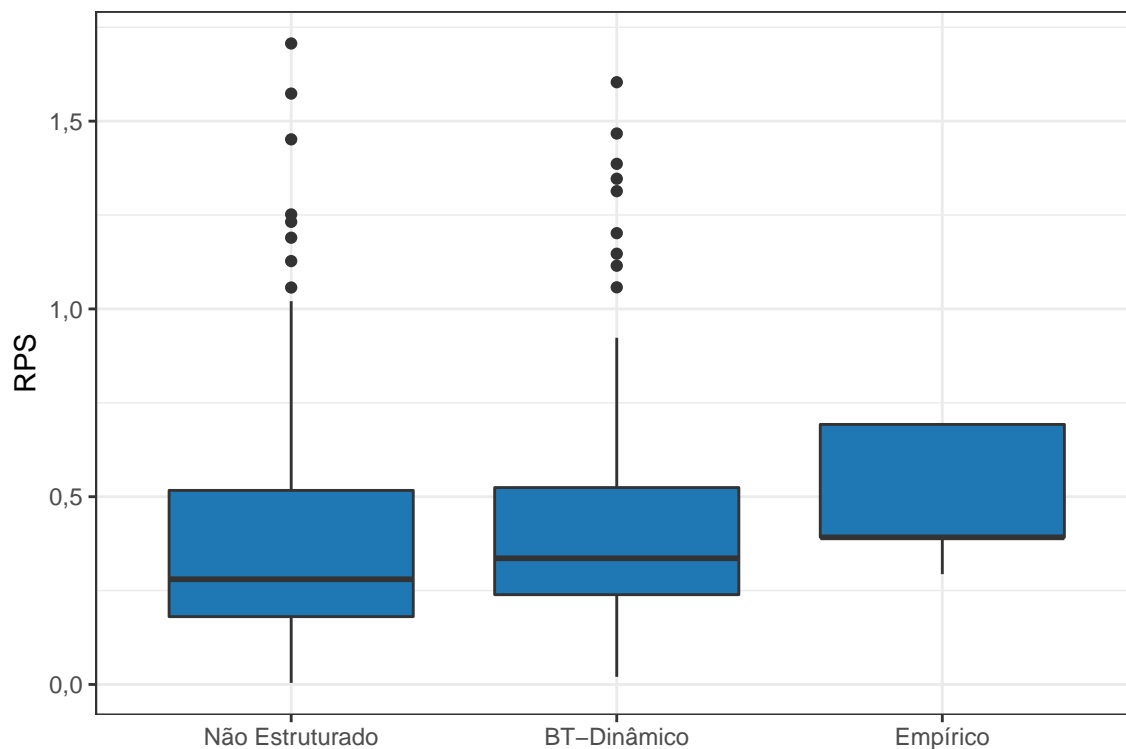


Figura 4.4: Boxplot do RPS para a *Bundesliga* de 2018 computado para o modelo não-estruturado, BT-Dinâmico e previsões baseadas em proporções empíricas

O RPS para o modelo dinâmico foi de 0,423 e o do não-estruturado foi de 0,362, 14,4% menor. O boxplot da [Figura 4.4](#) corresponde às estimativas do RPS obtidas para as partidas previstas. Observa-se que o modelo de BT-Dinâmico ficou similar ao não-estruturado, enquanto que o empírico é pior.

## 4.4.2 Brasileiro

Na temporada de 2019 do *Brasileirão* houveram  $k + 1 = 20$  equipes, com cada time disputando 38 jogos, sendo metade disputado como mandante e a outra metade como visitante. O número total de partidas da edição foi de  $N = 380$  e ocorreu entre 27 de abril e 8 de dezembro de 2019, em 104 dias distintos. O campeão foi o Flamengo (28V, 6E e 4D), com 90 pontos e o título foi conquistado com quatro rodadas de antecedência, antes mesmo de entrar em campo, pois o Palmeiras foi derrotado por  $2 \times 1$  pelo Grêmio. Nesta edição o Z4 ficou definido por: Avaí (3V, 11E e 24D), Chapecoense (7V, 11E e 20D), CSA (8E, 8E e 22D) e Cruzeiro (7E, 15E e 16D), rebaixado pela primeira vez em sua história.

### 4.4.2.1 Modelo de Bradley-Terry

As estimativas obtidas através do modelo de Bradley-Terry, que estima somente um valor de habilidade para cada equipe durante todo o campeonato está apresentado na [Figura 4.5](#).

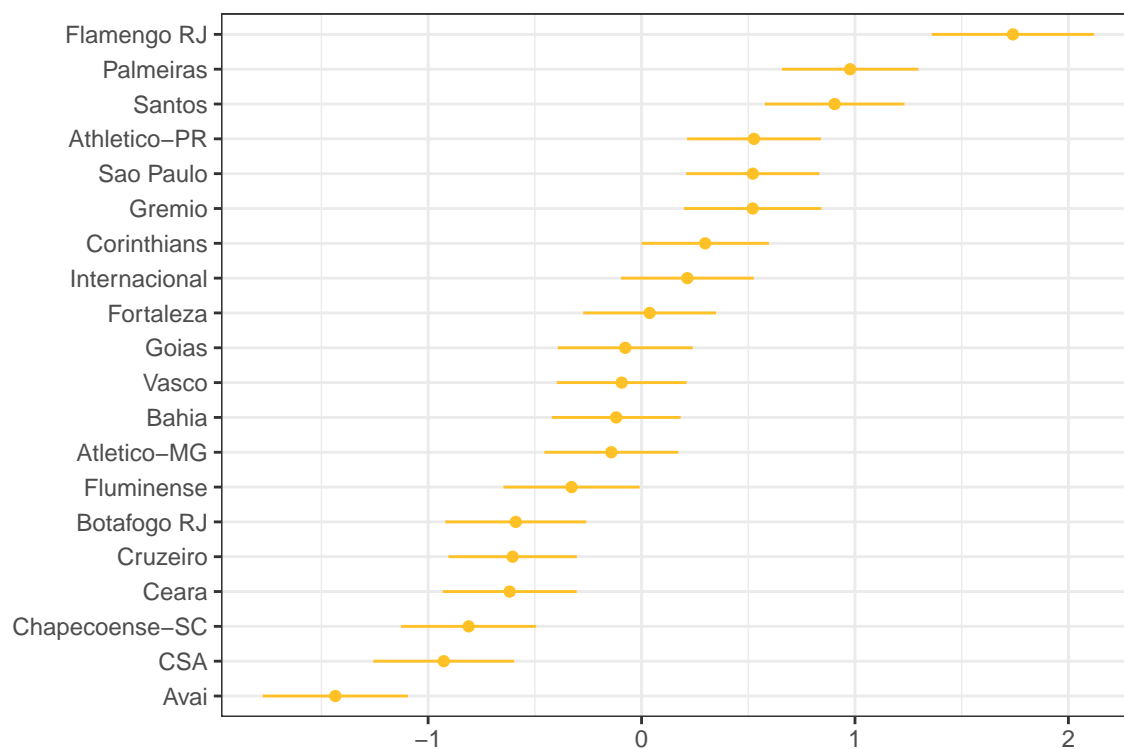


Figura 4.5: Habilidades e QSE estimados das 20 equipes da edição de 2019 do Brasileirão pelo modelo de Bradley-Terry

Não é nenhuma surpresa que o Flamengo obteve a maior estimativa. Além disso, o intervalo representado pelo QSE não tem intersecção com o do Palmeiras, indicando que são diferentes. O Avaí foi a pior equipe do torneio deste ano, o que já era esperado devido a péssima campanha feita durante toda a competição. Observa-se também que as habilidades estimadas para o Grêmio e para o São Paulo foram as mesmas, de 0,521.

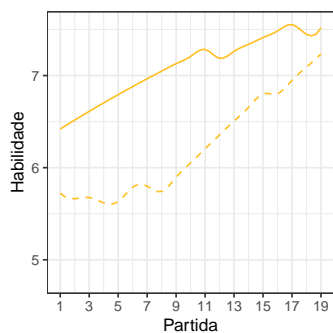
#### 4.4.2.2 Modelo de Bradley-Terry Dinâmico

O modelo de Bradley-Terry dinâmico também foi ajustado para as equipes do Brasileirão.

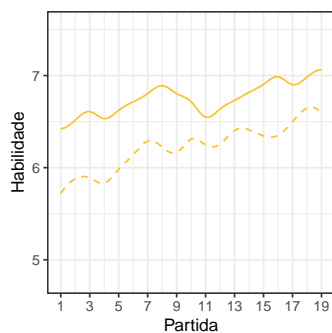
Os parâmetros estimados de suavização específicos do mandante e visitante,  $\hat{\lambda}_1 = 0,025$  e  $\hat{\lambda}_2 = 0,010$ , foram valores bem próximos de zero. O vetor de parâmetros de interesse estimado foi de  $\hat{\gamma} = (\hat{\beta}_1; \hat{\beta}_2; \hat{\delta})^T = (3,407; 6,928; 0,518)^T$ , com erros padrão de 0,639, 1,409 e 0,092, respectivamente.

Em geral as equipes possuem maiores habilidades dentro de casa, o que já era esperado para o futebol. Dito isto, as diferenças entre as curvas suavizadas da habilidade do Internacional como mandante e como visitante parece ser as mais afastadas entre os clubes da Figura 4.6. Durante toda a temporada o time gaúcho sofreu por problemas ligados exatamente a essa questão: a diferença nos resultados dentro e fora de casa. O aproveitamentos foi de 68,42% e 31,58%, respectivamente, o que mostra a fragilidade da equipe em partidas longe do Beira-Rio, que é confirmado pelo modelo através das estimações.

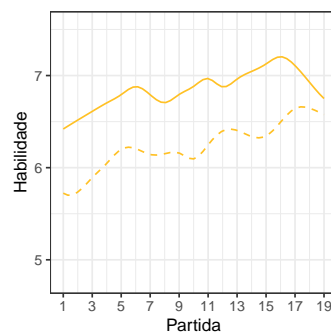




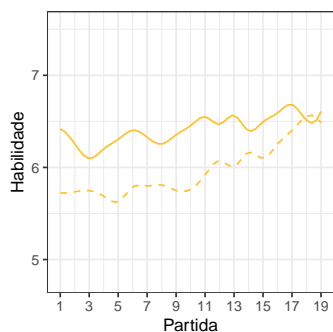
(a) Flamengo



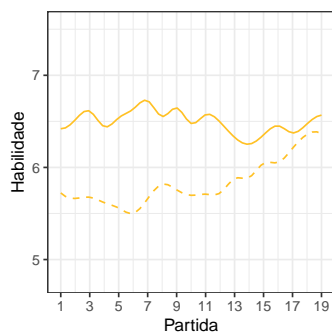
(b) Santos



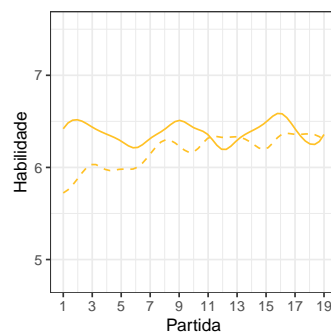
(c) Palmeiras



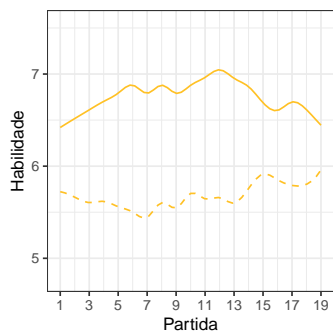
(d) Grêmio



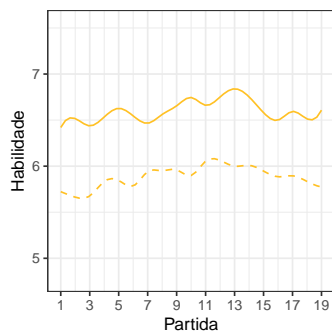
(e) Athletico PR



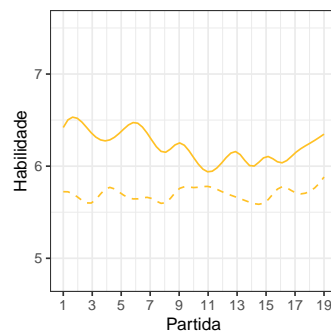
(f) São Paulo



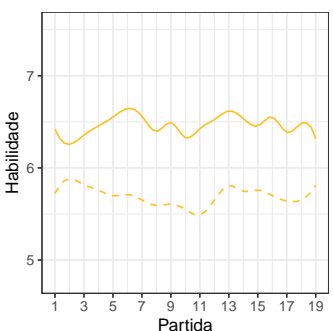
(g) Internacional



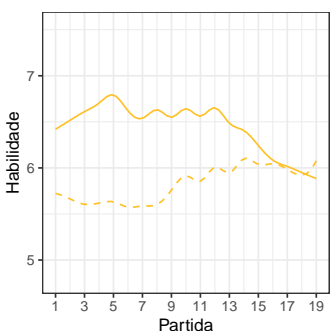
(h) Corinthians



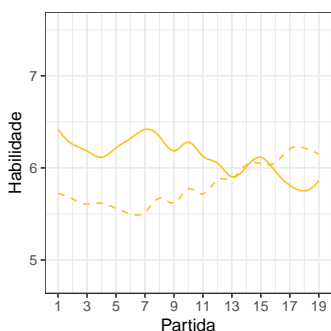
(i) Fortaleza



(j) Goiás



(k) Bahia



(l) Vasco

Figura 4.6: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2019 do Brasileirão - 1<sup>o</sup> ao 12<sup>o</sup> colocado

A crescente na habilidade do Flamengo dentro do Maracanã é explicada pela campanha invicta durante o torneio: foram 17 vitórias e 2 empates ( $4 \times 4$  contra o Vasco e  $0 \times 0$  contra o São Paulo).

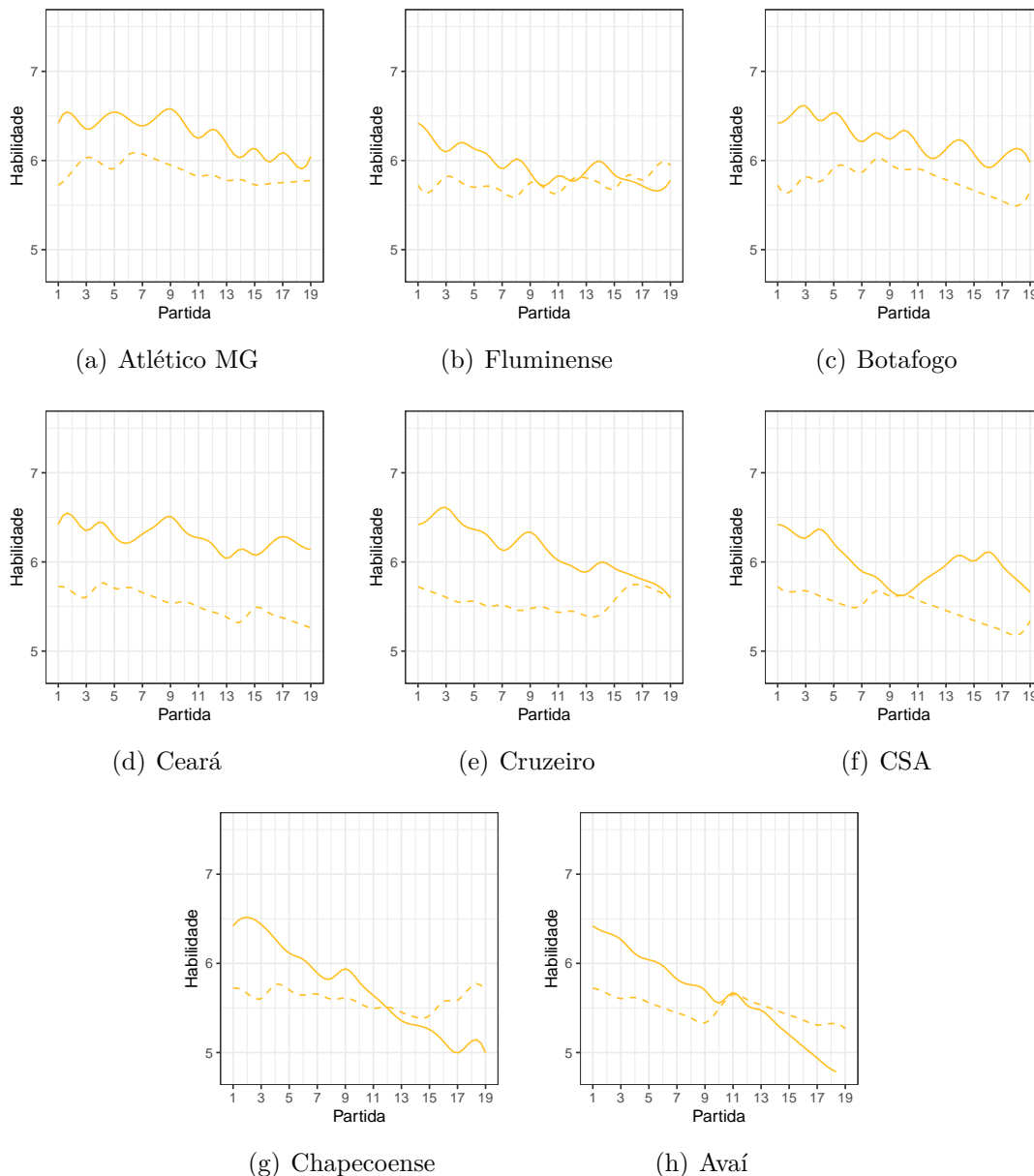


Figura 4.7: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2019 do Brasileirão - 13<sup>o</sup> ao 20<sup>o</sup> colocado

O comportamento da habilidade ao longo das partidas do Avaí e da Chapecoense são similares: a habilidade estimada como mandante de ambos os clubes foi enfraquecendo durante o torneio, uma consequência oriunda da falta de resultados positivos dessas equipes dentro de casa. Fora de casa, por outro lado, mostrou-se mais constante devido aos empates que essas equipes conseguiam segurar quando estavam em estádios adversários. O cruzeiro, que ficou em 16<sup>o</sup> na classificação final, foi rebaixado pela primeira vez em sua história. Sua curva de habilidade suavizada evidencia uma queda de rendimento dentro do Mineirão.

#### 4.4.2.3 Modelo Estático x Modelo Dinâmico

Como as duas versões do modelo de Bradley-Terry utilizam só os resultados das partidas para modelar a habilidade das equipes, é esperado que o *ranking* dos clubes sob esses métodos seja próxima da classificação dos pontos totais obtidos pelos times durante o campeonato, com pequenas modificações.

A Tabela 4.3 apresenta as habilidades estimadas através dos dois modelos de Bradley-Terry, estático e dinâmico, bem como as classificações de acordo com as estimativas.

Tabela 4.3: Tabela resumo das habilidades e QSE estimados pelo modelo BT-Estático e as habilidades suavizadas médias pelo BT-Dinâmico em termos gerais, como mandante e visitante do Brasileirão de 2019

Equipe	Modelo Estático			Modelo Dinâmico					
	Habilidade	QSE	Rank	Habilidade	Rank	Habilidade (Mandante)	Rank	Habilidade (Visitante)	Rank
1. Flamengo RJ	1,740	0,380	1	0,579	1	0,754	1	0,405	2
2. Santos	0,904	0,327	3	0,406	3	0,409	3	0,404	3
3. Palmeiras	0,977	0,320	2	0,462	2	0,506	2	0,418	1
4. Gremio	0,521	0,322	6	0,115	7	0,079	9	0,152	5
5. Athletico-PR	0,527	0,314	4	0,089	8	0,157	6	0,022	8
6. Sao Paulo	0,521	0,312	5	0,199	4	0,043	10	0,355	4
7. Internacional	0,215	0,311	8	0,136	6	0,405	4	-0,134	15
8. Corinthians	0,298	0,299	7	0,161	5	0,256	5	0,065	6
9. Fortaleza	0,038	0,311	9	-0,115	14	-0,117	14	-0,113	13
10. Goias	-0,076	0,316	10	0,002	11	0,122	7	-0,119	14
11. Bahia	-0,119	0,302	12	0,053	9	0,098	8	0,008	9
12. Vasco	-0,093	0,304	11	-0,110	13	-0,217	15	-0,003	10
13. Atletico-MG	-0,142	0,314	13	0,004	10	-0,043	11	0,052	7
14. Fluminense	-0,328	0,319	14	-0,232	16	-0,397	18	-0,067	12
15. Botafogo RJ	-0,589	0,330	15	-0,065	12	-0,092	13	-0,038	11
16. Ceara	-0,618	0,314	17	-0,172	15	-0,057	12	-0,288	18
17. Cruzeiro	-0,604	0,301	16	-0,239	17	-0,218	16	-0,260	17
18. CSA	-0,927	0,330	19	-0,331	18	-0,345	17	-0,316	19
19. Chapecoense-SC	-0,810	0,317	18	-0,421	19	-0,625	19	-0,217	16
20. Avai	-1,435	0,340	20	-0,523	20	-0,719	20	-0,327	20

Fica evidente a vulnerabilidade do Internacional diante de partidas como visitante. O *ranking* de habilidade médio para partidas fora de casa indica a 15<sup>a</sup> posição ao clube, muito inferior ao 7<sup>o</sup> lugar que a equipe terminou a temporada. Entretanto como mandante obteve a 4<sup>a</sup> melhor habilidade média do torneio, oposto a campanha do São Paulo, que ficou em 10<sup>o</sup>.

#### 4.4.2.4 Avaliação do modelo

Para avaliação foi utilizado o RPS. O modelo foi ajustado para os dados até o 52<sup>o</sup> dia de campeonato, que equivale ao primeiro turno do campeonato brasileiro dessa edição. Os resultados das partidas a partir do 53<sup>o</sup> dia são preditos e comparados com os resultados observados pelo RPS. O modelo então é reajustado incluindo também as partidas do dia 53 e é utilizado pra predizer os resultados do dia 54 e assim por diante até o último dia. As predições são comparadas também com as proporções empíricas de vitórias e derrotas do time mandante, computada pelas partidas dos primeiros 52 dias.

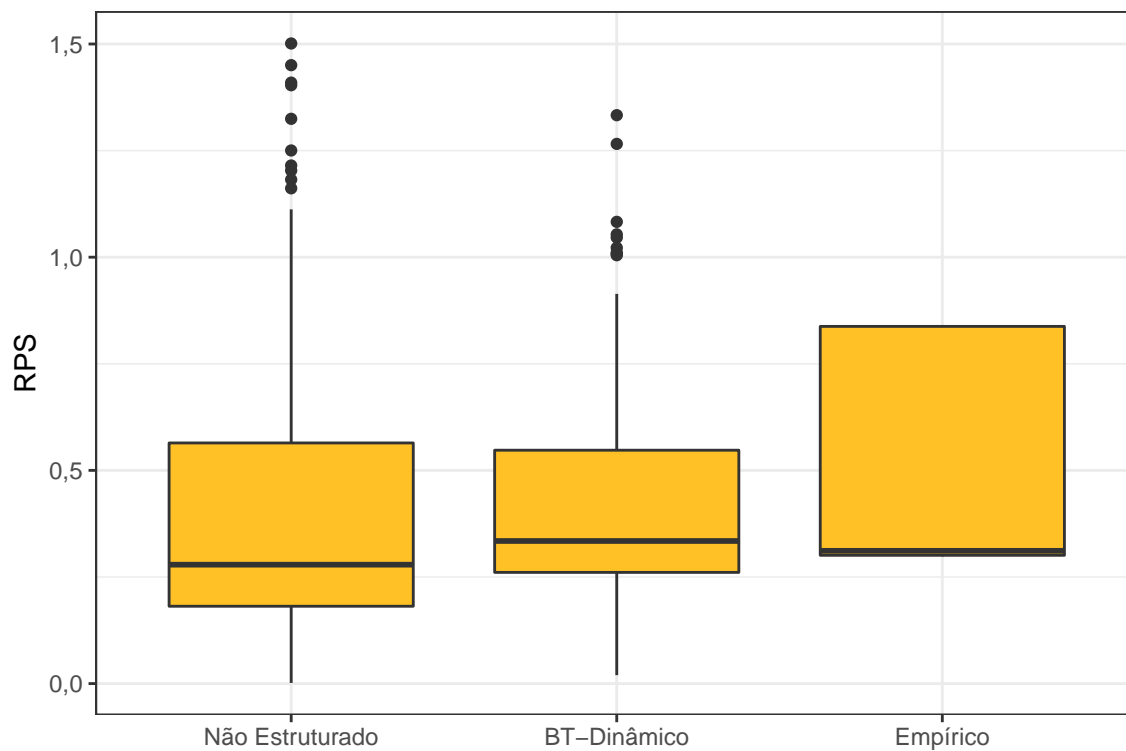


Figura 4.8: Boxplot do RPS para o Brasileirão de 2019 computado para o modelo não-estruturado, BT-Dinâmico e previsões baseadas em proporções empíricas

O RPS para o modelo dinâmico foi de 0,416 e o do não-estruturado foi de 0,364, 12,5% menor. O boxplot da [Figura 4.8](#) corresponde às estimativas do RPS obtidas para as partidas previstas. Observa-se que novamente o modelo de BT-Dinâmico ficou similar ao não-estruturado, enquanto que o empírico é pior.

## 5 Trabalhos Futuros

Visa-se como possíveis trabalhos futuros a expansão desta monografia em dois possíveis caminhos:

- A expansão da aplicação do coeficiente de sorte e habilidade ([Aoki, 2017](#)) e modelagem por meio do modelo de Bradley-Terry Dinâmico ([Cattelan et al., 2013](#)) a outras ligas, como por exemplo em campeonatos da América do Sul para análises acerca do papel da sorte e da habilidade em torneios da América Latina, principalmente no contexto do futebol brasileiro desde a implementação do sistema de pontos corridos.
- Outra abordagem seria na extensão do modelo proposto por [Cattelan et al. \(2013\)](#) para inclusão de covariáveis. Faz-se a ressalva de que as estatísticas das partidas disponíveis para acesso livre em sites como o [WhoScored \(2019\)](#) e [Football-Data \(2019\)](#) não trazem muitos conhecimentos relevantes em termos de resultado de um jogo, portanto necessita-se desenvolver formas de trabalhar com essas variáveis de modo que consiga-se extrair informações pertinentes.

## 6 Considerações Finais

Os resultados encontrados neste trabalho são, de forma geral, evidências que corroboram com o que já se conhece sobre futebol. A vantagem da equipe que disputa a partida como mandante é confirmada ao se analisar as probabilidades intrínsecas aos possíveis resultados de vitória, empate e derrota. Porém conforme mostrado, observa-se que ao longo dos anos essa vantagem tem diminuído dentro e entre os campeonatos aqui observados, o que pode ser devido ao fato de que o futebol vem se tornando um esporte em que o poder de compra é mais decisivo para definição do vencedor de uma partida, tornando-se assim as ligas menos competitivas.

Atualmente, com maior poder econômico e possibilidade de se gastar milhões em jogadores, as maiores equipes nacionais possuem um poder de mercado superior ao restante dos competidores das ligas, o que faz com que seja difícil não haver diferença na habilidade dos times. De acordo com o [Transfermarkt \(2020\)](#), entre os 25 clubes com maior poder de mercado no mundo atualmente, apenas um não joga nas cinco principais ligas da Europa: o Ajax, que ficou na 23<sup>a</sup> posição e disputa a *Eredivisie*, campeonato holandês e que se encontra nessa posição devido a ótima campanha que fez na última *Champions League*, chegando às semi-finais. Com relação ao restante tem-se: 37,5% (9 de 24) são da EPL: *Manchester City*, *Liverpool*, *Tottenham Hotspur*, *Chelsea*, *Manchester United*, *Arsenal*, *Everton*, *Leicester* e *Wolves*; 20,8% (5 de 24) da *Serie A*: *Juventus*, *Napoli*, *Internazionale*, *Milan* e *Roma*; 16,7% (4 de 24) da *La Liga*: *Real Madrid*, *Barcelona*, *Atlético de Madrid* e *Valencia*; 16,7% (4 de 24) da *Bundesliga*: *Bayern Munich*, *Borussia Dortmund*, *RB Leipzig* e *Bayer Leverkusen* e 8,3% (2 de 24) da *Ligue 1*: *Paris Saint-Germain* e *Lyon*.

Existe também o fato de que apesar de serem as equipes com maior valor de mercado, há grandes discrepâncias entre os clubes. Por exemplo, o *Manchester City*, que ocupa a 1<sup>a</sup> posição, vale €1.3 bilhões, enquanto que o *Wolves*, o 25<sup>o</sup>, vale €359.2 milhões, uma diferença de quase €1 bilhão.

Ao utilizar-se do coeficiente de sorte e habilidade proposto por [Aoki \(2017\)](#), pode-se notar que os valores obtidos em sua maioria foram mais próximos do valor 1 do que do 0. Em termos gerais, os campeonatos não são decididos por sorte e sim pelas habilidades dos clubes, o que é visivelmente observado nos torneios atuais, em que as equipes mais fortes são justamente as que ficam nas primeiras posições, enquanto que as mais “fracas”, normalmente as que vieram da segunda divisão, ficam nas posições inferiores. Notou-se um período de alta variabilidade na *Bundesliga* durante a década de 60, o que pode ser explicado pelo fato de ser as primeiras edições, então as equipes ainda estavam se ajustando.

As estimativas do coeficiente de habilidade mostram evidências de que a *Serie A*

e a EPL são os campeonatos em que a habilidade desempenha o maior papel. Cabe salientar que isso não significa que essas ligas são as mais habilidosas. De fato, é o oposto. Ao se examinar os resultados das últimas edições desses campeonatos, nota-se que a alta variabilidade nos pontos finais dos torneios é devido a grande diferença entre os extremos, em que equipes que terminam em primeiro chegam a fazer 100 pontos, enquanto que os lanternas não chegam a 20.

No caso da *Série A*, a dominância da *Juventus* sobre os outros clubes não é recente. São 35 *scudettos*, contra 15 títulos do *Milan* e *Inter*, cada um. Essa previsibilidade acaba impactando na visão dos telespectadores sobre o torneio, dando a impressão de ser um campeonato “chato” de se assistir, o que não é totalmente mentira. Em relação a EPL, mesmo que o título inglês geralmente fique nas mãos dos mesmos, a agressividade no estilo de jogo faz com que seja muito mais apreciado do que outras ligas, mesmo o campeonato como um todo não sendo tão imprevisível quanto parece.

Com a valorização dos jogadores ficou mais difícil de competir com clubes com grande poder de compra. Dificilmente um time que vale €91.25m irá se tornar vencedor de um campeonato com equipes que valem €579.80m, mesmo que tenha acontecido na EPL de 2015-16 quando o *Leicester* bateu times como o *Manchester City*, *Chelsea* e *Arsenal*, tornando-se campeão da *Premier League*.

Ao testar a significância dos coeficientes por meio de Monte Carlo (MC) observou-se que, no geral, os valores de  $\phi$  são estatisticamente diferentes de zero. Na edição de 1968 do campeonato alemão a hipótese de nulidade de  $\phi$  não foi rejeitada, além disso o coeficiente observado foi negativo.

Em seguida, buscou-se descobrir quais equipes destoam o campeonato a ponto de deixá-lo não aleatório. Foram retiradas 17,62% das equipes para a aleatorização do torneio. Destes, 62,23% foram equipes que se encontravam na parte de cima da tabela. Os resultados da porcentagem de retirada de cada equipe dentro de seu respectivo campeonato não se mostraram surpreendentes. Clubes como *Bayern München*, *Liverpool*, *Manchester United*, *Real Madrid*, *Barcelona* e *Juventus* foram retirados em pelo menos 50% das edições. Não é por acaso que esses clubes estão na lista de maiores e mais ricos do mundo.

Para o coeficiente de habilidade foi possível avaliar e confirmar que as equipes mais populares e ricas são as que mais destoam as ligas. A rotatividade de clubes causado pelo rebaixamento torna menos influente a participação destes nas saídas do algoritmo para aleatorização da temporada. Portanto fica claro a grande relevância que os “clubes grandes” exercem dentro de suas respectivas ligas.

A segunda abordagem foi aplicar o modelo de Bradley-Terry na última edição dos campeonatos para modelar a habilidade das equipes. Como o método utiliza os resultados das partidas na modelagem, o *ranking* apresentado pelo modelo não-estruturado mostrou-se ter alta relação com a classificação final dos campeonatos.

O modelo de Bradley-Terry dinâmico permite estimar habilidades das equipes em cada uma das partidas. Isso permite avaliar o progresso do time em diferentes aspectos: o geral (quando considera-se todas as  $2k$  jogos) e a habilidade temporal em termos de mandante e visitante. Com isso, pode-se comparar o *ranking* verdadeiro do torneio com o que foi encontrado pelo modelo de Bradley-Terry.

Através da modelagem utilizando as duas versões do Bradley-Terry notou-se que as equipes que se encontram em posições intermediárias na tabela geralmente possuem estimativas de habilidade muito próximas umas das outras. Esse resultado

é intuitivo e vai ao encontro com o observado no algoritmo de remoção de times que deixam o campeonato não aleatório.

Em termos de resultados observados nos campeonatos, as ligas em que a diferença de pontos entre os dois primeiros colocados é pequena, a habilidade estimada é parecida. É o caso do *Bayern* e *Dortmund* na *Bundesliga* e *Manchester City* e *Liverpool* na EPL. Se o clube venceu o torneio por uma margem de pontos alta, como por exemplo o PSG, então a habilidade estimada será bem afastada das demais equipes do campeonato. Comportamentos não usuais foram observados também, como por exemplo clubes com a curva de habilidade suavizada como visitante maior do que como mandante. No caso da *La Liga*, destaca-se o fato de que as habilidades quando a partida ocorre dentro de casa são tão superiores às fora de casa, que na maioria das vezes não há intersecções entre as duas curvas.

Como um tópico especial foi feita uma aplicação do modelo de Bradley-Terry no campeonato brasileiro deste ano. A motivação por trás dessa escolha dá-se pela grande discrepância encontrada em termos de qualidade dentro de campo entre o Flamengo e outros clubes do torneio, que remeteu um pouco ao que é encontrado na Europa atualmente, em que o time se sobressai sobre os demais, tornando assim o que é visto dentro de campo algo previsível em termos de resultados já que, de forma geral, quando uma equipe possui poder de compra muito superior aos demais, dificilmente ela será batida. O Internacional ocupou o 4º lugar na classificação considerando somente as partidas como mandante. Porém, foi o 15º entre os visitantes, o que não é uma surpresa já que durante toda a temporada, o clube mostrou-se ser um dos mais vulneráveis fora de casa, chegando a perder partidas contra equipes menores e abaixo na tabela.

Por fim, percebe-se que, no geral, as ligas não são competitivas há anos e talvez nunca tenham sido, apesar de haver campeonatos em que vários clubes possuem grande valor de mercado. É o caso da EPL, por exemplo, que teve apenas quatro vencedores distintos nos últimos dez anos, o que não é muito distante do que foi observado na *La Liga*, que obteve três vencedores diferentes nesse período, apesar de ser uma das ligas menos competitivas, já que geralmente é apenas uma competição entre *Real Madrid* e *Barcelona*. Observa-se que em alguns campeonatos existem hegemonias momentâneas: na *Ligue 1* a década de 2000 foi dominada pelo *Lyon*, já a década de 2010 foi pelo PSG, e que provavelmente se entenderá pelos próximos anos. Na *Serie A* as últimas três décadas foram dominadas por *Milan*, *Internazionale* e *Juventus*, respectivamente, apesar de que ao que tudo indica, a “velha senhora” não perderá o trono tão cedo. No caso dos campeonatos alemão e espanhol essa rotatividade não existe: o *Bayern München* é inevitavelmente o maior clube da *Bundesliga* e dificilmente uma outra equipe irá conseguir superá-lo, enquanto que na *La Liga* a disparidade entre *Real Madrid* e *Barcelona* em comparação com as outras equipes do torneio é, provavelmente inalcançável.

O que fica claro é que cada vez mais os clubes estão investindo milhões em contratações de jogadores e, com isso, são poucas as equipes que estão conseguindo acompanhar essa evolução, fazendo com que os times fiquem ainda mais destoantes do que já são. Os grandes clubes serão ainda mais grandes, enquanto que os pequenos serão cada vez mais pequenos. Quanto às equipes medianas, estas continuam nas posições intermediárias, lutando para, se derem sorte, conseguirem uma vaga na *Europe League*, ou muito raramente, uma vaga na *Champions League*, como é o caso da *Atalanta*, na temporada de 2018-19.



## Referências Bibliográficas

- Agresti, A. (2002). *Categorical Data Analysis*. Wiley Series In Probability and Statistics. John Wiley & Sons, Inc, Hoboken, New Jersey, second edition.
- Anderson, C. e Sally, D. (2013). *The Numbers Game: Why Everything You Know about Soccer Is Wrong*. Penguin Books.
- Aoki, R. Y. S. (2017). Um arcabouço computacional para caracterizar sorte e habilidade em ligas esportivas. Master's thesis, Universidade Federal de Minas Gerais.
- Aoki, R. Y. S., Assunção, R. M., e de Melo, P. O. S. V. (2016). Medindo o tamanho da caixinha de surpresas em ligas de futebol. In *31<sup>o</sup> Simpósio Brasileiro de Banco de Dados, SBBD 2016, Salvador, Bahia, Brasil, October 4-7, 2016.*, pages 211–216.
- Aylott, M. e Aylott, N. (2007). A meeting of social science and football: Measuring the effects of three points for a win. *Sport in Society*, 10(2):205–222.
- Barnard, G. (1963). Discussion of professor bartlett's paper. *Journal of The Royal Society*.
- Bell, A., Brooks, C., e Markham, T. (2013). The performance of football club managers: skill or luck? *Economics & Finance Research*, 1:19–30. DOI: 10.1080/21649480.2013.768829.
- Boas, O. V. (2019). Coluna onda esportiva: Futebol brasileiro x futebol europeu. <https://saocarlosemrede.com.br/coluna-onda-esportiva-futebol-brasileiro-x-futebol-europeu/>. Acesso em: 8 jun. 2019.
- Bradley, R. A. e Terry, M. E. (1952). Rank analysis of incomplete block designs: The method of paired comparisons. *Biometrika*, 39(3/4):324–345.
- Brooks, J., Kerr, M., e Gutttag, J. (2016). Developing a data-driven player ranking in soccer using predictive model weights. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, pages 49–55, New York, NY, USA. ACM.
- Buraimo, B. e Simmons, R. (2015). Uncertainty of outcome or star quality? television audience demand for english premier league football. *International Journal of the Economics of Business*, 22(3):449–469.

- Cattelan, M., Varin, C., e Firth, D. (2013). Dynamic bradley–terry modelling of sports tournaments. *Journal of the Royal Statistical Society Series C Applied Statistics*, 62:135–150.
- Chan, W., Courty, P., e Hao, L. (2008). Suspense: Dynamic Incentives in Sports Contests. *The Economic Journal*, 119(534):24–46.
- Constantinou, A. C. e Fenton, N. E. (2013). Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries. *Journal of Quantitative Analysis in Sports*, 9(1):37–50.
- Crocombe, M. (2018). The origin, history, and invention of soccer. <https://www.liveabout.com/who-invented-soccer-3556873>. Acesso em: 9 jun. 2019.
- Curley, J. P. (2016). *engsoccerdata: English Soccer Data 1871-2016*. R package version 0.1.5. <https://github.com/jalapic/engsoccerdata>.
- Czado, C., Gneiting, T., e Held, L. (2009). Predictive model assessment for count data. *Biometrics*, 65(4):1254–1261.
- DFML (2019). Deloitte football money league 2019. <https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/deloitte-football-money-league.html>. Acesso em: 13 set. 2019.
- Dias, T. e Lavinas, T. (2010). <http://globoesporte.globo.com/futebol/copa-do-mundo/noticia/2010/05/selecao-ja-tem-lista-para-copa.html>. Acesso em: 8 jun. 2019.
- Dilger, A. e Geyer, H. (2009). Are three points for a win really better than two? a comparison of german soccer league and cup games. *Journal of Sports Economics*, 10(3):305–318.
- Dobson, S. e Goddard, J. (2003). Persistence in sequences of football match results: A monte carlo analysis. *European Journal of Operational Research*, 148:247–256.
- Edwards, S. M. (2018). *lemon: Freshing Up your 'ggplot2' Plots*. R package version 0.4.2. <https://github.com/stefanedwards/lemon>.
- EPL (2019). Discover the origins and history of the top tier of english football. <https://www.premierleague.com/history>. Acesso em: 2 jun. 2019.
- Epoca (2014). Felipão convoca seleção brasileira para a copa 2014. <https://epoca.globo.com/vida/copa-do-mundo-2014/noticia/2014/05/bfelipao-convoca-selecao-brasileirab-para-copa-2014.html>. Acesso em: 8 jun. 2019.
- Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology (1962-1982)*, 8(6):985–987.
- Fahrmeir, L. e Tutz, G. (1994). Dynamic stochastic models for time-dependent ordered paired comparison systems. *Journal of the American Statistical Association*, 89(428):1438–1449.

- FIFA (2018). More than half the world watched record-breaking 2018 world cup. <https://www.fifa.com/worldcup/news/more-than-half-the-world-watched-record-breaking-2018-world-cup>. Acesso em: 20 abr. 2019.
- Firth, D. (2019). *qvcalc: Quasi Variances for Factor Effects in Statistical Models*. R package version 1.0.1. <https://davidfirth.github.io/qvcalc>.
- Firth, D. e De Menezes, R. X. (2004). Quasi[U+2010]variances. *Biometrika*, 91(1):65–80.
- FiveThirtyEight (2016). In the era of advanced stats, soccer still lags behind. <https://fivethirtyeight.com/features/in-the-era-of-advanced-stats-soccer-still-lags-behind/>. Acesso em: 9 jun. 2019.
- Football-Data (2019). <http://www.football-data.co.uk/data.php>. Acesso em: 29 abr. 2019.
- FoxSports (2017). Real madrid tem quase dez vezes o valor de mercado do grêmio. veja os números. <https://www.foxsports.com.br/news/335726-real-madrid-tem-quase-dez-vezes-o-valor-de-mercado-do-gremio-veja-os-numeros>. Acesso em: 8 jun. 2019.
- Freitas, L. F. (2017). A história da análise de desempenho no futebol – parte 1. <http://futebolanalitico.com.br/analise-estatistica/a-historia-da-analise-de-desempenho-parte-1/>. Acesso em: 9 jun. 2019.
- Gerrard, B. (2004). *Still Up for Grabs? Maintaining the Sporting and Financial Viability of European Club Soccer*. CT: Praeger Publishers, Westport. In R. Fort and J. Fizel (Eds.), *International Sports Economics Comparisons* (pp. 39-59).
- Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 21(2):331 – 340.
- Guterman, M. (2009). *O futebol explica o Brasil: Uma história da maior expressão popular do país*. Contexto.
- Harrison, J. e Kim, J. Y. (2019). *R Bindings for Selenium WebDriver*. R package version 1.7.5. <https://CRAN.R-project.org/package=RSelenium>.
- Helal, R. G., Lovisolo, H., e Soares, A. J. G. (2000). *A invenção do país do futebol: mídia, raça e idolatria*. Mauad, first edition.
- Hope, A. C. A. (1968). A simplified monte carlo significance test procedure. 30(3):582–598.
- Hunter, D. (2004). Mm algorithms for generalized bradley-terry models. *The Ann. Stat.*, 32.
- Hvattum, L. M. e Arntzen, H. (2010). Using elo ratings for match result prediction in association football. *International Journal of Forecasting*, 26(3):460 – 470. Sports Forecasting.

- Kenmare, J. (2018). The most watched league in european football revealed - the championship > la liga. <http://www.sportbible.com/football/news-take-a-bow-the-most-watched-league-in-european-football-championship-la-liga-20180117>. Acesso em: 20 abr. 2019.
- Knorr-Held, L. (2000). Dynamic rating of sports teams. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(2):261–276.
- Lang, D. T. (2019). *Tools for Parsing and Generating XML Within R and S-Plus*. R package version 3.98-1.20 <https://CRAN.R-project.org/package=XML>.
- Ley, C., de Wiele, T. V., e Eetvelde, H. V. (2019). Ranking soccer teams on the basis of their current strength: A comparison of maximum likelihood approaches. *Statistical Modelling*, 19(1):55–73.
- McHale, I. e Morton, A. (2011). A bradley-terry type model for forecasting tennis match results. *International Journal of Forecasting*, 27(2):619 – 630.
- Menke, J. E. e Martinez, T. R. (2008). A bradley-terry artificial neural network model for individual ratings in group competitions. *Neural Computing and Applications*, 17(2):175–186.
- Mills, J. (2005). *Charles Miller: o pai do futebol brasileiro*. Panda Books, first edition. ISBN: 8587537997.
- Montes, F., Sala-Garrido, R., e Usai, A. (2014). The lack of balance in the spanish first division football league. *European Sport Management Quarterly*, 14(3):282–298. DOI: 10.1080/16184742.2014.898678.
- OptaSports (2019a). *Advanced Metrics*. <https://www.optasports.com/services/analy-tics/advanced-metrics>. Acesso em: 9 jun. 2019.
- OptaSports (2019b). *Opta’s event definitions*. <https://www.optasports.com/news/opta-s-event-definitions/>. Acesso em: 31 out. 2019.
- Pelechrinis, K. e Papalexakis, E. (2016). The anatomy of american football: Evidence from 7 years of nfl game data. *PLOS ONE*, 11(12):1–17.
- Pham, P. (2016). Leicester city and the big trade – using statistics to succeed. <https://www.forbes.com/sites/peterpham/2016/05/30/leicester-city-and-the-big-trade-using-statistics-to-succeed/#4e4307597b47>. Acesso em: 13 dez. 2019.
- Phelan, G. C. e Whelan, J. T. (2017). Hierarchical bayesian bradley-terry for applications in major league baseball.
- Poli, R., Ravenel, L., e Besson, R. (2017). World expatriate footballers. <http://www.football-observatory.com/IMG/sites/mr/mr25/en/>. Acesso em: 8 jun. 2019.
- Pollard, R. (2008). Home advantage in football: A current review of an unsolved puzzle. *The Open Sports Sciences Journal*, 1:12–14.

- Prosperi, L. A. (2018). Confira os 23 convocados da seleção brasileira para a copa do mundo. <http://www.copa2014.gov.br/pt-br/noticia/confira-os-23-convocados-da-selecao-brasileira-para-a-copa-do-mundo>. Acesso em: 8 jun. 2019.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Version 3.6.0 - R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- RStudio (2018). *RStudio: Integrated Development Environment for R*. Version 1.1.463 - RStudio, Inc., Boston, MA. <http://www.rstudio.com/>.
- Schmicker, R. H. (2013). An application of satscan to evaluate the spatial distribution of corner kick goals in major league soccer. *Int. J. Comp. Sci. Sport*, 12(2).
- Schoenfeld, B. (2019). How data (and some breathtaking soccer) brought liverpool to the cusp of glory. <https://www.nytimes.com/2019/05/22/magazine/soccer-data-liverpool.html>. Acesso em: 13 dez. 2019.
- Schumaker, R. P., Jarmoszko, A. T., e Labeledz, C. S. (2016). Predicting wins and spread in the premier league using a sentiment analysis of twitter. *Decision Support Systems*, 88:76 – 84.
- Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., e Despouy, P. (2019). *plotly: Create Interactive Web Graphics via 'plotly.js'*. R package version 4.9.0. <https://CRAN.R-project.org/package=plotly>.
- Simon, R. (2017). Deserving to be lucky: Reflections on the role of luck and desert in sports. *Journal of the Philosophy of Sport*, 34(1):13–25. DOI: 10.1080/00948705.2007.9714706.
- Sire, C. e Redner, S. (2009). Understanding baseball team standings and streaks. *The European Physical Journal B*, 67(3):473–481.
- Spiegelhalter, D. (2007). Football leagues. <http://understandinguncertainty.org/no-de/314>. Acesso em: 16 jun. 2019.
- STATS (2019). <https://www.stats.com/football/>. Acesso em: 9 jun. 2019.
- Stein, L. (2019). Há 90 anos, a história se escrevia na espanha: La liga iniciava a sua primeira edição. <https://trivela.com.br/ha-90-anos-a-historia-se-escrevia-na-espanha-la-liga-iniciava-a-sua-primeira-edicao/>. Acesso em: 2 jun. 2019.
- Transfermarkt (2020). 100 most valuable teams in the world. <https://www.transfermarkt.com/spieler-statistik/wertvollstemannschaften/marktwertetop>. Acesso em: 09 jan. 2020.
- Tutz, G. e Schauburger, G. (2015). Extended ordered paired comparison models with application to football data from german bundesliga. *AStA Advances in Statistical Analysis*, 99(2):209–227.

- UEFA (2019a). Como é calculado o ranking a cinco anos das federações. <https://pt.uefa.com/memberassociations/uefarankings/country/about/>. Acesso em: 20 abr. 2019.
- UEFA (2019b). Ranking do país. <https://pt.uefa.com/memberassociations/uefaran-kings/country/#/yr/2019>. Acesso em: 9 jun. 2019.
- vanden Broucke, S. e Baesens, B. (2018). *Practical Web Scraping for Data Science: Best Practices and Examples with Python*.
- WhoScored (2019). <https://1xbet.whoscored.com/>. Acesso em: 06 dez. 2019.
- Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., e Woo, K. (2019). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. R package version 3.1.1. <https://CRAN.R-project.org/package=ggplot2>.
- Wickham, H. e Henry, L. (2019). *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*. R package version 0.8.3. <https://CRAN.R-project.org/package=tidyr>.
- Wills, C. E. (2017). The competitiveness of games in professional sports leagues. *Journal of Sports Analytics*, 3(2):103–117.
- Wood, G., Jordet, G., e Wilson, M. R. (2015). On winning the “lottery”: psychological preparation for football penalty shoot-outs. *Journal of Sports Sciences*, 33(17):1758–1765. DOI: 10.1080/02640414.2015.1012103.
- Xiao, N. (2018). *ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for 'ggplot2'*. R package version 2.9. <https://CRAN.R-project.org/package=ggsci>.
- Yan, T., Yang, Y., e Xu, J. (2012). Sparse paired comparisons in the bradley-terry model. *Statistica Sinica*, 22(3):1305–1318.

## Apêndice

### A Distribuições do Teste de Monte Carlo para $\phi$

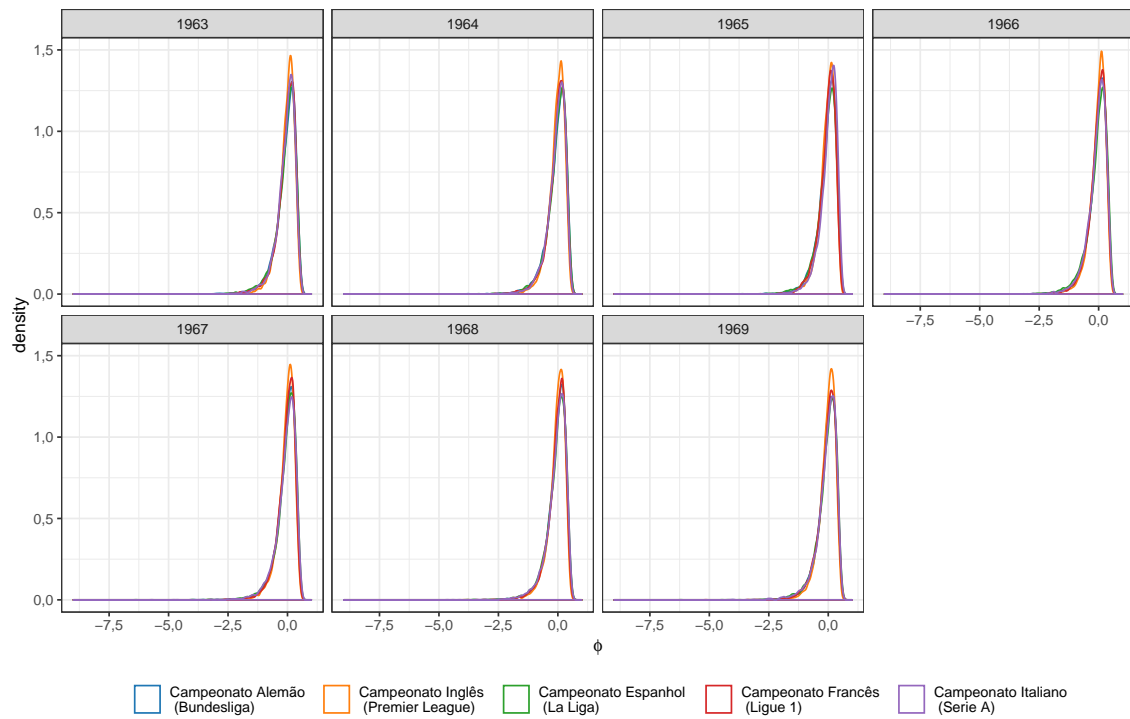


Figura A.1: A densidade do Teste de Monte Carlo para os cinco campeonatos (Década de 1960)

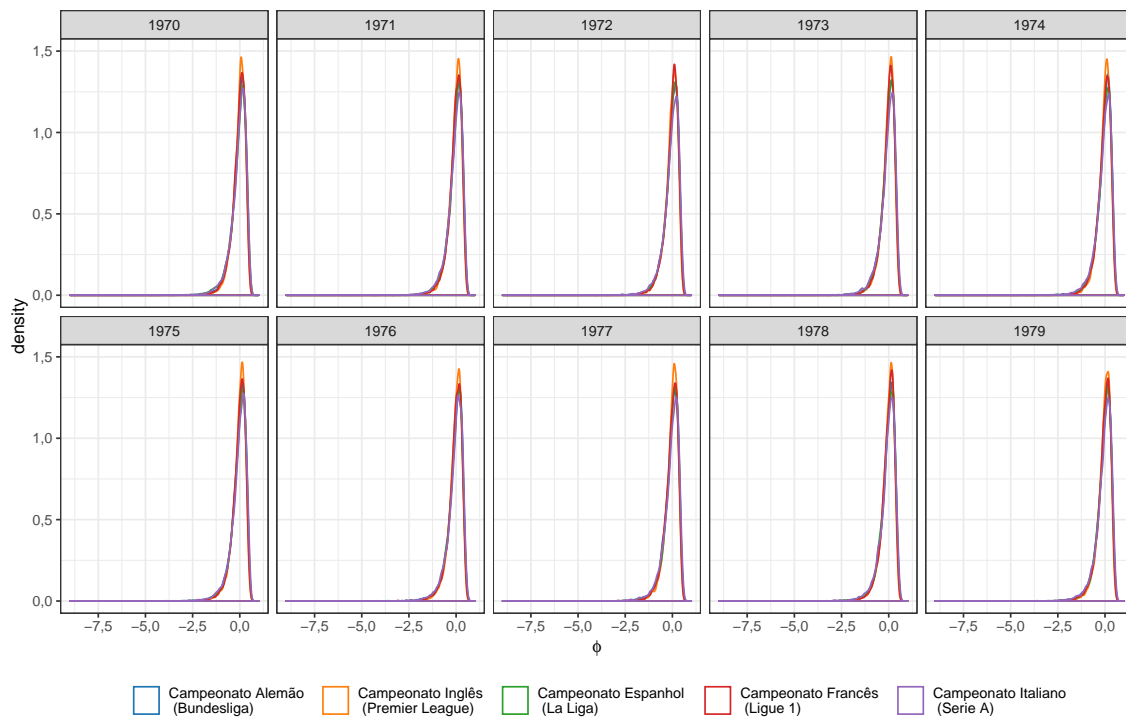


Figura A.2: A densidade do Teste de Monte Carlo para os cinco campeonatos (Década de 1970)

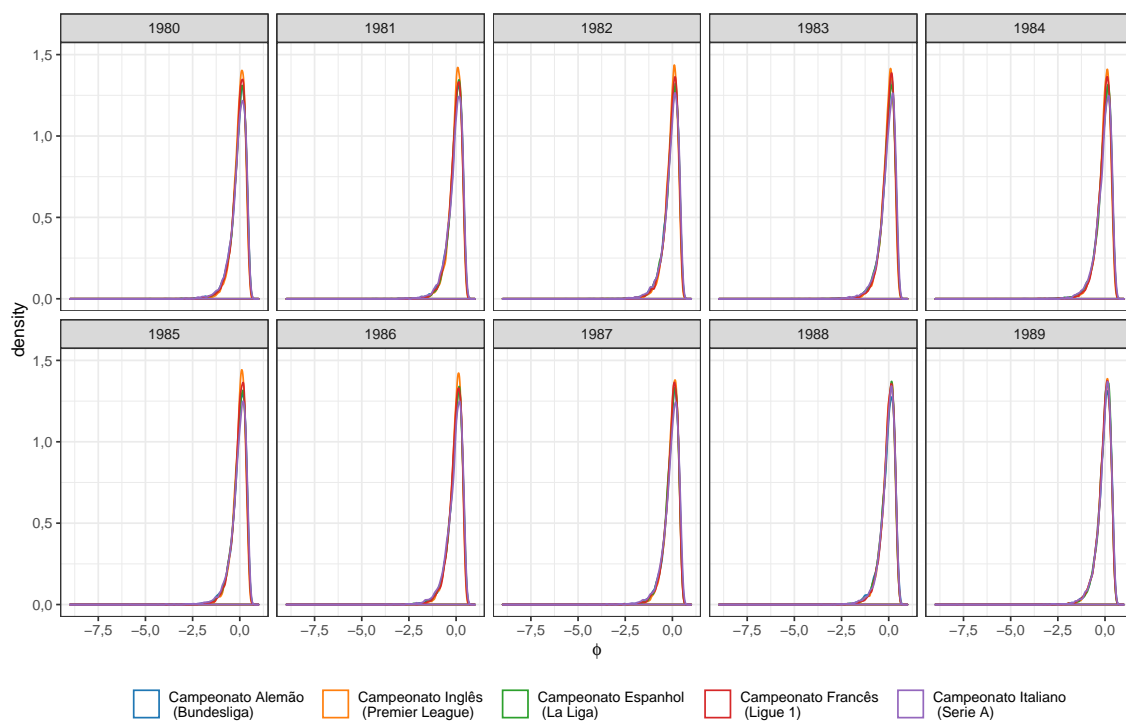


Figura A.3: Densidade do Teste de MC (Década de 1980)



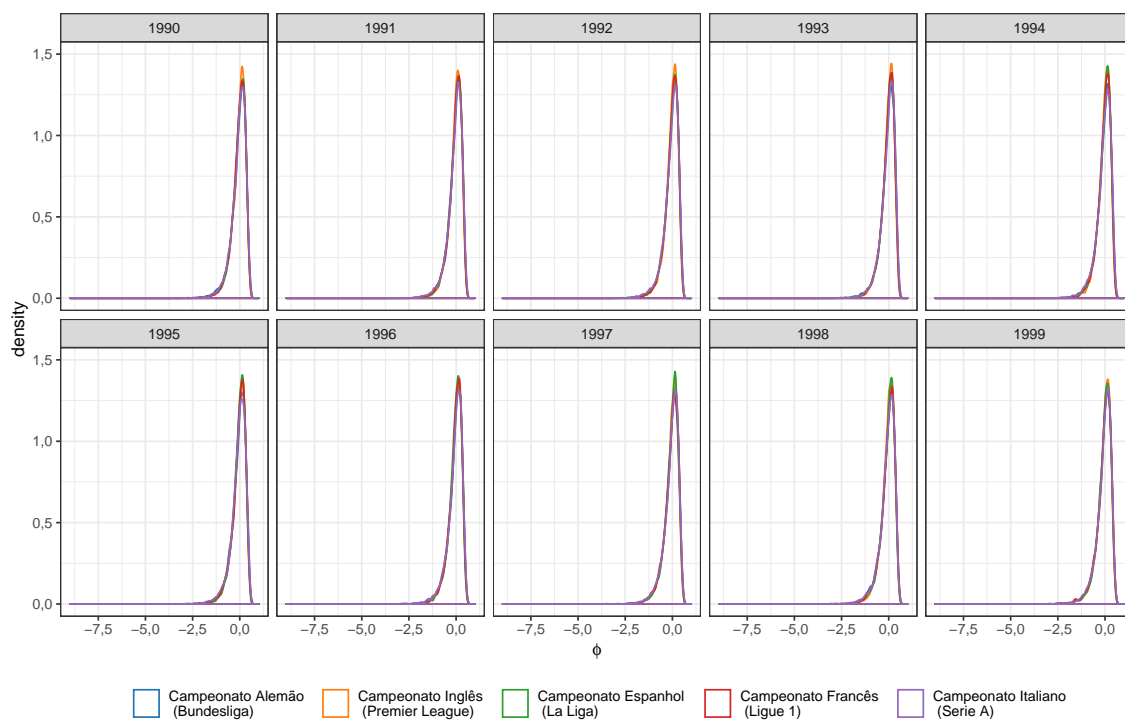


Figura A.4: A densidade do Teste de Monte Carlo para os cinco campeonatos (Década de 1990)

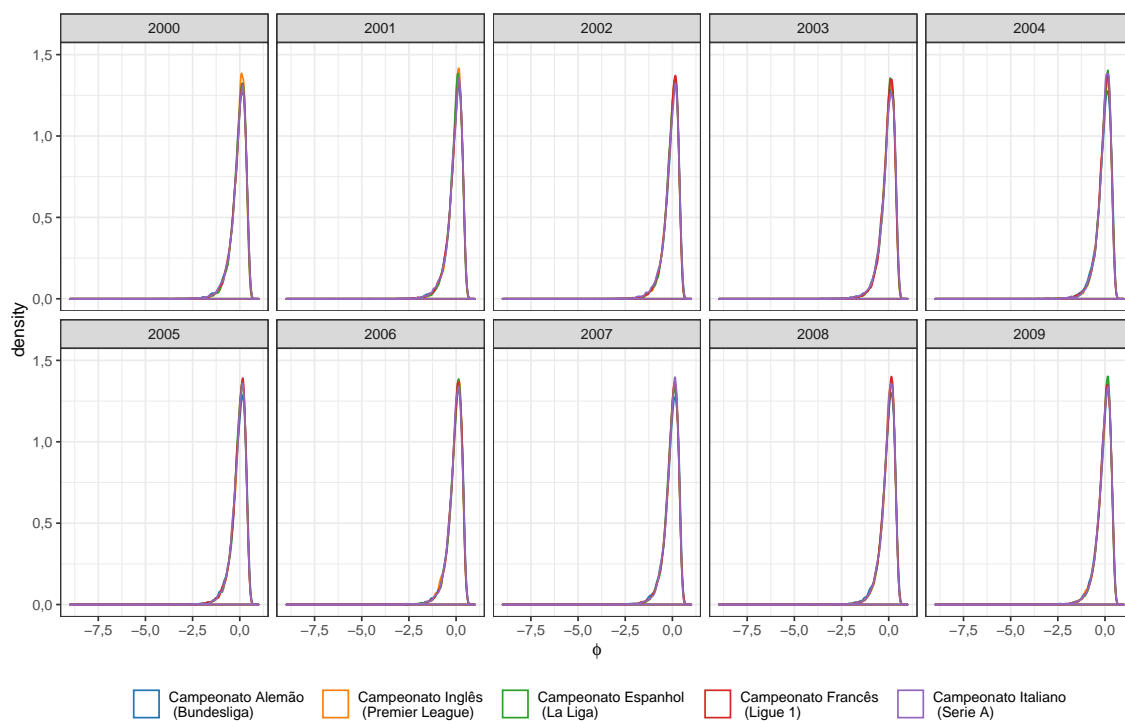


Figura A.5: A densidade do Teste de Monte Carlo para os cinco campeonatos (Década de 2000)

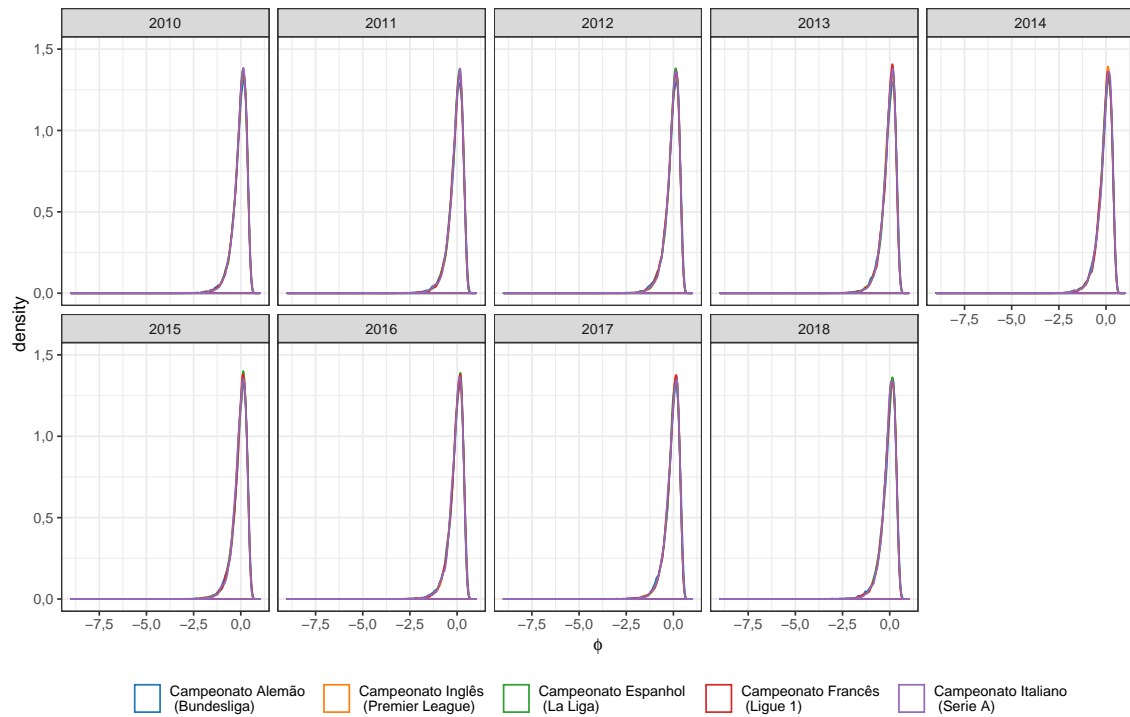


Figura A.6: A densidade do Teste de Monte Carlo para os cinco campeonatos (Década de 2010)

## B Estatísticas dos jogadores por partida

O conjunto de dados original obtido através do *webscraping* consiste em 31 variáveis referente às estatísticas dos jogadores na partida, uma variável de identificação do nome do jogador e uma do número do jogo. Suas descrições e definições encontradas em [OptaSports \(2019b\)](#) estão listadas abaixo:

- **Player:** nome do jogador,
- **Touches:** quantidade de toques dado pelo jogador. É a soma de todos os eventos em que um jogador toca a bola, exceto por bola aérea ou duelos perdidos,
- **TotalTackles:** quantidade de desarmes ganhos pelo jogador,
- **Interceptions:** quantidade de interceptações feitas pelo jogador,
- **Clearances:** quantidade de vezes que o jogador afastou a bola para longe do gol que está defendendo,
- **BlockedShots:** quantidade de vezes que o jogador – que está à frente do goleiro ou outros defensores – bloqueia um chute. Inclui bloqueios involuntários feitos em chutes dado jogador do próprio time,
- **Fouls:** quantidade de faltas cometidas pelo jogador,
- **Shots:** quantidade total de chutes feito pelo jogador,
- **ShotsOT:** quantidade de chutes no “alvo” feito pelo jogador. É considerado um chute no alvo quando: 1) é gol, independente da intenção do jogador e 2) uma tentativa clara de marcar que teria sido convertida se não fosse salva pelo goleiro ou bloqueado por um jogador que é o “último homem”,
- **KeyPasses:** quantidade de passes chave. É definido como o “‘passe final” feito de um jogador para outro que tenta fazer o gol, mas não marca,
- **Dribbles:** quantidade de dribles bem-sucedidos. É considerado drible quando o jogador está com a bola e vence o adversário, mantendo assim a posse,
- **Offsides:** quantidade de vezes que o jogador esteve em posição de impedimento,
- **Disp:** quantidade de vezes que o um jogador perdeu a posse de bola para um oponente, sem que tenha ocorrido um drible,
- **UnsTouches:** quantidade de vezes que o jogador controla mal a bola com um toque ruim,
- **Passes:** quantidade de passes intencionais feitos de um jogador para outro. Aqui desconsidera-se cruzamentos, passes chave e assistências, por exemplo,
- **PA.:** quantidade de passes bem-sucedidos feito pelo jogaoadr,

- **Crosses**: quantidade de cruzamentos feito pelo jogador. É qualquer bola jogada intencionalmente de uma posição ampla com a intenção de alcançar um companheiro de equipe em uma área específica na frente do gol,
- **AccCrosses**: quantidade de cruzamentos bem-sucedidos feito pelo jogador,
- **LB**: quantidade total de passes longos feito pelo jogador,
- **AccLB**: quantidade de passes longos bem-sucedidos feito pelo jogador,
- **ThB**: quantidade de passes que corta a defesa e o jogador do mesmo time precisa se movimentar para receber a bola,
- **AccThB**: quantidade de passes bem-sucedidos que corta a defesa e o jogador do mesmo time precisa se movimentar para receber a bola,
- **Mode**: variável dicotômica que indica se o time do jogador é o mandante (*Home*) ou visitante (*Away*) do jogo em questão,
- **Team**: clube em que o jogador atua,
- **Played**: variável dicotômica indicando se o jogador jogou a partida ou não,
- **Sub In**: variável dicotômica indicando se o jogador entrou no jogo por substituição ou não,
- **Sub Out**: variável dicotômica indicando se o jogador saiu do jogo por substituição ou não,
- **Position**: posição que o jogador atua, Tabelas
- **Age**: idade do jogador,
- **Min**: quantidade de minutos jogados pelo jogador na partida. É considerado o máximo de 90 minutos. Se um jogador entrou nos acréscimos, então são desconsiderados os minutos,
- **Game**: variável para identificar o número do jogo.

## C Código em R do Processo de *Webscraping*

O *webscraping* é uma forma de mineração de dados que consiste em baixar dados contidos em sites da *web* através de um código desenvolvido em uma linguagem de programação, neste caso em R, para capturar as informações utilizando o navegador *web* de forma automatizada (vanden Broucke e Baesens, 2018).

O código desenvolvido para obter os dados do site WhoScored (2019) é disponibilizado abaixo. Os pacotes utilizados foram o *RSelenium* (Harrison e Kim, 2019) e *XML* (Lang, 2019).

```
require(RSelenium) # utilizado para fazer o webscraping
require(XML) #utilizado para pegar os dados em uma tabela

# Abre o navegador da web
driver <- rsDriver(port = 4597L,
                   browser = c('chrome'), # navegador usado
                   chromeversion = "78.0.3904.70") # versao

# banco de dados contendo os links das 380 partidas no site
# com os dados das estatisticas dos jogadores
urls = read.table("Espanhol20172018.txt", header = FALSE)

# lista que salva todos os bancos com as estatisticas dos
# jogadores para cada partida
bancos <- vector("list", nrow(urls))

for (rep in 1:nrow(urls)){
url = as.character(urls[rep,]) # pega a url 'rep'

remDr <- driver[['client']]
remDr$navigate(url) # carrega o link
# da uma pausa de 3 segundos para esperar carregar
Sys.sleep(3)

# Procura um elemento na pagina.
webElem <- remDr$findElement(using = "xpath",
                              value = '//*[@id="content-2col-left"]/h2[1]/a')

# pega o codigo fonte do que foi pego em webElem e
# depois coloca as informacoes organizadas em uma lista
doc <- htmlParse(webElem$getPageSource()[[1]])
table_tmp <- readHTMLTable(doc)

# os proximos codigos fazem os mesmos procedimento
# pegando as informacoes dos quatro tipos de tabelas
# que tem para cada time:
# Summary, Offensive, Defensive e Passing

webElem <- remDr$findElement(using = "xpath",
                              value = '//*[@id="statistics-table-home-summary"]')
Sys.sleep(1)
```

```

doc <- htmlParse(webElem$pageSource()[[1]])
table_tmp <- readHTMLTable(doc)
Sys.sleep(1)

webElems <- remDr$findElements(using = "xpath",
                              value = '//*[@id="live-player-home-options"]/li[2]/a')
resHeaders <- unlist(lapply(webElems,
                            function(x) {x$getElementText()}))
webElem <- webElems[[which(resHeaders == resHeaders)]]
Sys.sleep(3)

webElem <- remDr$findElement(using = "xpath",
                             value = '//*[@id="statistics-table-home-offensive"]')
Sys.sleep(1)

webElems <- remDr$findElements(using = "xpath",
                              value = '//*[@id="live-player-away-options"]/li[2]/a')
resHeaders <- unlist(lapply(webElems,
                            function(x) {x$getElementText()}))
webElem <- webElems[[which(resHeaders == resHeaders)]]
Sys.sleep(3)

webElem <- remDr$findElement(using = "xpath",
                             value = '//*[@id="statistics-table-away-offensive"]')
Sys.sleep(1)

doc <- htmlParse(webElem$pageSource()[[1]])
table_tmp <- readHTMLTable(doc)
Sys.sleep(1)

webElems <- remDr$findElements(using = "xpath",
                              value = '//*[@id="live-player-home-options"]/li[3]/a')
resHeaders <- unlist(lapply(webElems,
                            function(x) {x$getElementText()}))
webElem <- webElems[[which(resHeaders == resHeaders)]]
Sys.sleep(3)
webElem <- remDr$findElement(using = "xpath",
                             value = '//*[@id="statistics-table-home-defensive"]')
Sys.sleep(1)

webElems <- remDr$findElements(using = "xpath",
                              value = '//*[@id="live-player-away-options"]/li[3]/a')
resHeaders <- unlist(lapply(webElems,
                            function(x) {x$getElementText()}))
webElem <- webElems[[which(resHeaders == resHeaders)]]
Sys.sleep(3)
webElem <- remDr$findElement(using = "xpath",
                             value = '//*[@id="statistics-table-away-defensive"]')
Sys.sleep(1)
doc <- htmlParse(webElem$pageSource()[[1]])
table_tmp <- readHTMLTable(doc)

```

```

Sys.sleep(1)

webElems <- remDr$findElements(using = "xpath",
                              value = '//*[@id="live-player-home-options"]/li[4]/a')
resHeaders <- unlist(lapply(webElems,
                           function(x) {x$getElementText()}))
webElem <- webElems[[which(resHeaders == resHeaders)]]
Sys.sleep(3)
webElem <- remDr$findElement(using = "xpath",
                             value = '//*[@id="statistics-table-home-passing"]')
Sys.sleep(1)

webElems <- remDr$findElements(using = "xpath",
                              value = '//*[@id="live-player-away-options"]/li[4]/a')
resHeaders <- unlist(lapply(webElems,
                           function(x) {x$getElementText()}))
webElem <- webElems[[which(resHeaders == resHeaders)]]
Sys.sleep(3)
webElem <- remDr$findElement(using = "xpath",
                             value = '//*[@id="statistics-table-away-passing"]')

Sys.sleep(3)

doc <- htmlParse(webElem$pageSource()[[1]])
table_tmp <- readHTMLTable(doc)
Sys.sleep(1)

print(rep)

# salva a lista completa com todas as tabelas
bancos[[rep]] <- table_tmp

}

# cria um banco de dados unico para a partida com todas as
# estatisticas dos dois times em uma unica tabela
for(rep in 1:nrow(urls)){

stats = as.data.frame(NULL) # nome do banco

# nome do time
Team <- c(rep(as.character(bancos[[rep]][[1]]$V1)[1],
              times = nrow(bancos[[rep]][[2]])),
         rep(as.character(bancos[[rep]][[1]]$V3)[1],
              times = nrow(bancos[[rep]][[10]])))

# se o time jogou em casa (home) ou fora (away)
Mode <- c(rep("Home", times = nrow(bancos[[rep]][[2]])),
         rep("Away", times = nrow(bancos[[rep]][[10]])))

# as estatisticas da lista do for anterior, retirando

```

```
# as repetidas que tem entre uma tabela e outra
aux = rbind(cbind(bancos[[rep]][[2]][,c(3,8:9)],
                 bancos[[rep]][[4]][,c(4:8)],
                 bancos[[rep]][[6]][,c(4:11)],
                 bancos[[rep]][[8]][,c(5:13)]),
            cbind(bancos[[rep]][[10]][,c(3,8:9)],
                 bancos[[rep]][[12]][,c(4:8)],
                 bancos[[rep]][[14]][,c(4:11)],
                 bancos[[rep]][[16]][,c(5:13)]))

stats <- as.data.frame(aux)

stats$Mode <- Mode
stats$Team <- Team

# salvando em txt o banco final
txt.name = paste0(rep, "_Espanhol_2017.txt", collapse = "")
write.table(x = stats, file = txt.name, sep = ",")

}
```



## D Habilidade das Equipes

### D.1 Premier League

Na temporada de 2018-19 da *English Premier League* houveram  $k + 1 = 20$  equipes, com cada time disputando 38 jogos, sendo metade disputado como mandante e a outra metade como visitante. O número total de partidas da edição foi de  $N = 380$  e ocorreu entre 10 de agosto de 2018 e 12 de maio de 2019, em 99 dias distintos. O campeão foi o *Manchester City* com 98 pontos (32V, 2E e 4D) e os rebaixados foram o *Cardiff* (10V, 4E e 24D), *Fulham* (7V, 5E e 26D) e *Huddersfield* (3V, 7E e 28D).

#### D.1.1 Modelo de Bradley-Terry

Fica evidente pela [Figura D.1](#) que as habilidades estimadas através do modelo de Bradley-Terry indica que o *Manchester City* e o *Liverpool* possuem habilidades muito diferente dos demais. É possível observar que, quando visto de forma ordenada, os clubes que estão no meio da tabela, no geral, possuem estimativas muito parecidas.

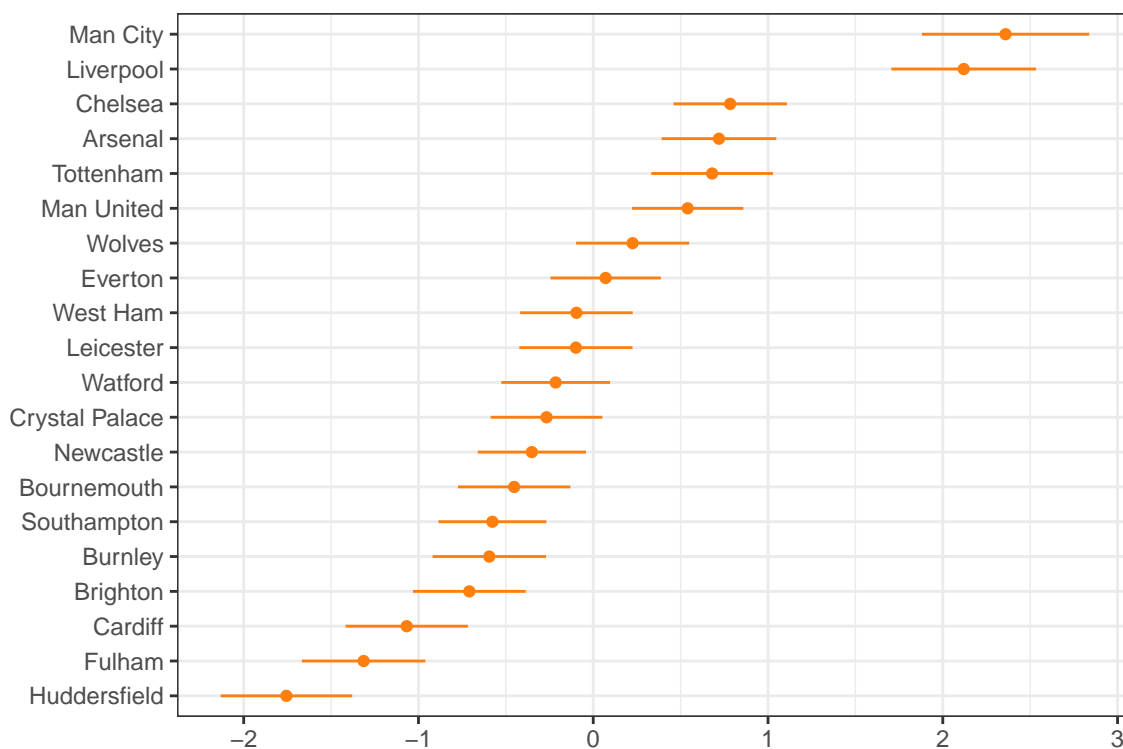


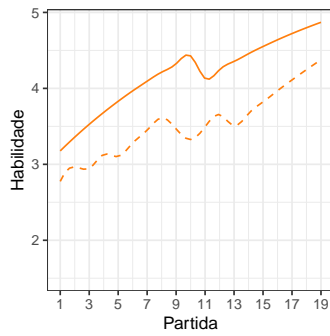
Figura D.1: Habilidades e QSE estimados das 20 equipes da edição de 2018 da *Premier League* pelo modelo de Bradley-Terry

O *Huddersfield*, lanterna da edição, possui habilidade estimada na edição abaixo de todas as equipes que ficaram fora da zona de rebaixamento ao final do campeonato. A campanha do time foi de apenas 3 vitórias e 7 empates entre os 38 jogos, com aproveitamento final de apenas 14,03%, obtendo 16 pontos ao final do campeonato.

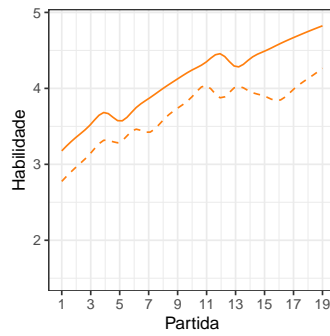
#### D.1.2 Modelo de Bradley-Terry Dinâmico

Os parâmetros estimados de suavização específicos do mandante e visitante,  $\hat{\lambda}_1 = 0,067$  e  $\hat{\lambda}_2 = 0,040$ , são valores bem próximos de zero. O vetor de parâmetros de

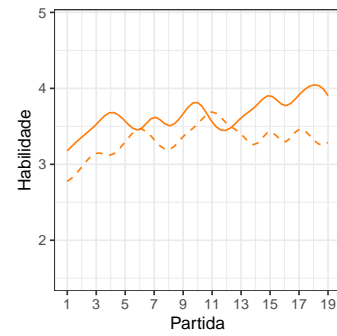
interesse estimado foi de  $\hat{\gamma} = (\hat{\beta}_1; \hat{\beta}_2; \hat{\delta})^T = (1,954; 2,494; 0,839)^T$ , com erros padrão de 0,286, 0,400 e 0,110, respectivamente.



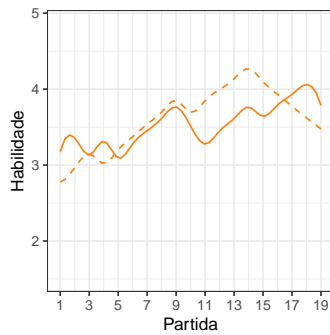
(a) Man City



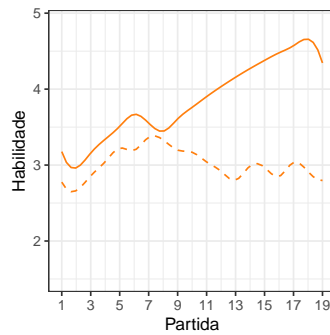
(b) Liverpool



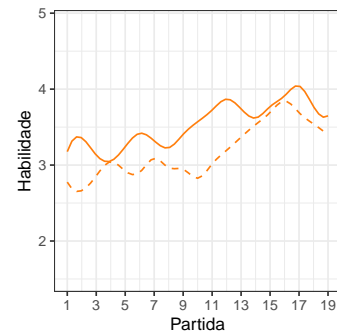
(c) Chelsea



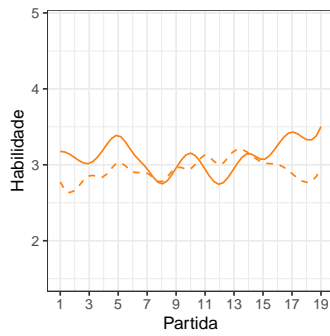
(d) Tottenham



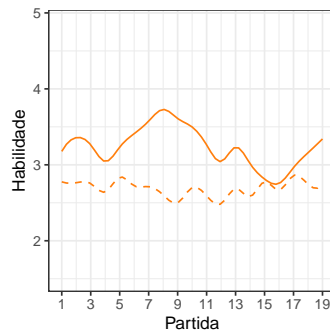
(e) Arsenal



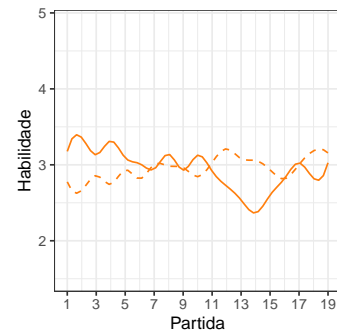
(f) Man United



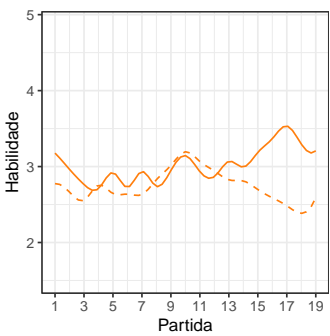
(g) Wolves



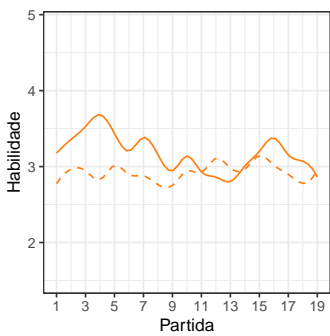
(h) Everton



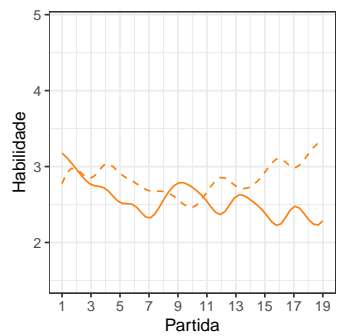
(i) Leicester



(j) West Ham



(k) Watford



(l) Crystal Palace

Figura D.2: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da *Premier League* - 1º ao 12º colocado

Os gráficos relacionados às habilidades suavizadas em termos de mandante e visitante para as equipes da EPL ao longo das partidas estão dispostos na [Figura D.2](#). Nota-se que a evolução temporal do *Manchester City* e do *Liverpool* é crescente para ambos tanto nas partidas dentro, quanto fora de casa. O *Tottenham*, entre as partidas 10 e 16, obteve uma habilidade estimada como visitante maior do que como mandante, o que não é muito comum de ocorrer, visto que sabe-se que há uma vantagem relacionada ao mando de campo no futebol. O mesmo ocorreu com o *Crystal Palace* na segunda metade do campeonato. O aproveitamento do clube foi de 42,98%, sendo que quando jogou no próprio estádio esse aproveitamento caiu para 35,09% e como visitante fica em 50,88%, acima do geral.

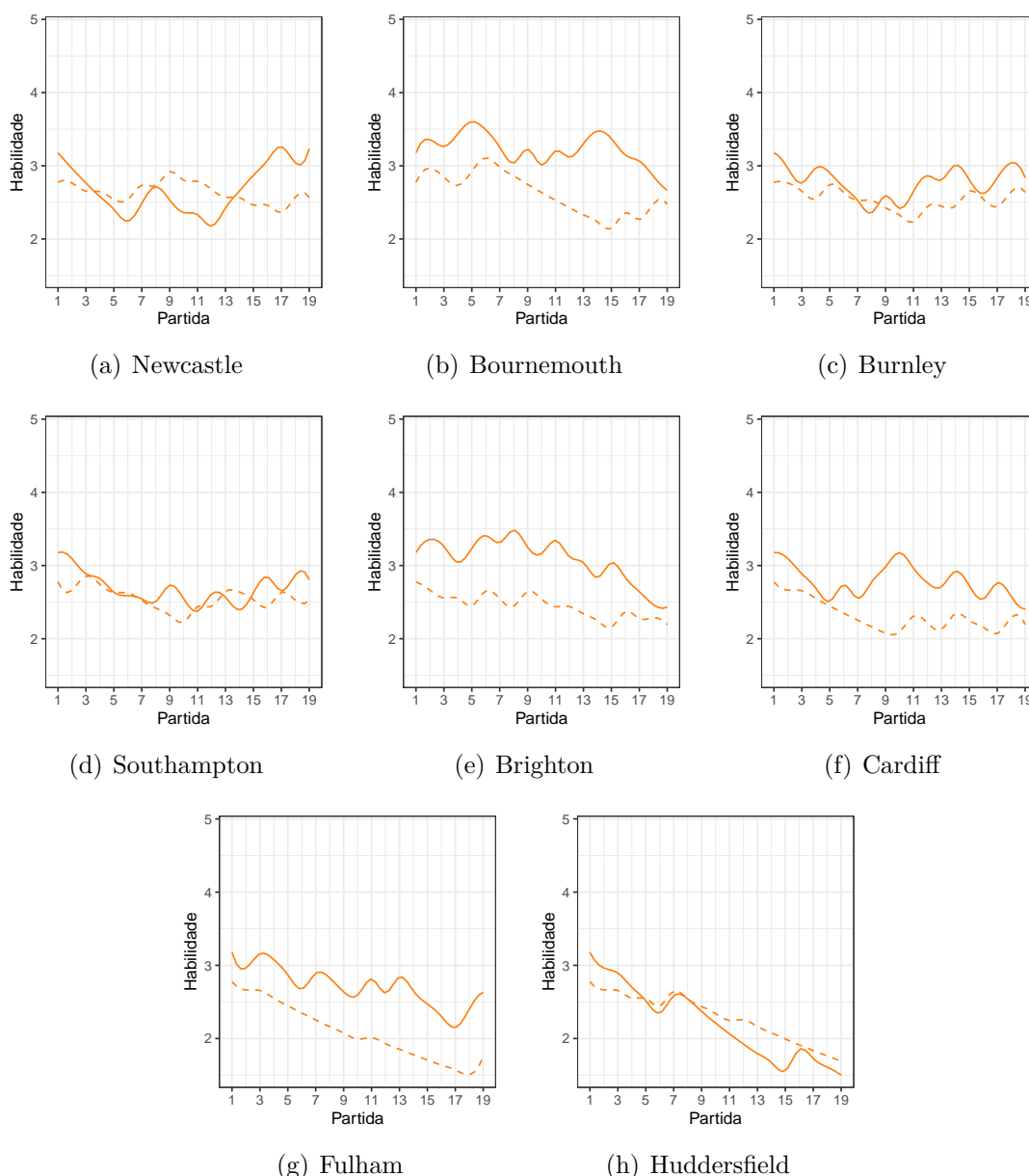


Figura D.3: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da *Premier League* - 13<sup>o</sup> ao 20<sup>o</sup> colocado

A habilidade estimada do *Arsenal* como mandante teve uma ascensão a partir da 8<sup>a</sup> partida, em que a equipe só obteve vitórias no *Emirates Stadium*, com exceção das

duas últimas, que foram uma derrota e um empate, respectivamente. As equipes da parte de baixo da tabela ao final do campeonato terão habilidades estimadas menores porque o desempenho delas foi inferior. A [Figura D.3](#) fornece as estimativas para os oito piores clubes da EPL.

O *Burnley* e o *Southampton* não parecem ter habilidades que se distinguem independentemente se estão jogando a partida como mandante ou visitante. Com relação ao *Fulham* e ao *Huddersfield*, é possível notar a queda brusca nas estimativas fora de casa e dentro de casa, respectivamente. O *Fulham* como visitante só obteve uma vitória, que ocorreu na 18ª rodada contra o *Wolves*. As duas vitórias do *Huddersfield* podem ser facilmente identificadas no gráfico porque é justamente onde tem os picos.

### D.1.3 Modelo Estático x Modelo Dinâmico

Uma comparação das habilidades estimadas através do modelo estático e as habilidades médias estimadas para o campeonato em termos gerais, mandante de visitante estão dispostas abaixo na [Tabela 1](#).

Tabela 1: Tabela resumo das habilidades e QSE estimados pelo modelo BT-Estático e as habilidades suavizadas médias pelo BT-Dinâmico em termos gerais, como mandante e visitante da *Premier League* de 2018

Equipe	Modelo Estático			Modelo Dinâmico					
	Habilidade	QSE	Rank	Habilidade	Rank	Habilidade (Mandante)	Rank	Habilidade (Visitante)	Rank
1. Man City	2,359	0,479	1	0,850	2	1,023	1	0,677	3
2. Liverpool	2,120	0,414	2	0,886	1	0,954	2	0,819	1
3. Chelsea	0,784	0,325	3	0,479	4	0,493	4	0,465	4
4. Tottenham	0,680	0,348	5	0,565	3	0,373	5	0,757	2
5. Arsenal	0,719	0,328	4	0,425	5	0,686	3	0,163	6
6. Man United	0,540	0,319	6	0,350	6	0,372	6	0,328	5
7. Wolves	0,225	0,323	7	0,027	8	-0,037	10	0,091	8
8. Everton	0,071	0,316	8	-0,035	9	0,081	7	-0,152	12
9. Leicester	-0,099	0,323	10	-0,049	10	-0,204	13	0,106	7
10. West Ham	-0,096	0,322	9	-0,116	12	-0,129	12	-0,103	11
11. Watford	-0,215	0,311	11	0,048	7	0,018	9	0,079	9
12. Crystal Palace	-0,267	0,319	12	-0,289	14	-0,587	19	0,010	10
13. Newcastle	-0,351	0,310	13	-0,331	15	-0,455	17	-0,206	13
14. Bournemouth	-0,452	0,321	14	-0,082	11	0,059	8	-0,223	14
15. Burnley	-0,595	0,325	16	-0,334	16	-0,373	15	-0,296	16
16. Southampton	-0,577	0,309	15	-0,371	17	-0,455	17	-0,286	15
17. Brighton	-0,709	0,323	17	-0,235	13	-0,078	11	-0,393	17
18. Cardiff	-1,067	0,350	18	-0,448	18	-0,367	14	-0,529	18
19. Fulham	-1,314	0,353	19	-0,593	19	-0,431	16	-0,756	20
20. Huddersfield	-1,756	0,377	20	-0,747	20	-0,943	20	-0,552	19

O modelo estático parece mais relacionado com a classificação final pela quantidade de pontos no campeonato. Quando analisado as estimativas médias para o modelo dinâmico, observa-se uma menor relação. Um exemplo é o *Watford*, que terminou o torneio na 11ª colocação, mas no *ranking* dinâmico ficou em 7ª. Em termos de habilidade média como mandante, o *Bournemouth* ficou em 8ª, apesar de ter finalizado a temporada em 14ª. Em termos de *ranking* pior do que a posição final, o *Everton* obteve um desempenho modesto como visitante (5V, 5E e 9D), ficando em 12ª pela classificação do modelo.

### D.1.4 Avaliação do modelo

Para avaliação foi utilizado o RPS. O modelo foi ajustado para os dados até o 48º dia de campeonato, que equivale ao primeiro turno do campeonato inglês dessa edição. Os resultados das partidas a partir do 49º dia são preditos e comparados com os resultados observados pelo RPS. O modelo então é reajustado incluindo também as partidas do dia 49 e é utilizado pra prever os resultados do dia 50 e assim por diante até o último dia. As previsões são comparadas também com as proporções empíricas de vitórias e derrotas do time mandante, computada pelas partidas dos primeiros 48 dias.

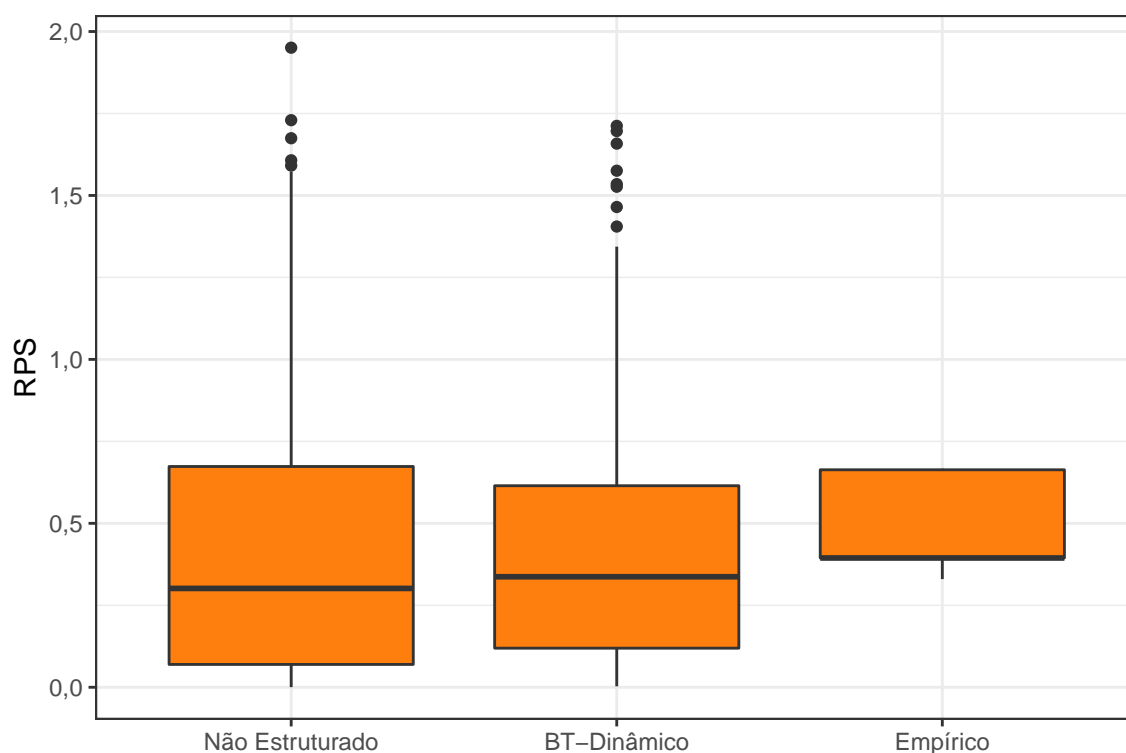


Figura D.4: Boxplot do RPS para a *Premier League* de 2018 computado para o modelo não-estruturado, BT-Dinâmico e previsões baseadas em proporções empíricas

O RPS para o modelo dinâmico foi de 0,423 e o do não-estruturado foi de 0,349, 17,5% menor. O boxplot da Figura D.4 corresponde às estimativas do RPS obtidas para as partidas preditas. Observa-se que novamente o modelo BT-Dinâmico ficou similar ao não-estruturado, enquanto que o empírico é pior.

## D.2 *La Liga*

Na temporada de 2018-19 da *La Liga* houveram  $k + 1 = 20$  equipes, com cada time disputando 38 jogos, sendo metade disputado como mandante e a outra metade como visitante. O número total de partidas da edição foi de  $N = 380$  e ocorreu entre 17 de agosto de 2018 e 26 de maio de 2019, em 129 dias distintos. O campeão foi o *Barcelona* com 87 pontos (26V, 9E e 3D) e os rebaixados foram o *Girona* (9V, 10E e 19D), *SD Huesca* (7V, 12E e 19D) e *Vallecano* (8V, 8E e 22D).

### D.2.1 Modelo de Bradley-Terry

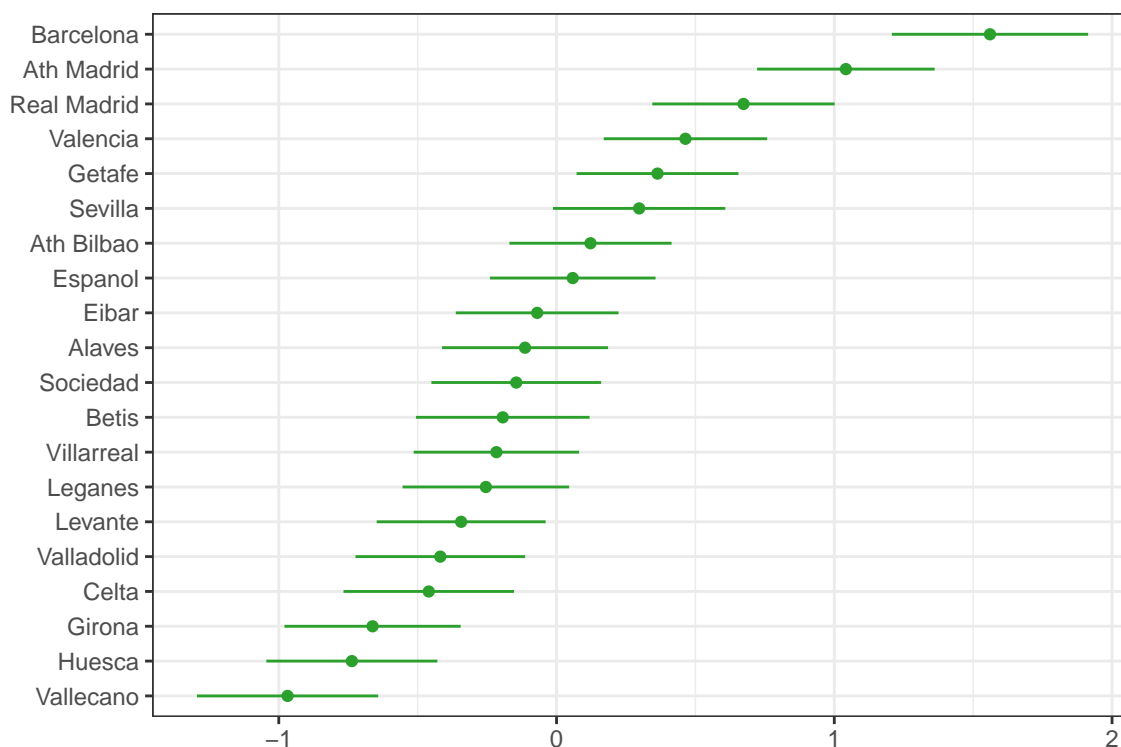


Figura D.5: Habilidades e QSE estimados das 20 equipes da edição de 2018 da *La Liga* pelo modelo de Bradley-Terry

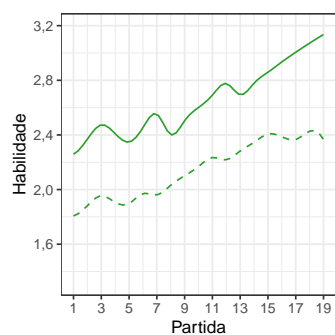
A Figura D.5 exposta acima fornece as estimativas para a habilidade das equipes na edição de 2018 do campeonato espanhol. O *Barcelona* possui habilidade muito superior aos demais, com exceção a do *Athletico Madrid*, que foi o vice-campeão. O *Vallecanao* teve a pior campanha do torneio, marcando apenas 32 pontos, um aproveitamento de 28,07%, não muito diferente do *Huesca* (28,95%).

Parece que os clubes do meio da tabela não diferem muito nas habilidades estimadas, o que faz sentido com o que foi observado no Capítulo 3, que mostrava que na retirada das equipes que destoavam o campeonato e os deixavam não aleatório, os clubes que ficavam em posições médias na classificação final dificilmente saíam (ver Figura 3.11) porque normalmente a diferença na habilidade dos times dentro de um torneio ocorre nos extremos.

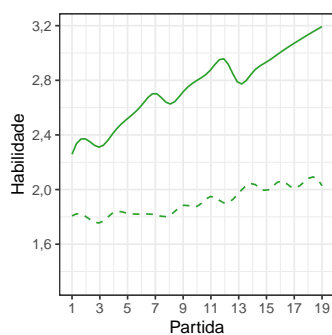
### D.2.2 Modelo de Bradley-Terry Dinâmico

Os parâmetros estimados de suavização específicos do mandante e visitante,  $\hat{\lambda}_1 = 0,059$  e  $\hat{\lambda}_2 = 0,025$ , foram valores bem próximos de zero. O vetor de parâmetros de interesse estimado foi de  $\hat{\gamma} = (\hat{\beta}_1; \hat{\beta}_2; \hat{\delta})^T = (1,378; 1,594; 0,455)^T$ , com erros padrão de 0,375, 0,527 e 0,086, respectivamente.

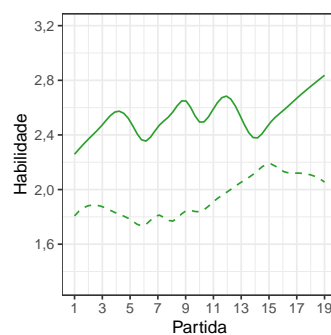
O modelo estático estima habilidades para as equipes em cada partida em termos de visitante e mandante separadamente. Os primeiros 12 colocados no campeonato espanhol estão representados na Figura D.6.



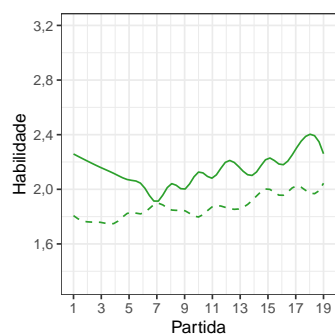
(a) Barcelona



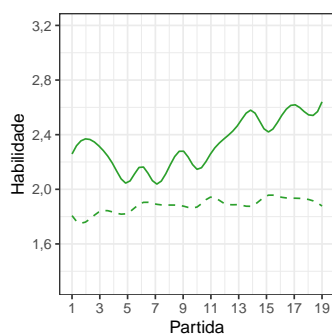
(b) Atletico Madrid



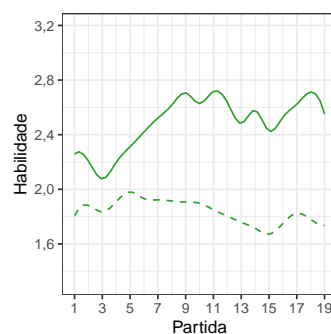
(c) Real Madrid



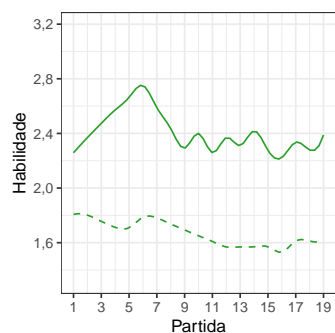
(d) Valencia



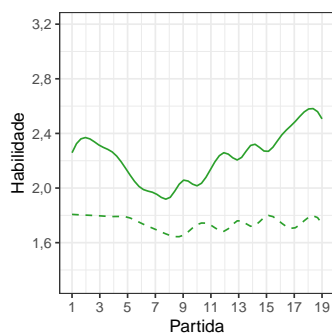
(e) Getafe



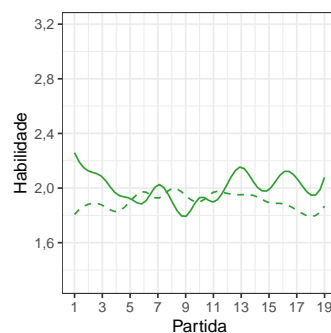
(f) Sevilla



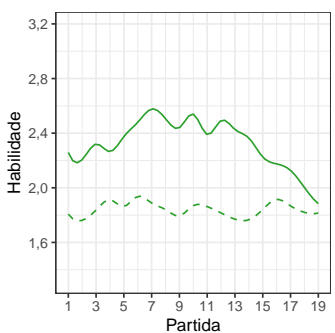
(g) Espanyol



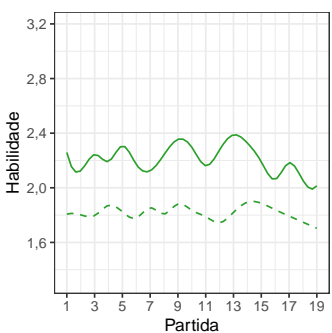
(h) Athletic Bilbao



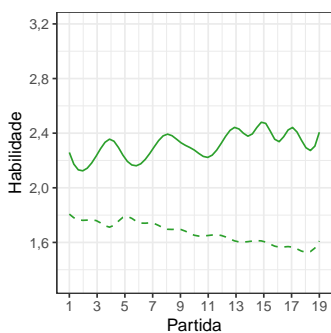
(i) Real Sociedad



(j) Alaves



(k) Betis



(l) Eibar

Figura D.6: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da *Premier League* - 1<sup>o</sup> ao 12<sup>o</sup> colocado

O que chama atenção nesse gráfico é que está muito mais visível a habilidade inferior quando os clubes estão disputando a partida como visitante. Na *Bundesliga* e na *Premier League* existiam mais intersecções nas habilidades estimadas das equipes. Observa-se que somente o *Real Sociedad*, que terminou o torneio em 9º lugar, obteve em algum momento do tempo habilidade superior fora de casa. A campanha do *Espanyol* como mandante (11V, 3E e 5D) e visitante (3V, 8E e 8D) explica porque as duas habilidades estimadas foram tão diferentes uma da outra. O aproveitamento do *Atletico* foi de 84,2% e 51,85% como mandante e visitante, respectivamente. As habilidades ao longo do campeonato demonstram essa diferença pois as curvas suavizadas estão afastadas uma da outra.

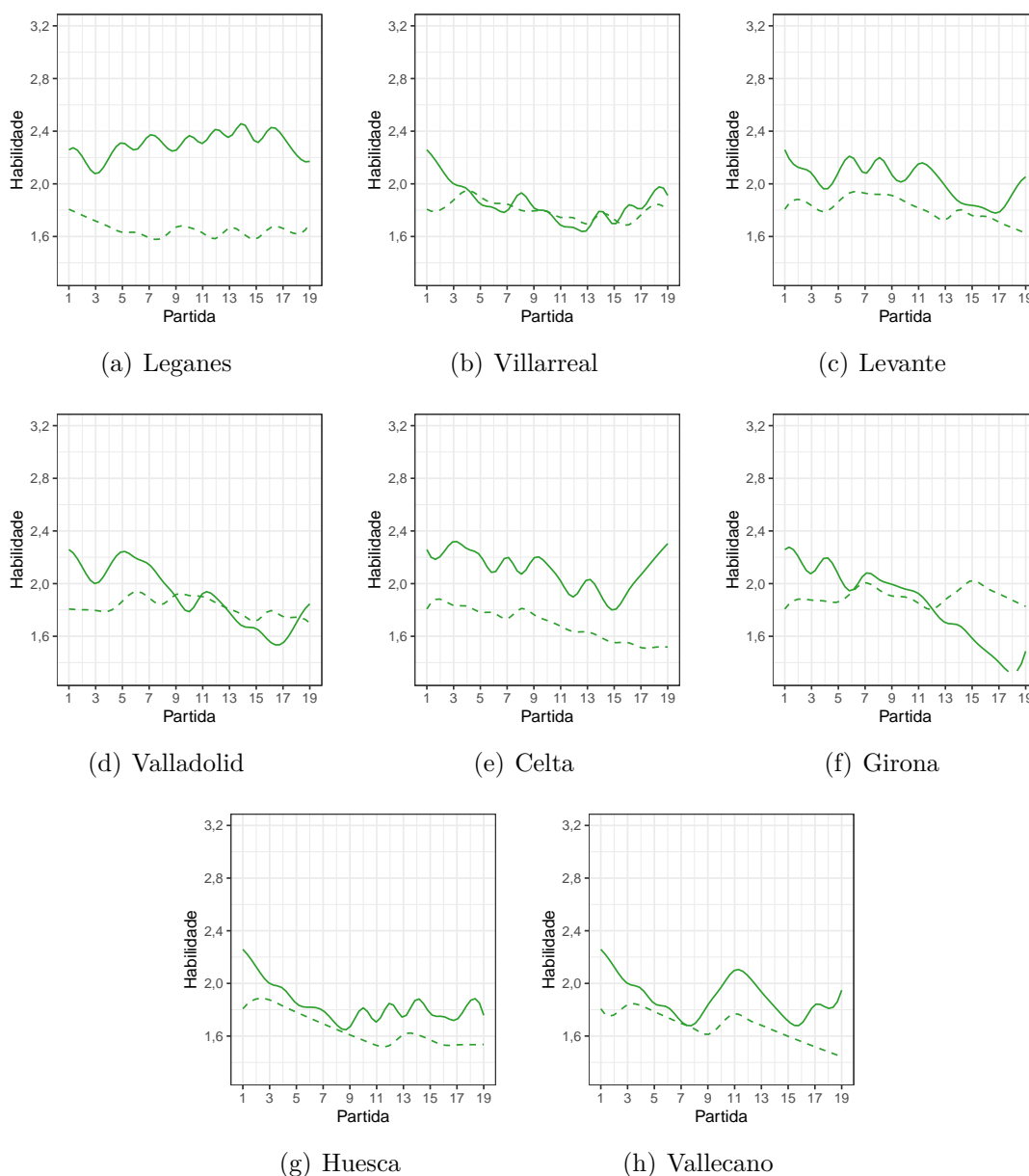


Figura D.7: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da *La Liga* - 13º ao 20º colocado

Nas equipes da ponta de baixo da tabela, observa-se pela [Figura D.7](#) que a diferença nas habilidades dentro e fora de casa não parecem ser tão diferentes, como é



o caso do *Villarreal* e do *Levante*. O declínio da habilidade do *Girona* como mandante condiz com o fato de que o clube obteve a pior campanha do campeonato em termos de pontuação dentro de casa, tendo somado apenas 15 pontos.

### D.2.3 Modelo Estático x Modelo Dinâmico

Os modelos BT-estático e o de BT-dinâmico geralmente chegam na mesma conclusão, porém o primeiro tende a concordar mais com a classificação final do campeonato.

Tabela 2: Tabela resumo das habilidades e QSE estimados pelo modelo BT-Estático e as habilidades suavizadas médias pelo BT-Dinâmico em termos gerais, como mandante e visitante da *La Liga* de 2018

Equipe	Modelo Estático			Modelo Dinâmico					
	Habilidade	QSE	Rank	Habilidade	Rank	Habilidade (Mandante)	Rank	Habilidade (Visitante)	Rank
1. Barcelona	1,561	0,353	1	0,385	1	0,438	2	0,331	1
2. Ath Madrid	1,042	0,320	2	0,316	2	0,534	1	0,098	3
3. Real Madrid	0,673	0,328	3	0,227	3	0,326	3	0,128	2
4. Valencia	0,464	0,294	4	0,001	8	-0,062	12	0,065	7
5. Getafe	0,364	0,291	5	0,098	5	0,127	6	0,069	6
6. Sevilla	0,297	0,310	6	0,149	4	0,274	4	0,024	9
7. Espanol	0,058	0,298	8	0,025	7	0,193	5	-0,144	18
8. Ath Bilbao	0,122	0,292	7	-0,023	11	0,022	10	-0,068	14
9. Sociedad	-0,145	0,306	11	-0,055	13	-0,200	15	0,090	4
10. Betis	-0,193	0,313	12	0,001	8	-0,001	11	0,003	11
11. Alaves	-0,113	0,299	10	0,069	6	0,109	7	0,030	8
12. Eibar	-0,070	0,293	9	-0,017	10	0,106	8	-0,140	17
13. Leganes	-0,254	0,300	14	-0,035	12	0,086	9	-0,155	19
14. Villarreal	-0,216	0,298	13	-0,181	18	-0,350	18	-0,012	13
15. Levante	-0,343	0,304	15	-0,091	14	-0,178	14	-0,003	12
16. Valladolid	-0,418	0,305	16	-0,143	17	-0,295	16	0,009	10
17. Celta	-0,460	0,307	17	-0,109	15	-0,098	13	-0,121	15
18. Girona	-0,662	0,317	18	-0,134	16	-0,356	19	0,088	5
19. Huesca	-0,737	0,308	19	-0,262	20	-0,366	20	-0,158	20
20. Vallecano	-0,968	0,326	20	-0,222	19	-0,308	17	-0,136	16

A [Tabela 2](#) traz as habilidades estimadas para a edição da *La Liga* de 2018 como uma estimativa pontual que representa a habilidade de determinada equipe dentro da competição. No caso do modelo dinâmico, pode-se também verificar essas estimativas médias em termos de mandante e visitante. Observa-se que se o *Girona* dependesse somente das partidas disputadas fora de casa, o clube não teria sido rebaixado, já que a habilidade média estimada foi a 5<sup>a</sup> maior do campeonato. Em termos de pontos absolutos foi a 7<sup>a</sup> melhor campanha, apesar de que o aproveitamento foi de 38,6%. Em termos gerais as equipes não conseguem pontuar muito como visitante. O *Barcelona* foi o maior pontuador, com 39 pontos (de 57 disponíveis).

### D.2.4 Avaliação do modelo

Para avaliação foi utilizado o RPS. O modelo foi ajustado para os dados até o 66<sup>o</sup> dia de campeonato, que equivale ao primeiro turno do campeonato espanhol dessa edição. Os resultados das partidas a partir do 67<sup>o</sup> dia são preditos e comparados com os resultados observados pelo RPS. O modelo então é reajustado incluindo também as partidas do dia 67 e é utilizado pra predizer os resultados do dia 68 e assim por

diante até o último dia. As previsões são comparadas também com as proporções empíricas de vitórias e derrotas do time mandante, computada pelas partidas dos primeiros 66 dias.

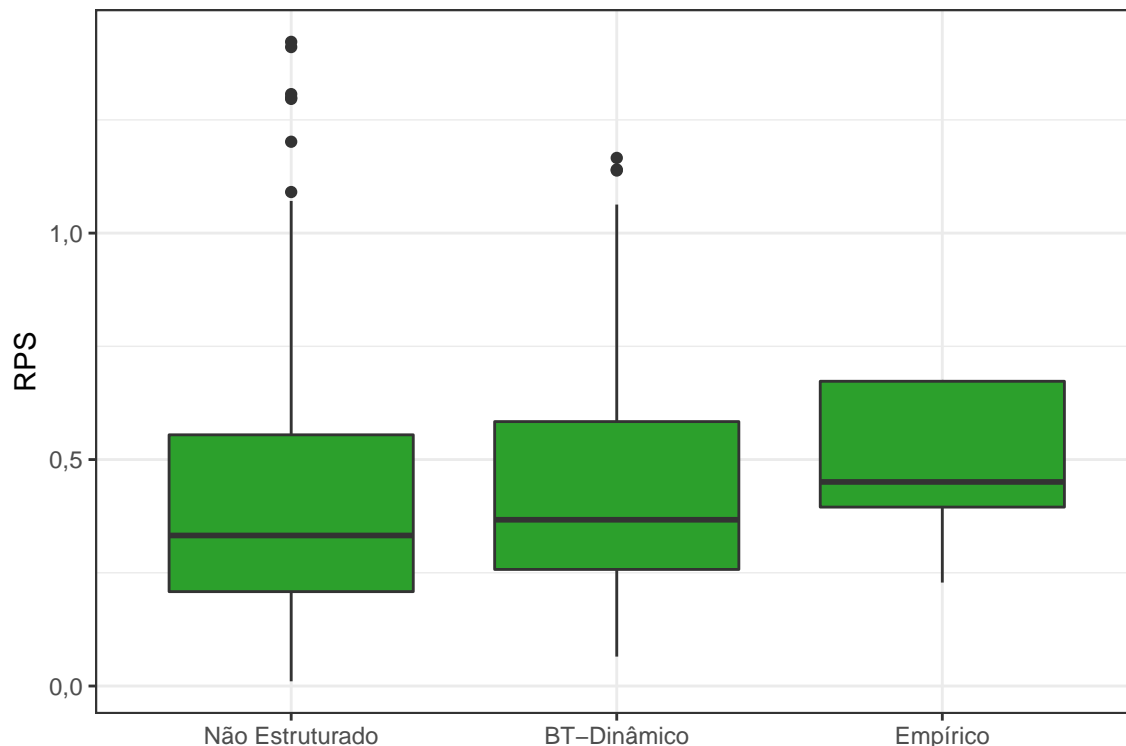


Figura D.8: Boxplot do RPS para a *La Liga* de 2018 computado para o modelo não-estruturado, BT-Dinâmico e previsões baseadas em proporções empíricas

O RPS para o modelo dinâmico foi de 0,432 e o do não-estruturado foi de 0,387, 10,42% menor. O boxplot da [Figura D.8](#) corresponde às estimativas do RPS obtidas para as partidas previstas. Observa-se que novamente o modelo BT-Dinâmico ficou similar ao não-estruturado, enquanto que o empírico é levemente pior.

### D.3 *Ligue 1*

Na temporada de 2018-19 da *Ligue 1* houveram  $k + 1 = 20$  equipes, com cada time disputando 38 jogos, sendo metade disputado como mandante e a outra metade como visitante. O número total de partidas da edição foi de  $N = 380$  e ocorreu entre 10 de agosto de 2018 e 20 de maio de 2019, em 119 dias distintos. O campeão foi o *Paris Saint-Germain* com 91 pontos (29V, 4E e 5D) e os rebaixados foram o *Caen* (7V, 12E e 19D) e *Guingamp* (5V, 12E e 21D). O clube que ficou na 18ª posição, *Dijon*, jogou os *play-offs* para evitar o rebaixamento e conseguiu se manter na primeira divisão do campeonato francês.

O PSG assegurou o bi-campeonato um mês antes (24 de abril) do término do torneio, pois o segundo colocado, *Lille* obteve um resultado desfavorável e portanto impossibilitando-o de alcançar o topo da tabela. O *Paris SG* ficou na liderança do campeonato desde a segunda rodada da edição, enquanto que o *Guingamp* esteve na zona de rebaixamento desde a terceira rodada. Estes são exemplos de dois extremos

que se mantiveram constantes dentro da competição, de forma positiva e negativa, respectivamente.

### D.3.1 Modelo de Bradley-Terry

A Figura D.9 que apresenta as habilidades estimadas pelo modelo de Bradley-Terry estático apresenta nenhuma surpresa.

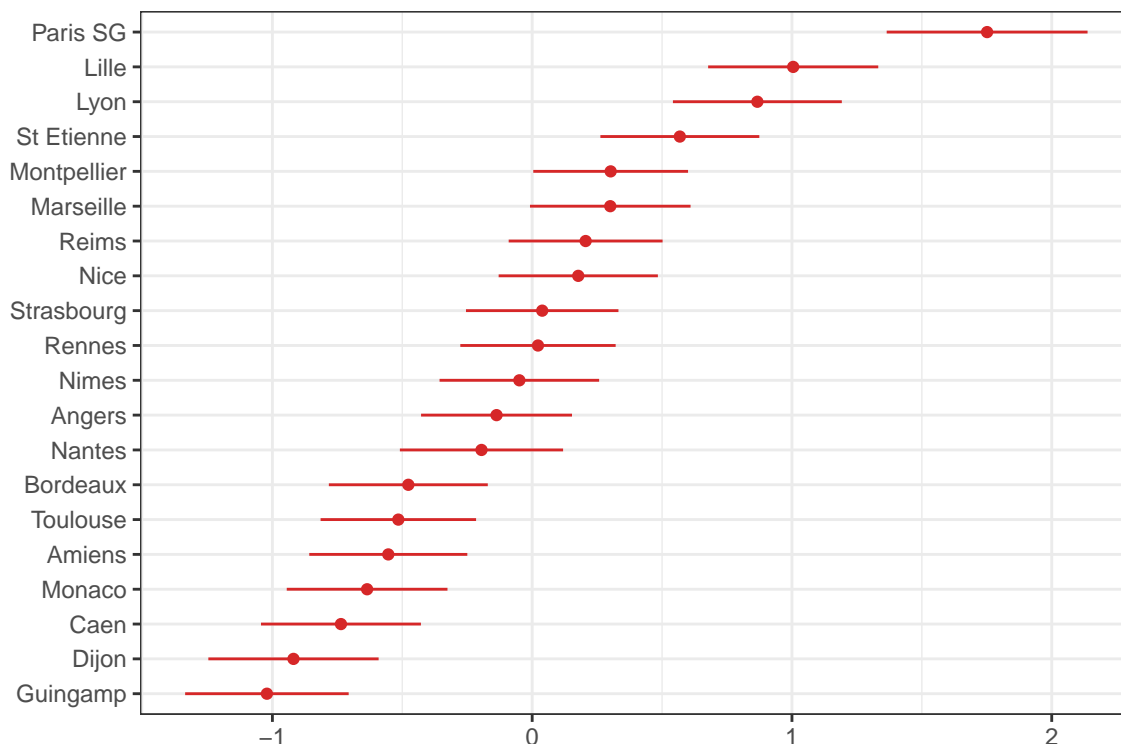


Figura D.9: Habilidades e QSE estimados das 20 equipes da edição de 2018 da *Ligue 1* pelo modelo de Bradley-Terry

Claramente o PSG é superior aos outros clubes da *Ligue 1* e isso fica evidente, já que não há intersecção no intervalo de habilidade dele com o *Lille* ou qualquer outra equipe dessa edição. Ao final do campeonato, a diferença na pontuação dessas duas equipes foi de 16 pontos. Quanto às menores habilidades, nota-se que apesar da péssima campanha do *Guingamp*, a estimativa dele não foi muito diferente do que foi para o *Dijon* ou *Caen*, por exemplo.

### D.3.2 Modelo de Bradley-Terry Dinâmico

O modelo de Bradley-Terry dinâmico que considera habilidades diferentes em cada partida em termos de mandante e visitante foi aplicado aos dados da *Ligue 1*.

Os parâmetros estimados de suavização específicos do mandante e visitante,  $\hat{\lambda}_1 = 0,082$  e  $\hat{\lambda}_2 = 0,038$ , foram valores bem próximos de zero. O vetor de parâmetros de interesse estimado foi de  $\hat{\gamma} = (\hat{\beta}_1; \hat{\beta}_2; \hat{\delta})^T = (1,182; 1,343; 0,453)^T$ , com erros padrão de 0,274, 0,383 e 0,086, respectivamente.

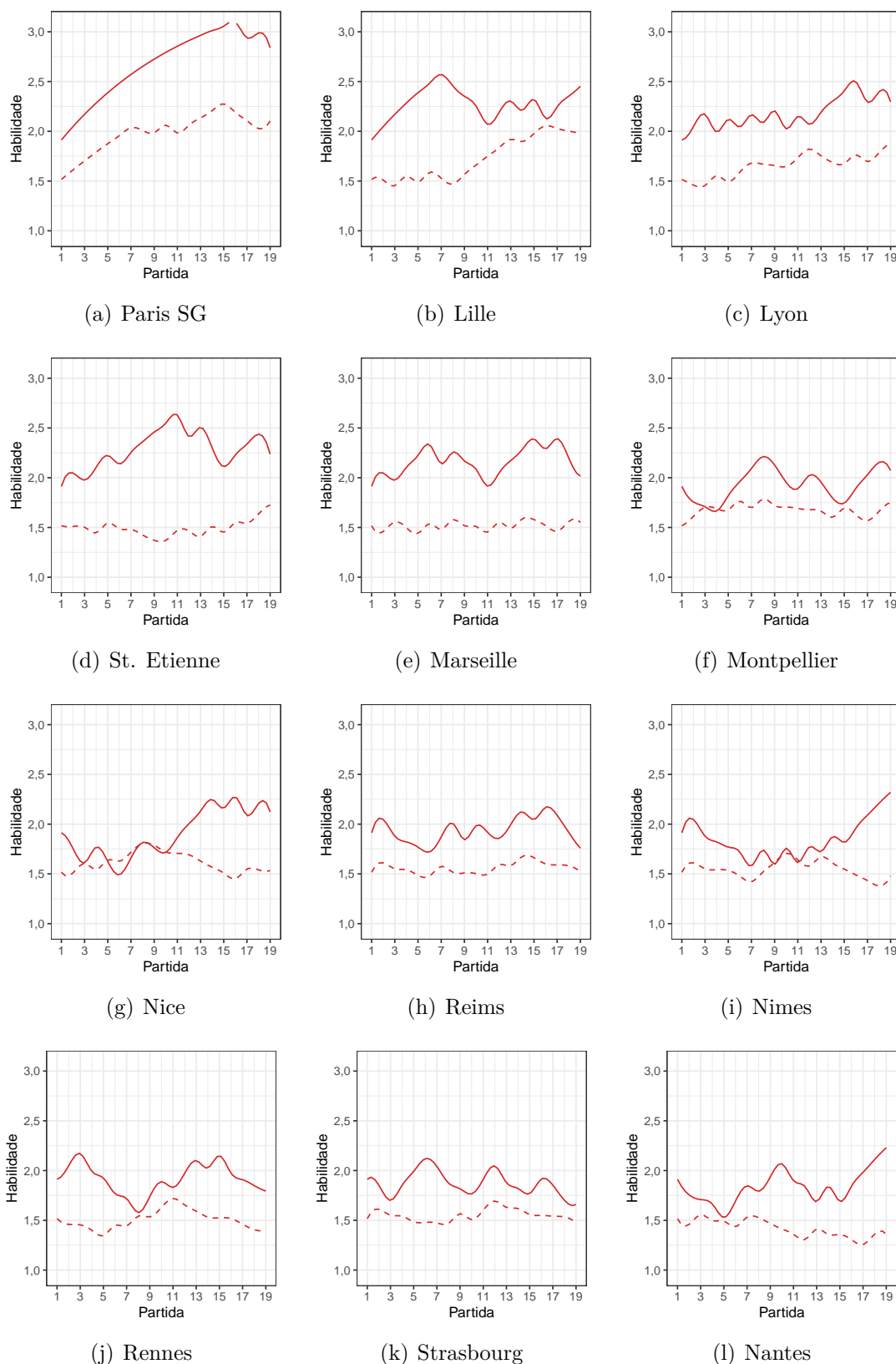


Figura D.10: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da *Ligue 1* - 1<sup>o</sup> ao 12<sup>o</sup> colocado

Pela [Figura D.10](#) é possível notar a ascendência instantânea na habilidade estimada do PSG, que só foi regredir um pouco em jogos dentro de casa após a 15<sup>a</sup>

partida, em que o clube obteve dois empates, na 16<sup>a</sup> a 18<sup>a</sup> rodada. O *St. Etienne* e o *Marseille* tiveram habilidades como visitante estáveis, enquanto que como mandante houve mais variação entre uma partida e outra, além das estimativas serem maiores. A campanha do *St. Etienne* foi de 13V, 2E e 4D (em casa) e 6V, 7E e 6D (fora de casa), um aproveitamento de 71,93% e 43,86%, respectivamente. Para o *Marseille* foi de 64,91% e 42,1%, o que consiste em 11V, 4E e 4D como mandante e 7V, 3E e 9D como visitante.

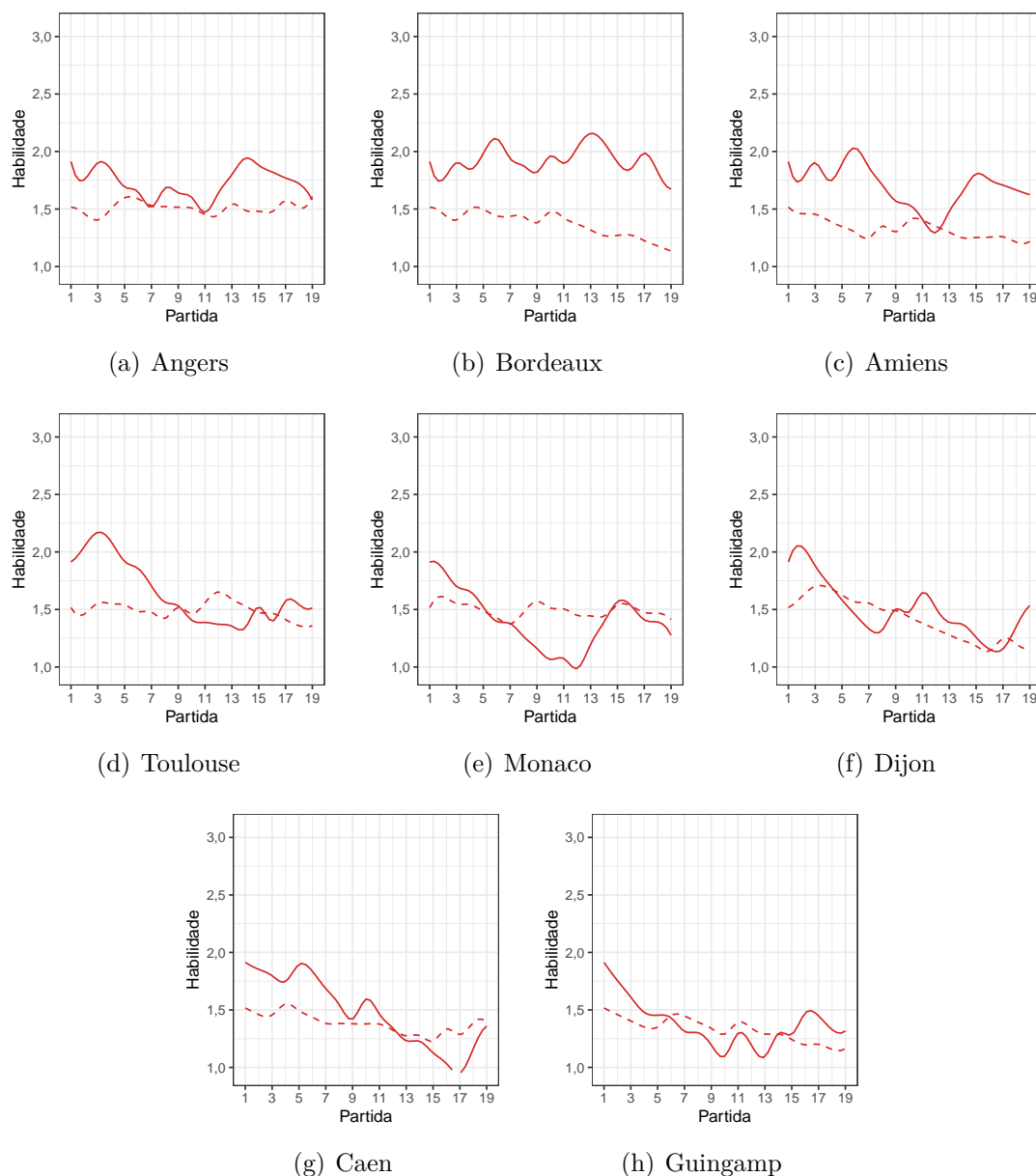


Figura D.11: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da *Ligue 1* - 13<sup>o</sup> ao 20<sup>o</sup> colocado

As outras equipes representadas na [Figura D.11](#) possuem comportamentos parecidos, que em alguns momentos chegam a ter oscilações tão grandes na habilidade dentro de casa, que acabam por terem estimativas como visitante mais altas do que como mandante, o que em termos gerais não é muito comum, por várias razões já citadas anteriormente no [Capítulo 3](#).

O *Monaco*, por exemplo, foi vencer sua primeira partida dentro do *Stade Louis II* somente na 12<sup>a</sup> rodada. Antes disso foram 4 empates e 7 derrotas. Para os clubes que ficaram nas últimas posições da classificação final do campeonato é difícil encontrar comportamentos diferentes do observado porque atualmente a discrepância entre as equipes é tão alta que os piores clubes dentro de um torneio são muito inferiores aos demais e por consequência a quantidade de vitórias é muito baixa, portanto a habilidade estimada ao decorrer do tempo será cada vez menor.

### D.3.3 Modelo Estético x Modelo Dinâmico

A [Tabela 3](#) fornece as estimativas do modelo estático, além do modelo dinâmico de forma geral, visitante e mandante. O *ranking*, que fornece informação da ordem crescente com relação aos clubes e suas habilidades também encontram-se na tabela.

Tabela 3: Tabela resumo das habilidades e QSE estimados pelo modelo BT-Estático e as habilidades suavizadas médias pelo BT-Dinâmico em termos gerais, como mandante e visitante da *Ligue 1* de 2018

Equipe	Modelo Estático			Modelo Dinâmico					
	Habilidade	QSE	Rank	Habilidade	Rank	Habilidade (Mandante)	Rank	Habilidade (Visitante)	Rank
1. Paris SG	1,751	0,387	1	0,621	1	0,789	1	0,453	1
2. Lille	1,005	0,327	2	0,296	2	0,395	3	0,197	2
3. Lyon	0,867	0,325	3	0,214	3	0,299	4	0,130	4
4. St Etienne	0,569	0,306	4	0,182	4	0,397	2	-0,033	12
5. Marseille	0,301	0,309	6	0,134	5	0,277	5	-0,010	9
6. Montpellier	0,302	0,298	5	0,101	6	0,056	6	0,146	3
7. Nice	0,177	0,306	8	0,056	7	0,028	10	0,085	5
8. Reims	0,206	0,296	7	0,042	8	0,055	7	0,030	6
9. Nimes	-0,049	0,307	11	-0,005	11	-0,026	12	0,015	8
10. Rennes	0,022	0,299	10	0,005	9	0,034	9	-0,024	11
11. Strasbourg	0,039	0,293	9	0,004	10	-0,011	11	0,020	7
12. Nantes	-0,195	0,314	13	-0,072	13	-0,035	13	-0,108	15
13. Angers	-0,137	0,291	12	-0,084	14	-0,151	14	-0,016	10
14. Bordeaux	-0,477	0,306	14	-0,058	12	0,043	8	-0,159	18
15. Amiens	-0,554	0,304	16	-0,188	16	-0,178	15	-0,198	19
16. Toulouse	-0,515	0,299	15	-0,135	15	-0,236	16	-0,034	13
17. Monaco	-0,635	0,309	17	-0,254	18	-0,471	19	-0,037	14
18. Dijon	-0,919	0,328	19	-0,245	17	-0,373	17	-0,116	16
19. Caen	-0,736	0,308	18	-0,269	19	-0,395	18	-0,143	17
20. Guingamp	-1,021	0,315	20	-0,346	20	-0,495	20	-0,198	19

De todos os modelos ajustados, é a primeira vez que uma equipe fica em 1<sup>o</sup> lugar em todos os quatro *rankings*, conforme fornecido pela [Tabela 3](#). Fora isso, o que mais chama a atenção é o fato de que o *St. Etienne* ficou em 12<sup>o</sup> no *ranking* do modelo dinâmico referente a habilidade como visitante, o que é discrepante com as posições ocupadas no restante da tabela e também na classificação geral do campeonato.

### D.3.4 Avaliação do modelo

Para avaliação foi utilizado o RPS. O modelo foi ajustado para os dados até o 51<sup>o</sup> dia de campeonato, que equivale ao primeiro turno do campeonato francês dessa edição. Os resultados das partidas a partir do 52<sup>o</sup> dia são preditos e comparados com os resultados observados pelo RPS. O modelo então é reajustado incluindo também as partidas do dia 52 e é utilizado pra predizer os resultados do dia 53 e assim por diante até o último dia. As predições são comparadas também com as proporções

empíricas de vitórias e derrotas do time mandante, computada pelas partidas dos primeiros 51 dias.

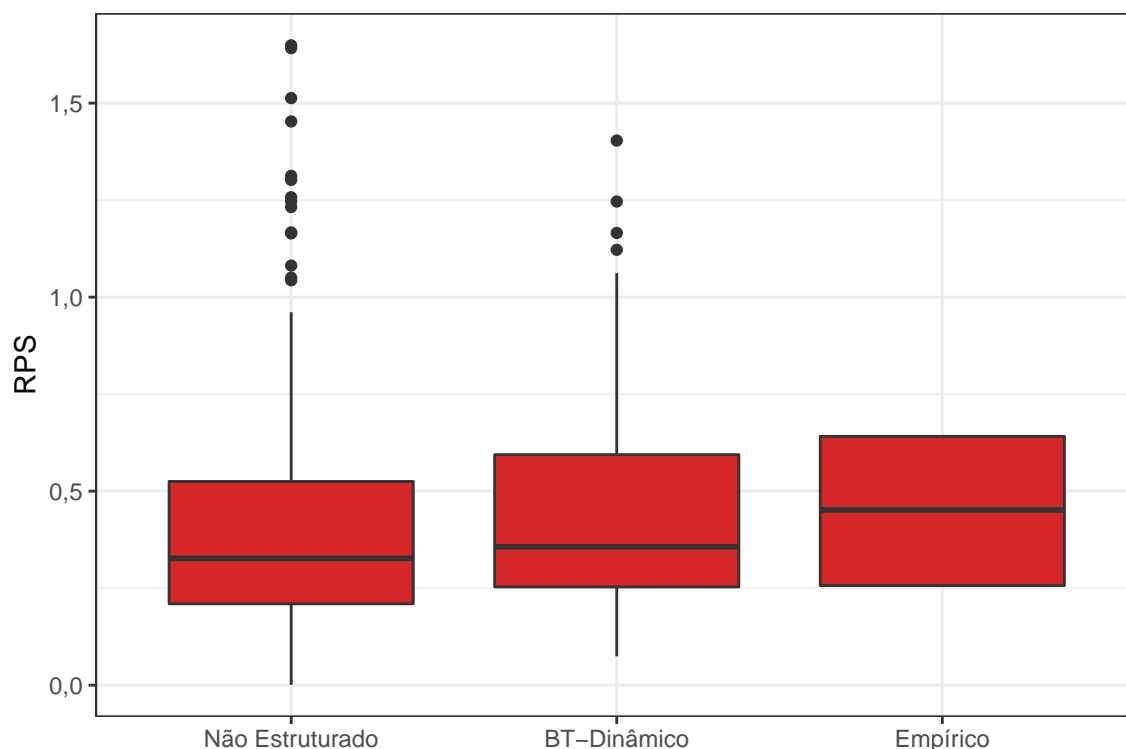


Figura D.12: Boxplot do RPS para a *Ligue 1* de 2018 computado para o modelo não-estruturado, BT-Dinâmico e previsões baseadas em proporções empíricas

O RPS para o modelo dinâmico foi de 0,431 e o do não-estruturado foi de 0,378, 12,3% menor. O boxplot da [Figura D.12](#) corresponde às estimativas do RPS obtidas para as partidas previstas. Observa-se que novamente o modelo BT-Dinâmico ficou similar ao não-estruturado, porém dessa vez o modelo empírico não ficou muito diferente dos outros dois, apesar de ainda ser o pior.

#### D.4 *Serie A*

Na temporada de 2018-19 da *Serie A* houveram  $k + 1 = 20$  equipes, com cada time disputando 38 jogos, sendo metade disputado como mandante e a outra metade como visitante. O número total de partidas da edição foi de  $N = 380$  e ocorreu entre 18 de agosto de 2018 e 26 de maio de 2019, em 114 dias distintos. A campeã pelo 8º ano seguido foi a *Juventus* com 90 pontos (28V, 6E e 4D) e os rebaixados foram o *Empoli* (10V, 8E e 20D), *Frosinone* (5V, 10E e 23D) e o *Chievo* (2V, 14E e 22D).

Ao fim do torneio o *Chievo* obteve 20 pontos, porém foi punido com a perda de três pontos por fraudar contabilidade, mas independentemente ele seria o lanterna do campeonato. Em muitos aspectos a *Serie A* é parecida com a *Ligue 1*, por exemplo, tanto a *Juventus* quanto o *Chievo* ficaram 37 rodadas (desde a segunda) na posição em que terminaram nessa edição.

A *Serie A*, apesar de possuir muitas equipes expressivas, não parece ser um torneio competitivo. Nesta edição a *Juventus* levantou a taça de campeã pela oitava

vez seguida e 35<sup>a</sup> no geral. O *Milan* e a *Inter* venceram o campeonato italiano 18 vezes cada somente. Clubes como a *Roma* e o *Napoli* são expressivos nacionalmente, mas não são vencedores.

#### D.4.1 Modelo de Bradley-Terry

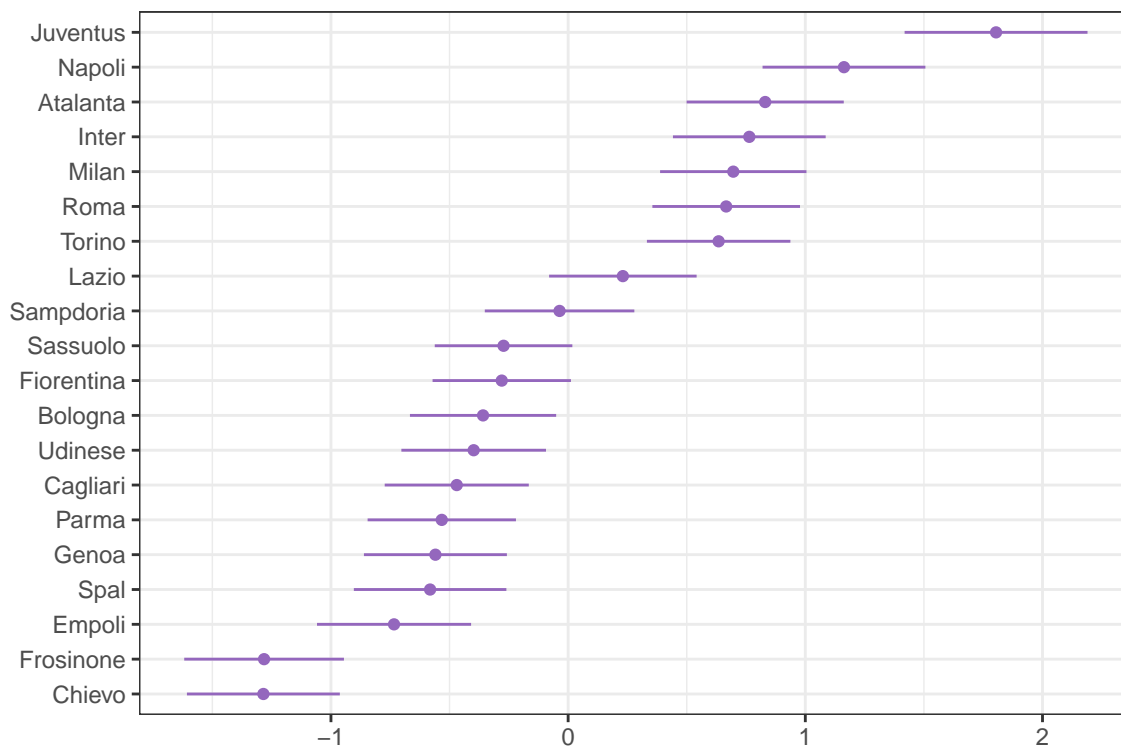


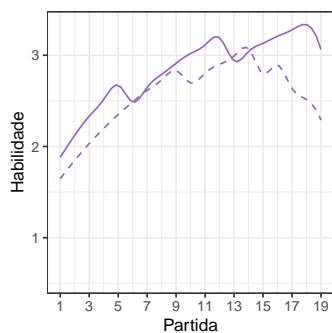
Figura D.13: Habilidades e QSE estimados das 20 equipes da edição de 2018 da *Serie A* pelo modelo de Bradley-Terry

A Figura D.13 apresenta as habilidades estimadas para cada equipe. Primeiramente observa-se que parece possuir quatro conjuntos de clubes com habilidades parecidas. A *Juventus* claramente apresenta uma estimativa bem distante dos demais, com exceção do *Napoli* que apesar do valor pontual da habilidade estar bastante distante, os QSE de ambas as equipes possuem intersecções. Na outra extremidade encontram-se o *Chievo* e *Frosinone*, ambos os clubes que foram rebaixados pois ficaram nas últimas duas posições e obtiveram estimativas bastante próximas. Outro conglomerado que se pode observar é que entre a *Atalanta* e o *Torino* não há muita diferença nas habilidades. Ao observar a classificação final do campeonato, nota-se que a diferença entre esses clubes foi de apenas seis pontos. Entre o *Sassuolo* e o *Spal* também não há grande discrepâncias nas estimativas apresentadas no gráfico e a diferença na pontuação entre esses times também foi de seis pontos.

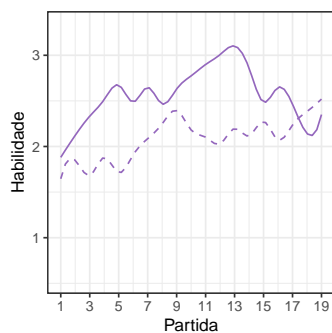
#### D.4.2 Modelo de Bradley-Terry Dinâmico

Os parâmetros estimados de suavização específicos do mandante e visitante,  $\hat{\lambda}_1 = 0,131$  e  $\hat{\lambda}_2 = 0,090$ , foram valores bem próximos de zero. O vetor de parâmetros de interesse estimado foi de  $\hat{\gamma} = (\hat{\beta}_1; \hat{\beta}_2; \hat{\delta})^T = (1,243; 1,239; 0,087)^T$ , com erros padrão de 0,189, 0,239 e 0,087, respectivamente.

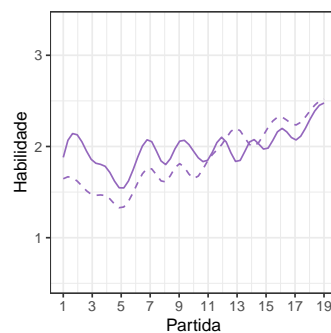




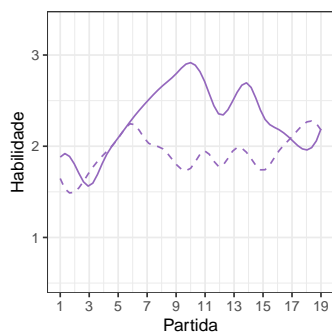
(a) Juventus



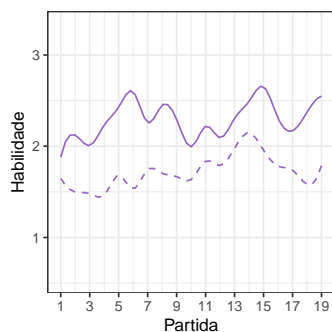
(b) Napoli



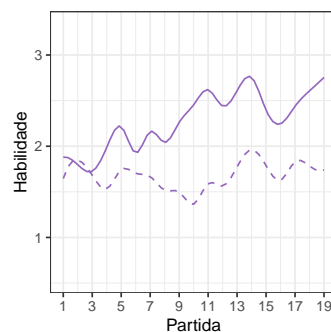
(c) Atalanta



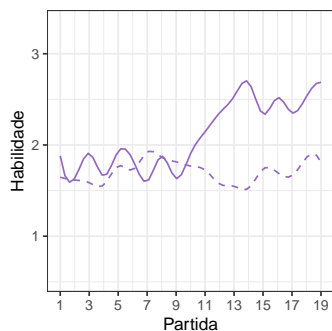
(d) Inter



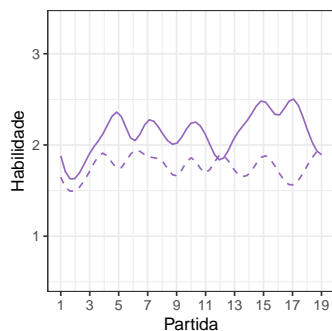
(e) Milan



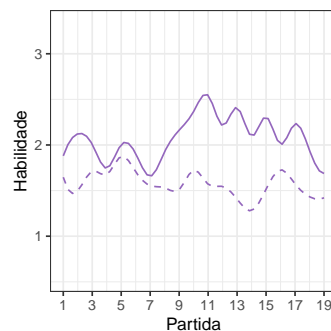
(f) Roma



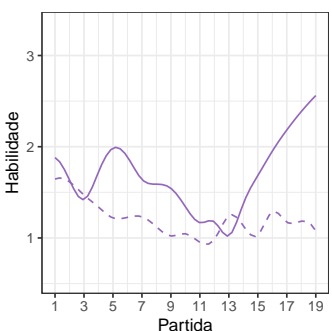
(g) Torino



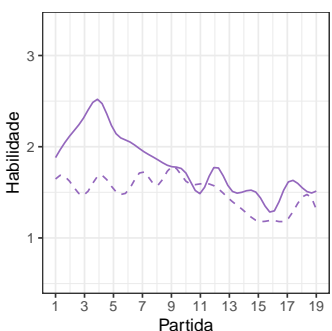
(h) Lazio



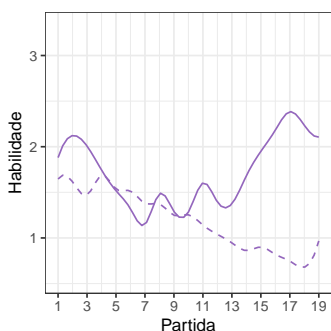
(i) Sampdoria



(j) Bologna



(k) Sassuolo



(l) Udinese

Figura D.14: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da *Serie A* - 1<sup>o</sup> ao 12<sup>o</sup> colocado

Sem surpresa nenhuma os resultados dispostos na [Figura D.14](#). A *Juventus* claramente com as habilidades estimadas crescendo ao longo das partidas, mas dando uma decrescida no final quando o título já tinha sido garantido. De fato, as últimas cinco partidas acarretaram em três empates e duas derrotas. A *Atalanta*, a maior supresa do campeonato, terminou o torneio em terceiro lugar, garantindo assim uma vaga para a UCL. Observa-se que as habilidades estimadas, tanto considerando mandante quanto visitante, parecem seguir uma tendência positiva, sendo que em alguns momentos do tempo a vantagem de fora de casa foi maior do que dentro de casa.

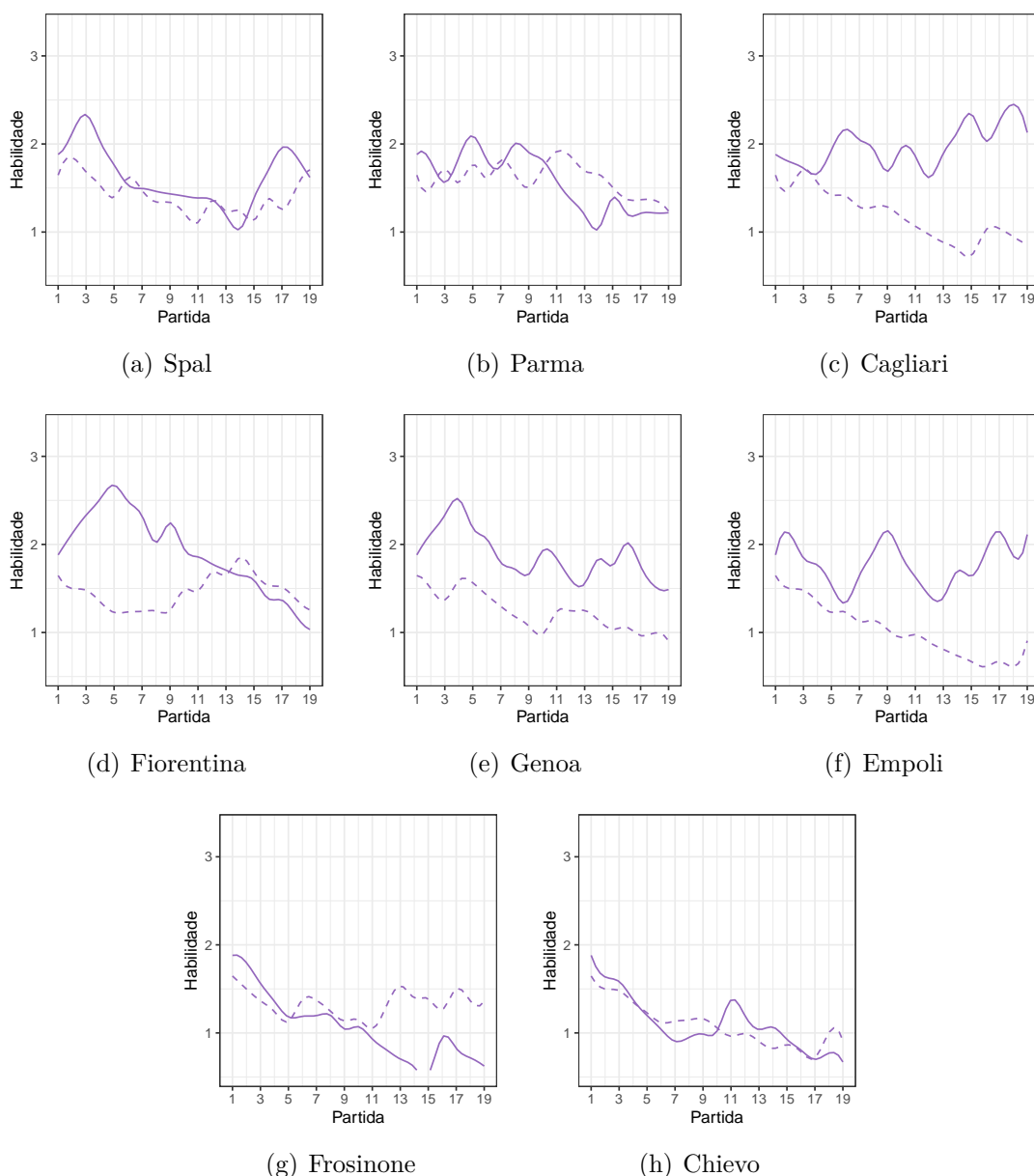


Figura D.15: Habilidades suavizadas das equipes em jogos dentro de casa (—) e jogos fora de casa (- - -) na edição de 2018 da *Serie A* - 13<sup>o</sup> ao 20<sup>o</sup> colocado

A campanha da *Udinese* como visitante consistiu em 3 vitórias, 5 empates e 11 derrotas, um aproveitamento de 24,56%. A habilidade estimada para essa equipe ajuda a entender esses resultados desfavoráveis fora de casa ao longo do campeonato,

já que uma derrota em determinada partida irá fazer com que a estimativa diminua em relação á estimativa anterior.

Na parte final da classificação, as oito últimas equipes estão sendo apresentadas na [Figura D.15](#). Times como o *Geonoo*, *Empoli* e *Cagliari* também tiveram dificuldades em pontuar fora de casa, o que reflete diretamente na curva da habilidade como visitante desses clubes.

A queda na habilidade da *Fiorentina* como mandante a partir da nona rodada é explicada pela quantidade de empates e derrotas obtidas pela equipe. Em 19 rodadas, foram apenas cinco vitórias, sendo que quatro delas foram no começo da temporada. Foram 11 rodadas sem vencer dentro de casa e se estendeu até o termino da competição. Essa jejum contribuiu para que o clube tivesse as estimações de habilidade reduzidas.

Em termos de zona de rebaixamento, fica evidente que o desempenho do *Frosinone* dentro do estádio adversário é superior ao desempenho no próprio estádio. Em casa o clube conseguiu apenas uma vitória, em um jogo que foi  $3 \times 2$  contra o Parma no dia 3 de Abril.

#### D.4.3 Modelo Estético x Modelo Dinâmico

Dentre as edições de 2018 apresentadas neste trabalho sob enfoque do modelo de Bradley-Terry o campeonato italiano é o menos surpreendente em termos de resultados.

Tabela 4: Tabela resumo das habilidades e QSE estimados pelo modelo BT-Estático e as habilidades suavizadas médias pelo BT-Dinâmico em termos gerais, como mandante e visitante da *Serie A* de 2018

Equipe	Modelo Estático			Modelo Dinâmico					
	Habilidade	QSE	Rank	Habilidade	Rank	Habilidade (Mandante)	Rank	Habilidade (Visitante)	Rank
1. Juventus	1,804	0,386	1	0,941	1	0,894	1	0,988	1
2. Napoli	1,163	0,343	2	0,572	2	0,621	2	0,523	2
3. Atalanta	0,831	0,331	3	0,186	7	0,053	9	0,319	4
4. Inter	0,764	0,322	4	0,357	3	0,349	4	0,364	3
5. Milan	0,696	0,309	5	0,258	4	0,355	3	0,162	6
6. Roma	0,666	0,311	6	0,234	5	0,348	5	0,120	8
7. Torino	0,635	0,302	7	0,163	8	0,175	7	0,150	7
8. Lazio	0,231	0,311	8	0,200	6	0,192	6	0,208	5
9. Sampdoria	-0,036	0,315	9	0,075	9	0,133	8	0,018	10
10. Bologna	-0,359	0,308	12	-0,292	17	-0,238	16	-0,346	16
11. Sassuolo	-0,272	0,290	10	-0,098	11	-0,133	13	-0,062	11
12. Udinese	-0,398	0,305	13	-0,282	16	-0,210	15	-0,355	17
13. Spal	-0,582	0,322	17	-0,219	15	-0,302	17	-0,135	13
14. Parma	-0,533	0,313	15	-0,155	12	-0,358	18	0,047	9
15. Cagliari	-0,470	0,304	14	-0,162	13	0,044	10	-0,367	18
16. Fiorentina	-0,280	0,292	11	-0,065	10	-0,030	11	-0,099	12
17. Genoa	-0,560	0,301	16	-0,190	14	-0,059	12	-0,322	15
18. Empoli	-0,734	0,325	18	-0,337	18	-0,147	14	-0,527	20
19. Frosinone	-1,282	0,337	19	-0,546	19	-0,868	20	-0,223	14
20. Chievo	-1,285	0,322	20	-0,641	20	-0,819	19	-0,462	19

Os *rankings* das habilidades apresentados na [Tabela 4](#) apresentam poucas modificações ao serem comparados com a classificação geral ao final da *Serie A* de 2018. O resultado mais extremo encontrado é o caso do *Parma* que terminou a competição em 14<sup>o</sup> lugar, mas pelo modelo estático ficou na 15<sup>a</sup> posição, enquanto que no dinâmico ficou em 12<sup>o</sup>. Quando feito a estimação da habilidade geral média em termos

de mandante e visitante, este acabou por ficar em 18<sup>o</sup> e 9<sup>o</sup>, respectivamente. O aproveitamento dentro de casa do clube, 38,6%, condiz com a posição estabelecida no *ranking* de habilidade do mandante.

#### D.4.4 Avaliação do modelo

Para avaliação foi utilizado o RPS. O modelo foi ajustado para os dados até o 53<sup>o</sup> dia de campeonato, que equivale ao primeiro turno do campeonato italiano dessa edição. Os resultados das partidas a partir do 54<sup>o</sup> dia são preditos e comparados com os resultados observados pelo RPS. O modelo então é reajustado incluindo também as partidas do dia 54 e é utilizado pra prever os resultados do dia 55 e assim por diante até o último dia. As previsões são comparadas também com as proporções empíricas de vitórias e derrotas do time mandante, computada pelas partidas dos primeiros 53 dias.

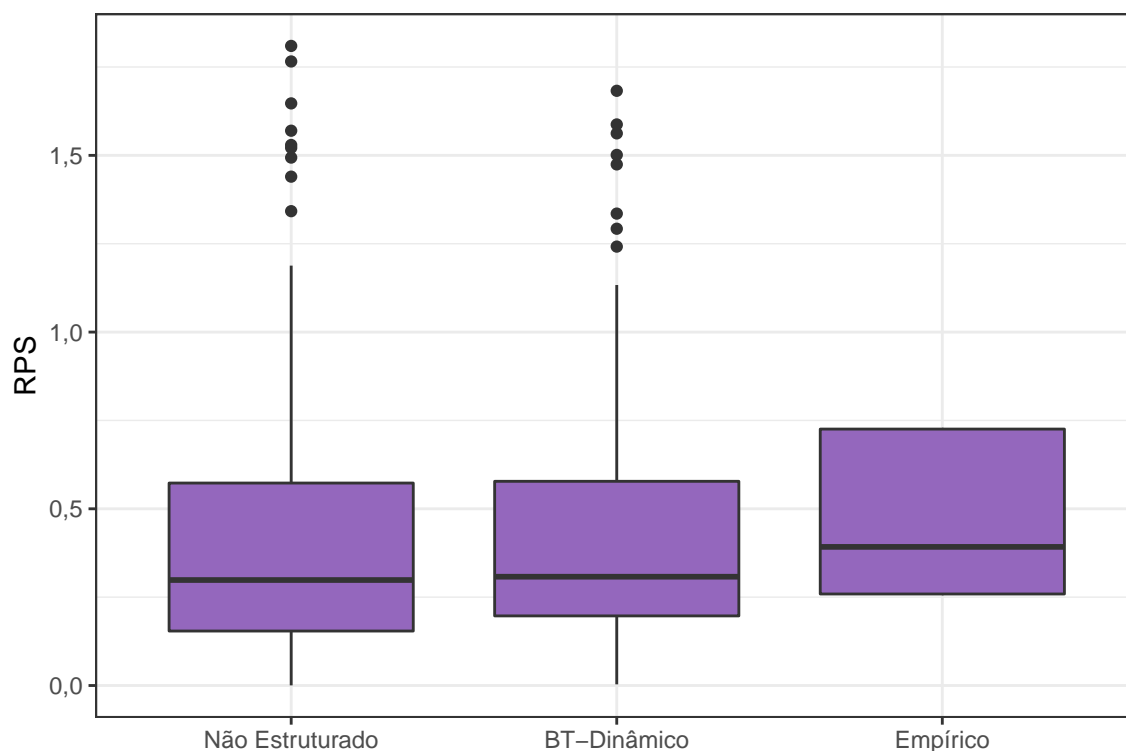


Figura D.16: Boxplot do RPS para a *Serie A* de 2018 computado para o modelo não-estruturado, BT-Dinâmico e previsões baseadas em proporções empíricas

O RPS para o modelo dinâmico foi de 0,423 e o do não-estruturado foi de 0,349, 17,49% menor. O boxplot da [Figura D.16](#) corresponde às estimativas do RPS obtidas para as partidas preditas. Observa-se que novamente o modelo BT-Dinâmico ficou similar ao não-estruturado, enquanto que o empírico é um pouco pior.