



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de São José do Rio Preto

Arthur Necchi Corrêa

**Probabilidade no Futebol: Um Incentivo ao Estudo
da Teoria Estatística**

São José do Rio Preto
2019

Arthur Necchi Corrêa

**Probabilidade no Futebol: Um Incentivo ao Estudo
da Teoria Estatística**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Matemática em Rede Nacional, junto ao Programa de Pós-Graduação em Mestrado Profissional em Matemática em Rede Nacional, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Orientador: Prof. Dr. João Carlos Ferreira Costa

São José do Rio Preto
2019

C824p

Corrêa, Arthur Necchi

Probabilidade no futebol : um incentivo ao estudo da Teoria Estatística / Arthur Necchi Corrêa. -- São José do Rio Preto, 2019
68 p. : tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp),
Instituto de Biociências Letras e Ciências Exatas, São José do Rio
Preto

Orientador: João Carlos Ferreira Costa

1. Modelagem. 2. Simulação. 3. Previsão Esportiva. 4. Futebol. 5.
Probabilidade. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Arthur Necchi Corrêa

**Probabilidade no Futebol: Um Incentivo ao Estudo
da Teoria Estatística**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Matemática em Rede Nacional, junto ao Programa de Pós-Graduação em Mestrado Profissional em Matemática em Rede Nacional, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Comissão Examinadora

Prof. Dr. João Carlos Ferreira Costa
UNESP – Câmpus de São José do Rio Preto
Orientador

Prof. Dr. Jéfferson Luiz Rocha Bastos
UNESP – Câmpus de São José do Rio Preto

Prof. Dr. Aldicio José Miranda
UFU – Universidade Federal de Uberlândia

São José do Rio Preto
06 de maio de 2019

Dedicatória

À minha esposa, meu amor e companheira, Ana Flávia, por ter me inspirado e apoiado desde o início, além de sempre acreditar nos meus sonhos, me incentivando e ajudando em todos os momentos.

Agradecimentos

Ao meu orientador, Prof. Dr. João Carlos, o qual tenho o prazer de conhecer a mais de uma década. Pessoa que merece toda admiração e respeito, detentor de uma excelente didática, que sempre me inspirou no meio acadêmico. Agradeço pela orientação, paciência e sugestões para a realização deste trabalho.

Aos meus pais, Eliete e Orides, e meu irmão, Edgar, por sempre incentivarem os estudos, compartilhando cada conquista.

À minha família e amigos, pelo apoio e pela compreensão do tempo de convívio muitas vezes sacrificado para realização deste trabalho.

Aos meus colegas do Mestrado, especialmente à Larissa Oliveira, companheira desde os tempos de graduação, cujo apoio e amizade estiveram presentes em todos os momentos.

Aos professores do programa, pelos conhecimentos compartilhados.

Aos professores da banca de defesa, Prof. Dr. Jéfferson Luiz Rocha Bastos e Prof. Dr. Aldicio José Miranda, pelas preciosas sugestões para melhoria deste trabalho.

À todos que participaram desta fase de minha vida, mesmo que indiretamente, meu muito obrigado!

Resumo

O desinteresse pela Matemática pode estar atrelado à falta de articulação entre teoria e prática. Apresentá-la de forma contextualizada é uma maneira positiva para o ensino dessa ciência. Diante disso, esse trabalho traz um incentivo ao estudo da modelagem matemática em um campeonato de futebol, usando técnicas da teoria Estatística. A partir de um modelo de previsão de resultados no futebol, presente na literatura, foi proposto um novo modelo baseado no primeiro, que mostrou-se uma alternativa viável para previsão esportiva. Além disso, exibimos uma aplicação prática dos modelos, no campeonato Brasileiro de 2018 - Série A, e sugerimos algumas atividades para serem desenvolvidas em sala de aula para o Ensino Médio.

Palavras-chave: Modelagem; Simulação; Previsão Esportiva; Futebol; Probabilidade; Medida de De Finetti.

Abstract

The lack of interest in Mathematics may be related to the lack of articulation between theory and practice. Presenting it in a contextualized way is a positive way to teach this science. Therefore, this work brings an incentive to the study of mathematical modeling in a soccer league, using Statistical theory techniques. Based on a soccer results prediction model presented in the literature, a new model based on the first one was proposed, which proved to be a viable alternative for sports forecasting. In addition, we showed a practical application of the models, in the 2018 Campeonato Brasileiro Série A, and we suggest some activities to be developed in the classroom for high school.

Keywords: Modelling; Simulation; Sports Forecasting; Soccer; Probability; De Finetti Measure.

Lista de Tabelas

3.1	Estrutura dos Confrontos.	29
3.2	Jogos do Campeonato Fictício.	33
3.3	Classificação final do Campeonato Fictício.	33
3.4	Probabilidade dos jogos para o modelo PFS.	36
3.5	Probabilidade dos jogos para o modelo PFC.	37
4.1	Classificação final da Série A do Campeonato Brasileiro de 2018.	39
4.2	Esquema das rodadas utilizadas para ajuste dos modelos e previsões.	41
4.3	Classificação das distâncias de De Finetti para 19 rodadas de ajuste.	43
4.4	Medida de De Finetti para cada rodada com base em um ajuste de 19 rodadas.	43
4.5	Classificação das distâncias de De Finetti para 28 rodadas de ajuste.	44
4.6	Medida de De Finetti para cada rodada com base em um ajuste de 19 rodadas.	44
4.7	Distâncias de De Finetti para a 33 ^a rodada.	45
4.8	Classificação da Série A do Campeonato Brasileiro de 2018 com 19 rodadas.	47
4.9	Média de pontos considerando 19 rodadas para os ajustes, após 10.000 simulações.	48
4.10	Resumo da classificação de cada time (em %) com as 19 primeiras rodadas para os ajustes, após 10.000 simulações.	49
4.11	Classificação da Série A do Campeonato Brasileiro de 2018 com 28 rodadas.	50
4.12	Média de pontos considerando 28 rodadas para os ajustes, após 10.000 simulações.	51
4.13	Resumos da classificação de cada time (em %) com as 28 primeiras rodadas para os ajustes, após 10.000 simulações.	52
5.1	Frequência relativa (em %) observada por quantidade de repetição no lançamento de um dado, via simulação.	56
5.2	Frequência relativa (em %) observada por quantidade de repetição em uma partida de futebol e pontos médios obtidos, via simulação.	57

Sumário

1	Introdução	11
1.1	Objetivos do Trabalho	13
1.1.1	Objetivo geral	13
1.1.2	Objetivos específicos	13
1.2	Organização do Trabalho	13
2	Introdução à Teoria das Probabilidades	15
2.1	Modelagem	15
2.2	Introdução à Teoria dos Conjuntos	16
2.2.1	Operações com conjuntos	17
2.3	Espaço Amostral e Eventos	19
2.4	Noções Fundamentais de Probabilidade	21
2.5	Espaços Amostrais Finitos	23
2.6	Lei dos Grandes Números	24
2.6.1	Lei Fraca e Lei Forte - formalização	24
3	Modelos de Previsão de Futebol	27
3.1	Especificações dos Modelos	27
3.2	Modelo para Previsão	28
3.3	Construção de um Novo Modelo	30
3.4	Campeonato Fictício	33
4	Aplicação para Dados Reais e um Estudo de Simulação	38
4.1	Dados da Série A do Campeonato Brasileiro de 2018	38
4.1.1	Formato da competição e posições relevantes	40
4.2	Método Utilizado	40
4.3	Medidas de Qualidade Preditiva	41
4.3.1	Medida de De Finetti	41
4.3.2	Erro Quadrático Médio	42
4.4	Análise da Qualidade Preditiva	42
4.5	Simulação	45

4.5.1	Resultados	46
5	Propostas de Atividades em Sala de Aula	53
5.1	Atividade 1 - Modelagem no Futebol	53
5.2	Atividade 2 - Qualidade Preditiva	54
5.3	Atividade 3 - Aplicação da Lei dos Grandes Números	55
5.3.1	Lançamento de uma moeda	55
5.3.2	Simulação do lançamento de um dado	55
5.3.3	Simulação de uma partida de futebol	56
5.4	Atividade 4 - Simulação de um campeonato	58
6	Considerações Finais	59
	Referências Bibliográficas	61
	Apêndice A - Simulação para um Campeonato de Pontos Corridos	64

Capítulo 1

Introdução

O número de estudantes que não gostam ou temem a Matemática e saem da escola matematicamente despreparados tem suscitado grande preocupação, dada a crescente importância do raciocínio matemático e da “alfabetização quantitativa” para a vida e o trabalho das pessoas (Steen, 1997; Boaler, 2015). De acordo com o Indicador de Alfabetismo Funcional - 2018, somente no Brasil, cerca de 29% das pessoas pertencentes à faixa etária de 15 a 64 anos são consideradas analfabetas funcionais, por não conseguirem identificar operações matemáticas elementares para a resolução de problemas do cotidiano e por possuir dificuldades em localizar informações explícitas em um texto curto.

Apesar do desenvolvimento da educação matemática nas últimas décadas, o ensino tradicional dessa disciplina ainda permanece no campo da abstração sem articulação entre teoria e prática (Rosen, 2001). Em consequência desse fato, os estudantes desenvolvem ideias negativas sobre a Matemática e veem o assunto como algo distante do seu dia a dia.

Apresentar a Matemática de forma contextualizada é uma característica positiva para o ensino dessa ciência. Segundo Gutstein (2003), convidar os alunos a considerar matematicamente ideias complexas com implicações no mundo real muda fundamentalmente as orientações dos estudantes em relação à Matemática. Diante disso, a proposta deste trabalho é um convite interativo para o aluno conhecer um pouco da modelagem matemática em um campeonato de futebol. A escolha por esse esporte fundamenta-se no fato de ele ser um dos mais populares no mundo e a modalidade esportiva mais praticada no Brasil.

Nos últimos anos, um crescente interesse por modelagem de previsões relacionadas ao futebol vem sendo observado. O entretenimento gerado por esse esporte e as apostas acerca dos resultados das partidas são os motivos que levaram a tal aumento. As apostas tornaram-se tradição na Europa, principalmente na Inglaterra, onde os valores envolvidos têm aumentado consideravelmente. É nesse contexto que surgem os modelos estatísticos que procuram fornecer boas previsões para resultados de jogos.

As primeiras tentativas de modelagem de dados no futebol baseiam-se principalmente no número de gols marcados em uma partida. Moroney (1956) analisou brevemente uma forma de modelar o número de gols e sugeriu que melhorias poderiam ser obtidas. Inicial-

mente, [Pollard \(1986\)](#) e, mais tarde, [Clarke e Norman \(1995\)](#) verificaram que times que jogam em casa possuem uma taxa de vitória maior que a dos visitantes.

O modelo apresentado por [Lee \(1997\)](#) parte do pressuposto de que as quantidades de gols feitos por cada time em uma partida são independentes entre si, mesmo identificando a existência de uma correlação (relativamente baixa) entre o número de gols marcados pelos dois oponentes. [Maher \(1982\)](#) discutiu essa questão mencionando uma solução alternativa, que não foi utilizada principalmente porque exigia um grande custo computacional.

A partir da década de 90, com o desenvolvimento de recursos computacionais, surgiram os primeiros modelos que consideravam uma correlação entre os gols ([Kocherlakota e Kocherlakota, 1992](#)). [Karlis e Ntzoufras \(2003\)](#) utilizaram essa técnica, argumentando que mesmo um pequeno valor para o parâmetro de dependência levava a uma previsão mais precisa do número de empates.

[Dixon e Coles \(1997\)](#) sugeriram que a suposição de independência é razoável, exceto para alguns resultados. Para alteração desse pressuposto, estenderam modelos anteriores introduzindo uma dependência para os placares baixos (de 0-0 até 1-1), os quais são mais frequentes, pelo fato de o jogo começar em 0-0. Além disso, propuseram uma alteração através de uma função de peso exponencial dependente do tempo, diminuindo a relevância de partidas mais antigas, de acordo com o tempo em que as partidas foram realizadas, atribuindo peso maior para partidas recentes.

Como nesses modelos as forças de ataque e defesa de cada equipe são estáticas ao longo do tempo, [Rue e Salvesen \(2000\)](#) incorporaram a estrutura de [Dixon e Coles \(1997\)](#) em um modelo dinâmico, para estudar as propriedades de variação no tempo de uma equipe. Em suas análises, truncaram o número de gols em um máximo de 5, argumentando que uma quantidade superior não fornece informação adicional sobre o ataque e defesa de um time. Outros autores também utilizaram abordagens dinâmicas, como [Crowder et al. \(2002\)](#) e [Owen \(2011\)](#), os quais desenvolveram modelos com base no apresentado por [Maher \(1982\)](#), e [Koopman e Lit \(2013\)](#), que consideraram correlação entre o número de gols.

Partindo de abordagens distintas, no lugar de modelar o número de gols feito por cada equipe, [Karlis e Ntzoufras \(2009\)](#) concentraram-se na diferença do saldo de gols entre os oponentes, com uma modelagem similar àquela proposta por [Lee \(1997\)](#), e [Brillinger \(2008\)](#) propôs modelar diretamente as probabilidades de vitória, empate e derrota para obter a pontuação final das equipes e, assim, obter, a probabilidade de qualquer time ser campeão ou ficar entre os quatro primeiros colocados.

As modelagens dos autores supracitados exigem um conhecimento estatístico mais aprofundado, evidenciando que modelar fenômenos do nosso cotidiano nem sempre é uma tarefa fácil. Neste trabalho concentrar-nos-emos em abordar a modelagem de forma tangível, utilizando e aprimorando o modelo proposto por [Lima et al. \(2012\)](#), que modela as probabilidades de vitória, empate e derrota das equipes, em uma partida de futebol, com base nas frequências desses resultados em disputas anteriores.

Com intuito de ilustrar os modelos apresentados e desenvolvidos, uma aplicação em um campeonato de futebol de pontos corridos, Série A do Campeonato Brasileiro de 2018, foi realizada. No desenvolvimento dessa pesquisa, foi utilizado o *software* Excel, por não demandar um conhecimento aprofundado de técnicas e lógica de programação, permitindo o seu uso em sala de aula.

1.1 Objetivos do Trabalho

1.1.1 Objetivo geral

A proposta deste trabalho é abordar a modelagem de forma tangível, com a finalidade de desenvolver a percepção da importância da Matemática por meio de uma aplicação em um evento cotidiano, um campeonato de futebol.

1.1.2 Objetivos específicos

- Apresentar um modelo da literatura e propor um aprimoramento por meio da construção de um novo modelo probabilístico de previsão de resultados para um campeonato de pontos corridos no futebol;
- Mostrar, de forma prática, uma aplicação dos modelos na Série A do Campeonato Brasileiro de Futebol de 2018 e apresentar testes de qualidade preditiva, que visam averiguar a eficácia da previsão de resultados futuros;
- Realizar um estudo de simulação com base em jogos ocorridos;
- Propor atividades para o desenvolvimento em sala de aula.

1.2 Organização do Trabalho

No [Capítulo 2](#), são abordados conceitos de modelagem matemática e probabilidade, com intuito de facilitar o entendimento dos demais capítulos. No [Capítulo 3](#), será apresentado um modelo da literatura e proposto um aprimoramento por meio da construção de um novo modelo probabilístico de previsão de resultados para um campeonato de pontos corridos no futebol, além da utilização de um campeonato fictício para ilustrar os métodos desenvolvidos. Já no [Capítulo 4](#) é realizada uma aplicação dos modelos para a Série A do Campeonato Brasileiro de 2018, a fim de realizar análises de resultados para dados reais, e são apresentados critérios de seleção de modelo e métricas para comparar os modelos, do ponto de vista preditivo, sendo elas o erro quadrático médio e a medida de De Finetti ([De Finetti, 1972](#)). Além disso, um estudo de simulação é realizado como forma alternativa para verificar a eficácia dos modelos desenvolvidos. Baseado nos conteúdos abordados,

no [Capítulo 5](#) são propostas atividades que podem ser realizadas em sala de aula para alunos do Ensino Médio. Finalmente, no [Capítulo 6](#) serão apresentadas as conclusões e as propostas futuras.

Capítulo 2

Introdução à Teoria das Probabilidades

Este capítulo discorre sobre modelagem matemática e noções fundamentais da teoria das Probabilidades. São abordadas, também, definições, propriedades e teoremas acerca da teoria dos conjuntos, espaço amostral, eventos e uma breve introdução à lei dos grandes números, com a finalidade de propiciar embasamento teórico necessário para a compreensão dos próximos capítulos. Além de exemplos clássicos encontrados na literatura, são apresentados alguns exemplos envolvendo o futebol. Os conceitos e as demonstrações exibidos neste capítulo podem ser melhor explorados nos livros “Probabilidade: Aplicações à Estatística” ([Meyer, 1970](#)) e “Probabilidade: um curso em nível intermediário” ([James, 1996](#)).

2.1 Modelagem

A motivação da modelagem matemática está atrelada à necessidade de antecipar os resultados dos fenômenos aos quais somos expostos no dia a dia. A modelagem matemática nos possibilita transformar problemas do cotidiano em problemas matemáticos, permitindo resolver e interpretar as suas soluções na linguagem do mundo real ([Bassanezi, 2009](#)).

Os fenômenos podem ser descritos ou replicados através de experimentos, os quais são classificados em determinísticos (não aleatórios) e probabilísticos (aleatórios). Os experimentos determinísticos são caracterizados a priori, isto é, o resultado é conhecido antes mesmo que ele ocorra, com base em condições preestabelecidas. Como exemplos de experimentos determinísticos temos: gravitação, causa-efeito e aceleração.

Neste trabalho, iremos abordar experimentos probabilísticos, que são aqueles cujo o resultado não sabemos a priori, ou seja, o resultado empírico não é previsível. Como exemplos de experimentos probabilísticos pode-se citar o resultado do lançamento de uma moeda, lançamento de um dado e o resultado de uma partida de futebol, objeto deste

estudo.

2.2 Introdução à Teoria dos Conjuntos

O estudo sobre teoria dos conjuntos é atribuído ao russo George Ferdinand Cantor (1845 – 1918). Embora esta teoria seja extensa, para expor os conceitos básicos do modelo probabilístico que desejamos desenvolver, apenas algumas noções fundamentais serão necessárias.

Podemos definir conjunto como sendo um agrupamento de elementos com características comuns. Usualmente, conjuntos são representados por letras maiúsculas, por exemplo A , B , Ω , etc. Existem algumas maneiras de expressar um conjunto A :

- (a) Descrição dos elementos: $A = \{1,2,3,4\}$;
- (b) Por meio de palavras: A é formado por todos os números reais entre 0 e 1, inclusive;
- (c) Simbolicamente: $A = \{x \in \mathbb{R} | 0 \leq x \leq 1\}$; isto é, A é o conjunto de todos os x , onde x é um número real entre 0 e 1, inclusive.

Um conjunto pode apresentar um número infinito ou finito de elementos, sendo denominados por conjunto infinito e conjunto finito, respectivamente. Os conjuntos infinitos podem ser segregados em conjuntos enumeráveis e conjuntos não enumeráveis. De forma particular, um conjunto finito pode conter apenas um elemento ou não possuir elementos, classificados por conjunto unitário e conjunto vazio (denotado por \emptyset ou $\{\}$), respectivamente.

Os objetos que individualmente formam o conjunto A são denominados elementos de A . Denota-se por $a \in A$ quando a pertence ao conjunto A e por $a \notin A$, caso contrário.

Dizemos que o conjunto A está contido no conjunto B , $A \subseteq B$, se, e somente se, todos os elementos do conjunto A pertencem ao conjunto B . Simbolicamente,

$$A \subseteq B \iff \forall a, \text{ se } a \in A \text{ então } a \in B.$$

Caso contrário, se existe pelo menos um elemento a tal que $a \in A$ e $a \notin B$, então $A \not\subseteq B$, isto é, A não está contido em B .

Definição 2.1 (Igualdade de conjuntos) *Se $A \subseteq B$ e $B \subseteq A$ então A e B contêm os mesmos elementos. Dizemos que A e B são iguais e indicamos por $A = B$. Caso contrário, $A \neq B$.*

Definição 2.2 (Subconjunto próprio) *Um conjunto A é um subconjunto próprio de um conjunto B se $A \subseteq B$ e $A \neq B$, denotado por $A \subset B$.*

Definição 2.3 (Conjunto Complementar) *O complementar de um conjunto A é representada por A^C , formando um novo conjunto cujos elementos são todos aqueles que não pertencem ao conjunto A . Formalmente, escrevemos “ A complementar” da seguinte maneira:*

$$A^C = \{a \mid a \notin A\}.$$

2.2.1 Operações com conjuntos

Definição 2.4 (União) *A união de dois conjuntos A e B é representada por $A \cup B$, formando um novo conjunto cujos elementos são aqueles que pertencem a um dos dois conjuntos, ou ambos, isto é, os elementos pertencem a pelo menos um dos conjuntos. Formalmente, escrevemos “ A união B ” da seguinte maneira:*

$$A \cup B = \{a \mid a \in A \text{ ou } a \in B\}.$$

Propriedade 2.1 $A \cup A = A$.

Demonstração. $a \in (A \cup A) \iff a \in A \text{ ou } a \in A \iff a \in A$. ■

Propriedade 2.2 $A \cup \emptyset = A$.

Demonstração. $a \in (A \cup \emptyset) \iff a \in A \text{ ou } a \in \emptyset \iff a \in A$. ■

Propriedade 2.3 $A \cup B = B \cup A$.

Demonstração. $a \in A \cup B \iff a \in A \text{ ou } a \in B \iff a \in B \text{ ou } a \in A \iff a \in B \cup A$. ■

Propriedade 2.4 $A \cup (B \cup C) = (A \cup B) \cup C$.

Demonstração. $a \in A \cup (B \cup C) \iff a \in A \text{ ou } (a \in B \text{ ou } a \in C) \iff (a \in A \text{ ou } a \in B) \text{ ou } a \in C \iff a \in (A \cup B) \cup C$. ■

Propriedade 2.5 $A \subset (A \cup B)$.

Demonstração. $a \in A \implies a \in A \text{ ou } a \in B \iff a \in A \cup B$. ■

Propriedade 2.6 $A \subset B$ se, e somente se, $A \cup B = B$.

Demonstração.

(\implies) Suponha que $A \subset B$. Então,

$$a \in A \cup B \iff a \in A \text{ ou } a \in B \stackrel{A \subset B}{\iff} a \in B \text{ ou } a \in B \iff a \in B.$$

(\impliedby) Por outro lado, se $A \cup B = B$, então

$$a \in A \implies a \in A \text{ ou } a \in B \iff a \in (A \cup B) \stackrel{A \cup B = B}{\iff} a \in B$$
. ■

Definição 2.5 (Intersecção) *A intersecção de dois conjuntos A e B é representada por $A \cap B$, formando um novo conjunto cujos elementos são aqueles que pertencem simultaneamente aos dois conjuntos. Formalmente, escrevemos “A intersecção B” da seguinte maneira:*

$$A \cap B = \{a \mid a \in A \text{ e } a \in B\}.$$

Propriedade 2.7 $A \cap A = A$.

Demonstração. $a \in (A \cap A) \iff a \in A \text{ e } a \in A \iff a \in A$. ■

Propriedade 2.8 $A \cap \emptyset = \emptyset$.

Demonstração. Suponha que exista um elemento a tal que $a \in (A \cap \emptyset)$, então $a \in A$ e $a \in \emptyset$, absurdo! Logo, não existe a tal que $a \in (A \cap \emptyset)$, ou seja, $A \cap \emptyset = \emptyset$. ■

Propriedade 2.9 $A \cap B = B \cap A$.

Demonstração. $a \in A \cap B \iff a \in A \text{ e } a \in B \iff a \in B \text{ e } a \in A \iff a \in B \cap A$. ■

Propriedade 2.10 $A \cap (B \cap C) = (A \cap B) \cap C$.

Demonstração. $a \in A \cap (B \cap C) \iff a \in A \text{ e } (a \in B \text{ e } a \in C) \iff (a \in A \text{ e } a \in B) \text{ e } a \in C \iff a \in (A \cap B) \cap C$. ■

Propriedade 2.11 $(A \cap B) \subset A$.

Demonstração. $a \in (A \cap B) \iff a \in A \text{ e } a \in B \implies a \in A$. ■

Propriedade 2.12 $A \subset B$ se, e somente se, $A \cap B = A$.

Demonstração.

(\implies) Suponha que $A \subset B$. Então,

$$a \in A \xrightarrow{A \subset B} a \in A \text{ e } a \in B \iff a \in A \cap B.$$

(\impliedby) Por outro lado, se $A \cap B = A$, então

$$a \in A \xrightarrow{A \cap B = A} a \in (A \cap B) \iff a \in A \text{ e } a \in B \implies a \in B. \quad \blacksquare$$

Definição 2.6 (Diferença) *A diferença de dois conjuntos A e B é representada por $A - B$, formando um novo conjunto cujos elementos pertencem ao conjunto A , mas não pertencem ao conjunto B . Formalmente, escrevemos “a diferença entre A e B ” da seguinte maneira:*

$$A - B = \{a \mid a \in A \text{ e } a \notin B\}.$$

Definição 2.7 (Conjuntos Disjuntos) *Dois conjuntos A e B são denominados disjuntos quando não possuem elementos em comum, isto é, $A \cap B = \emptyset$.*

Definição 2.8 (Conjuntos Disjuntos - Generalização) *Uma família de conjuntos é dita disjunta dois a dois ou mutuamente disjunta se, dados dois conjuntos quaisquer da família, eles forem disjuntos. Mais formalmente falando, seja A_n , $n \in \mathbb{N}$, uma família de conjuntos disjuntos, então:*

$$A_i \cap A_j = \emptyset, \forall i, j \leq n, i \neq j.$$

2.3 Espaço Amostral e Eventos

Definição 2.9 (Espaço Amostral) *Dado um experimento ε , o espaço amostral é o conjunto de todos os possíveis resultados do experimento. Usualmente, esse conjunto é representado por S ou Ω .*

Como o espaço amostral é um conjunto, ele também pode ser classificado em finito, infinito enumerável ou infinito não numerável. Salientamos que o resultado de um experimento não é necessariamente um número.

Exemplo 2.1 *Lançar uma moeda e verificar a face voltada para cima. O espaço amostral é o conjunto $\Omega = \{\text{cara}, \text{coroa}\}$.*

Exemplo 2.2 *Lançar um dado de seis faces e observar o número obtido na face superior. O espaço amostral é o conjunto $\Omega = \{1, 2, 3, 4, 5, 6\}$.*

Exemplo 2.3 *Realizar uma partida de futebol e observar o número total de gols. O espaço amostral é o conjunto $\Omega = \{0, 1, 2, \dots\}$.*

Exemplo 2.4 *Realizar uma partida de futebol e observar o resultado da disputa. O espaço amostral é o conjunto $\Omega = \{\text{vitória do mandante}, \text{empate}, \text{derrota do mandante}\}$.*

Definição 2.10 (Evento) *Um evento é um subconjunto do espaço amostral associado à um experimento, em outras palavras, é um conjunto de resultados possíveis. De forma geral, os eventos serão denotados por letras maiúsculas.*

Exemplo 2.5 *Ocorrer face par no lançamento de um dado de seis faces.*

Exemplo 2.6 *Ocorrer 3 faces caras em 3 lançamentos de uma moeda.*

Exemplo 2.7 *Ocorrer empate em uma partida de futebol.*

Note que o próprio espaço amostral é um evento, também conhecido como evento certo, enquanto que o conjunto \emptyset é denominado de evento impossível.

Definição 2.11 (Eventos mutuamente excludentes) *Dois eventos A e B são ditos mutuamente excludentes se eles não puderem ocorrer juntos, ou seja, $A \cap B = \emptyset$.*

Exemplo 2.8 *Em uma partida de futebol, os eventos “mandante ganhar” e “visitante marcar um gol” podem ocorrer simultaneamente, não caracterizando eventos mutuamente excludentes.*

Exemplo 2.9 *Em uma partida de futebol, os eventos “mandante ganhar” e “visitante não sofrer gols” são eventos mutuamente excludentes, pois não podem ocorrer simultaneamente.*

Definição 2.12 (Frequência relativa) *Seja ε um experimento, com espaço amostral Ω , repetido n vezes. A frequência relativa f_A do evento $A \subseteq \Omega$ é a razão entre o número de vezes que o evento ocorre, n_A , e o número de repetições ocorridas, n . Assim, $f_A = \frac{n_A}{n}$.*

Considere A e B eventos de um espaço amostral Ω associado a um experimento ε repetido n vezes. A seguir são apresentadas propriedades acerca de frequência relativa.

Propriedade 2.13 $0 \leq f_A \leq 1$.

Demonstração.

$$0 \leq n_A \leq n \implies \frac{0}{n} \leq \frac{n_A}{n} \leq \frac{n}{n} \implies 0 \leq f_A \leq 1. \blacksquare$$

Propriedade 2.14 $f_A = 1$ se, e somente se, A ocorrer em todas as n repetições.

Demonstração.

$$f_A = 1 \iff \frac{n_A}{n} = 1 \iff n_A = n. \blacksquare$$

Propriedade 2.15 $f_A = 0$ se, e somente se, A nunca ocorrer nas n repetições.

Demonstração.

$$f_A = 0 \iff \frac{n_A}{n} = 0 \iff n_A = 0. \blacksquare$$

Propriedade 2.16 *Se A e B forem eventos mutuamente excludentes e se $f_{(A \cup B)}$ for a frequência relativa associada ao evento $A \cup B$, então $f_{(A \cup B)} = f_A + f_B$.*

Demonstração. Suponha A e B eventos mutuamente excludentes. Então nas n_A ocorrências do evento A , não é observada nenhuma ocorrência do evento B . Do mesmo modo, nas n_B ocorrências do evento B , não se observa ocorrências do evento A . Dessa forma, o número de ocorrências do evento $A \cup B$ é $n_A + n_B$, assim

$$f_{(A \cup B)} = \frac{n_A + n_B}{n} = \frac{n_A}{n} + \frac{n_B}{n} = f_A + f_B. \blacksquare$$

2.4 Noções Fundamentais de Probabilidade

A probabilidade é o ato de atribuímos pesos aos eventos, a fim de avaliar quão verossímil será a ocorrência de cada evento do experimento. Uma possível maneira de tratar a questão seria repetir o experimento um número grande de vezes, calcular f_A e utilizar este número.

As demonstrações dos teoremas exibidos nesta seção podem ser encontradas na bibliografia base deste capítulo. A seguir serão apresentados os termos básicos associados à modelagem dos experimentos probabilísticos.

Definição 2.13 (Probabilidade) *Seja ε um experimento e Ω um espaço amostral. A probabilidade, que será denotada por \mathbb{P} , é uma função $\mathbb{P} : \Omega \rightarrow [0,1]$ que satisfaz os seguintes axiomas:*

Axiomas da Probabilidade:

1. $0 \leq \mathbb{P}(A) \leq 1$ para todo A eventos de Ω .
2. $\mathbb{P}(\Omega) = 1$.
3. Se A e B forem eventos mutuamente excludentes, então

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

4. Se A_1, A_2, \dots , forem, dois a dois, eventos mutuamente excludentes, então

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Qualquer função \mathbb{P} que atribua pesos a eventos associados a um espaço amostral e que satisfaça os axiomas 1, 2, 3 e 4 supracitados será denominada probabilidade.

Teorema 2.1 $\mathbb{P}(\emptyset) = 0$.

Demonstração. Para qualquer evento A , $A = A \cup \emptyset$. Como A e \emptyset são mutualmente excludentes, podemos escrever

$$\mathbb{P}(A) = \mathbb{P}(A \cup \emptyset) = \mathbb{P}(A) + \mathbb{P}(\emptyset) \implies \mathbb{P}(\emptyset) = 0. \blacksquare$$

Teorema 2.2 *Se A^C for o evento complementar de A , então $\mathbb{P}(A) = 1 - \mathbb{P}(A^C)$.*

Demonstração. Seja Ω o espaço amostral, temos que $\Omega = A \cup A^C$, sendo essa união disjunta, pois $A \cap A^C = \emptyset$. Logo,

$$\mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^C) \implies \mathbb{P}(A) = \mathbb{P}(\Omega) - \mathbb{P}(A^C) \implies \mathbb{P}(A) = 1 - \mathbb{P}(A^C). \blacksquare$$

Teorema 2.3 *Se A e B forem dois eventos quaisquer, então*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Demonstração. Temos que $(A \cup B) = A \cup (B - A)$, sendo A e $(B - A)$ mutualmente excludentes. Logo,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \cup (B - A)) = \mathbb{P}(A) + \mathbb{P}(B - A).$$

Além disso, $B = (B - A) \cup (A \cap B)$, com $(B - A)$ e $(A \cap B)$ mutualmente excludentes, então

$$\mathbb{P}(B) = \mathbb{P}((B - A) \cup (A \cap B)) = \mathbb{P}(B - A) + \mathbb{P}(A \cap B) \iff \mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Dos resultados acima, podemos concluir que

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \blacksquare$$

Teorema 2.4 *Se A , B e C forem três eventos quaisquer, então*

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

Demonstração. Temos que

$$A \cup B \cup C = (A \cup B) \cup C = (A \cup B) \cup (C - (A \cup B)),$$

sendo esta união disjunta. Então pelo axioma 3 da probabilidade, temos que

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A \cup B) + \mathbb{P}(C - (A \cup B)). \quad (2.10)$$

Pelo Teorema 2.3, na equação (2.10) temos que

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) + \mathbb{P}(C - (A \cup B)).$$

Além disso, $C = (C - (A \cup B)) \cup (C \cap (A \cup B))$, sendo que esta união é disjunta, portanto

$$\mathbb{P}(C - (A \cup B)) = \mathbb{P}(C) - \mathbb{P}(C \cap (A \cup B)). \quad (2.11)$$

Também temos que $C \cap (A \cup B) = (C \cap A) \cup (C \cap (B - A))$ e esta união é disjunta. Daí

$$\mathbb{P}(C \cap (A \cup B)) = \mathbb{P}(A \cap C) + \mathbb{P}(C \cap (B - A)). \quad (2.12)$$

Finalmente, $C \cap B = (A \cap B \cap C) \cup (C \cap (B - A))$, com $(A \cap B \cap C)$ e $(C \cap (B - A))$

mutuamente excludentes, o que implica que

$$\mathbb{P}(C \cap (B - A)) = \mathbb{P}(B \cap C) - \mathbb{P}(A \cap B \cap C). \quad (2.13)$$

Logo, por (2.10), (2.11), (2.12) e (2.13), concluímos que

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C). \blacksquare$$

Teorema 2.5 *Se $A \subset B$, então $\mathbb{P}(A) \leq \mathbb{P}(B)$.*

Demonstração. Temos que se $A \subset B$ então $B = A \cup (B - A)$, sendo que esta união é disjunta. Portanto, utilizando o axioma 3 da probabilidade, segue que $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B - A)$. Como $\mathbb{P}(B - A) \geq 0$, temos então que $\mathbb{P}(B) \geq \mathbb{P}(A)$. ■

Teorema 2.6 *Se $A \subset B$, então $\mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A)$.*

Demonstração. Observe que, como $A \subset B$, $B = A \cup (B - A)$, e, ainda, que $A \cap (B - A) = \emptyset$. Assim, pelo o axioma 3 da probabilidade, segue que $\mathbb{P}(B) = \mathbb{P}(A \cup (B - A)) = \mathbb{P}(A) + \mathbb{P}(B - A) \implies \mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A)$ ■

2.5 Espaços Amostrais Finitos

Definição 2.14 (Eventos Equiprováveis) *Se um experimento ε tem como espaço amostral $\Omega = \{e_1, e_2, \dots, e_n\}$, com um número finito de elementos, dizemos que os eventos elementares e_i são equiprováveis se todos tiverem a mesma probabilidade de ocorrer, ou seja,*

$$\mathbb{P}(\{e_i\}) = \frac{1}{n}.$$

Dessa forma, podemos definir a probabilidade de um evento $A = \{e_{j_1}, e_{j_2}, \dots, e_{j_k}\}$, composto por $k \leq n$ elementos, como sendo:

$$\mathbb{P}(A) = \frac{\text{número de casos favoráveis a } A}{\text{número de casos possíveis de } \Omega} = \frac{k}{n}. \quad (2.14)$$

Exemplo 2.15 *Considere o experimento “lançar um dado” e seja A o evento “obter face par”. O espaço amostral do experimento é $\Omega = \{1, 2, 3, 4, 5, 6\}$ e o evento em questão é dado por $A = \{2, 4, 6\}$. Desse modo, pela equação (2.14),*

$$\mathbb{P}(A) = \frac{\text{número de casos favoráveis a } A}{\text{número de casos possíveis de } \Omega} = \frac{3}{6} = 0,5.$$

2.6 Lei dos Grandes Números

A Lei dos Grandes Números possui grande importância, sendo considerada uma das principais leis assintóticas da Estatística e o princípio de simulações. Essa lei surgiu devido à necessidade de estimar a probabilidade de ocorrência de um evento natural. Para a sua definição formal são necessários conhecimentos prévios sobre vários conceitos estatísticos (tais como: esperança, variância, espaço de probabilidades, independência de variáveis aleatórias, distribuições de probabilidade, entre outros). Estes conceitos serão omitidos aqui, sendo que iremos apenas enunciar os resultados que possibilitarão a formalização da Lei dos Grandes Números. Entretanto, é possível abordar essa lei de modo introdutório e intuitivo, conforme o exemplo a seguir:

Exemplo 2.16 *Considere o lançamento de um dado equilibrado, cujas faces são numeradas de 1 a 6. A probabilidade de obter o número 3 é de $1/6$. Agora, repetindo o experimento uma quantidade suficientemente grande de vezes é possível notar que a frequência relativa da face 3, em relação ao número de repetições do experimento, aproxima-se da probabilidade calculada.*

Em suma, a Lei dos Grandes Números nos diz que quanto maior o número de repetições de um experimento, mais o resultado experimental aproxima-se da probabilidade esperada (veja [Tabela 5.1](#) e [Tabela 5.2](#)).

2.6.1 Lei Fraca e Lei Forte - formalização

Esta subseção requer conhecimentos prévios de Estatística. A formalização da Lei dos Grandes Números pode ser dividida em duas partes: Lei Fraca (de Chebyshev e de Khintchine) e Lei Forte (de Kolmogorov), sendo que a principal diferença entre elas se dá no tipo de convergência, como veremos a seguir.

Definição 2.15 *Considere ε um experimento e Ω o espaço amostral associado a esse experimento. Uma função X , que associa a cada elemento $\omega \in \Omega$ um número real, $X(\omega)$, é denominada variável aleatória. Ou seja, variável aleatória (v.a.) é um característico numérico do resultado de um experimento.*

Definição 2.16 (Convergência em probabilidade) *Sejam $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias e X uma variável aleatória definidas no mesmo espaço de probabilidade. Dizemos que X_n converge em probabilidade para X e denotamos $X_n \xrightarrow{P} X$ se, para todo $\delta > 0$,*

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \delta) = 0.$$

A principal ideia da convergência em probabilidade é que, quando n é arbitrariamente grande, a probabilidade da diferença $|X_n - X|$ ser maior do que qualquer número positivo δ tende a zero.

Definição 2.17 (Convergência quase certa) *Sejam $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias e X uma variável aleatória definidas no mesmo espaço de probabilidade. Dizemos que X_n converge quase certamente para X , isto é, $X_n \xrightarrow{q.c.} X$ se*

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} X_n = X \right) = 1$$

ou, equivalentemente,

$$\mathbb{P} \left(\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\} \right) = 1.$$

Proposição 2.1 *Se uma sequência de variáveis aleatórias $\{X_n\}$ sobre um espaço de probabilidade converge quase certamente para uma variável aleatória X então $X_n \rightarrow X$ em probabilidade.*

Teorema 2.7 (Lei Fraca de Chebyshev) *Sejam $\{X_i\}$ uma sequência enumerável de variáveis aleatórias independentes duas a duas. Suponha que a sequência $\{X_i\}$ tem variância finita e uniformemente limitada, ou seja, existe uma constante $c \in \mathbb{R}$ tal que $\text{Var}[X_i] \leq c$. Então a sequência $\{X_i\}$ satisfaz a Lei Fraca dos Grandes Números:*

$$\frac{S_n - \mathbb{E}[S_n]}{n} \xrightarrow{P} 0$$

onde $S_n = \sum_{i=1}^n X_i$ e $\mathbb{E}[S_n]$ é a esperança da v.a. S_n .

Teorema 2.8 (Lei Fraca de Khintchine) *Sejam $\{X_i\}$ uma sequência enumerável de variáveis aleatórias independentes e identicamente distribuídas e integráveis com média μ . Então $\{X_i\}$, $i \in \mathbb{N}$, satisfazem a Lei Fraca dos Grandes Números:*

$$\frac{S_n}{n} \xrightarrow{P} \mu.$$

Corolário 2.1 (Lei Fraca dos Grandes Números de Bernoulli) *Seja $\{X_i\}$ uma sequência de ensaios de Bernoulli independentes, com mesma probabilidade de sucesso em que*

$$E[X_i] = p \text{ e } S_n = \sum_{i=1}^n X_i,$$

então

$$\frac{S_n}{n} \xrightarrow{P} p.$$

Teorema 2.9 (Primeira Lei Forte de Kolmogorov) *Sejam $\{X_i\}$ uma sequência de variáveis aleatórias independentes e integráveis, e suponha que*

$$\sum_{i=1}^{\infty} \frac{\text{Var}[X_i]}{i^2} < \infty,$$

então

$$\sum_{i=1}^n \frac{X_i - \mathbb{E}[X_i]}{n} \xrightarrow{q.c.} 0.$$

Teorema 2.10 (Lei Forte de Kolmogorov) *Seja $\{X_i\}$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas e com $\mathbb{E}[X_i] = \mu < \infty$. Então*

$$\sum_{i=1}^n \frac{X_i}{n} \xrightarrow{q.c.} \mu.$$

Proposição 2.2 (Lema de Borel-Cantelli) *Sejam A_1, A_2, \dots eventos aleatórios num espaço de probabilidade.*

(a) *Se $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, então $\mathbb{P}(A_n \text{ infinitas vezes}) = 0$;*

(b) *Se $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ e os A_n são independentes, então $\mathbb{P}(A_n \text{ infinitas vezes}) = 1$.*

Proposição 2.3 *Consideremos uma sequência de ensaios de Binomiais independentes com probabilidade p_n de sucesso no n -ésimo ensaio. Seja $X_n = 1$ se o n -ésimo ensaio é sucesso e $X_n = 0$ em caso de fracasso. Então vale o seguinte:*

- *Se $\sum_{n=1}^{\infty} p_n = +\infty$, então $\mathbb{P}\left(\sum_{n=1}^{\infty} X_n = \infty\right) = 1$;*
- *Se $\sum_{n=1}^{\infty} p_n < \infty$, então $\mathbb{P}\left(\sum_{n=1}^{\infty} X_n < \infty\right) = 1$.*

Em outras palavras, a Proposição 2.3 diz que

- $\sum_{n=1}^{\infty} p_n < \infty \implies$ um número finito de sucessos, quase certamente;
- $\sum_{n=1}^{\infty} p_n = \infty \implies$ um número infinito de sucessos, quase certamente.

Exemplo 2.17 *Um clássico exemplo metafórico encontrado na literatura afirma que um macaco digitando aleatoriamente em uma máquina de escrever, por um intervalo de tempo infinito, irá, quase certamente, criar um texto qualquer escolhido, como por exemplo a obra completa de William Shakespeare.*

De fato, embora ínfima, é razoável assumir que o macaco possui uma probabilidade $p > 0$ de realizar essa façanha em um ensaio. Pela Proposição 2.3, tomando $p_n = p$, para todo n , há probabilidade 1 dele escrever as obras de Shakespeare um número infinito de vezes.

Exemplo 2.18 *Uma adaptação do exemplo anterior nos diz que se infinitos macacos digitarem simultaneamente em infinitas máquinas de escrever, pelo menos um irá, quase certamente, produzir a obra completa de William Shakespeare perfeitamente sem erros.*

Capítulo 3

Modelos de Previsão de Futebol

No decorrer deste capítulo, será apresentado um modelo de previsão para resultados de uma partida de futebol. Tendo como referência esse modelo, será apresentado e construído um novo modelo, que levará em conta o fator de mando de campo, a fim de aferir com maior precisão os resultados futuros. Para ilustrar os métodos desenvolvidos neste trabalho, um campeonato fictício de pontos corridos será utilizado.

3.1 Especificações dos Modelos

Vários fatores em conjunto definem o desempenho de um time, tais como as habilidades individuais de cada jogador, o tipo de treinamento oferecido pelo clube, entre outros. Entretanto, essas informações podem ser resumidas na observação, apenas, das quantidades de vitórias, empates e derrotas.

A fim de se obter um modelo simples, porém bom o suficiente para prever o resultado de uma partida de futebol, alguns fatos podem ser levados em consideração:

- o modelo adotado deve considerar a habilidade de ambas as equipes que disputam uma determinada partida;
- a habilidade de uma determinada equipe deve ser avaliada por meio de uma medida da capacidade de vitória, empate ou derrota;
- as equipes mandantes, em geral, tem alguma vantagem por estarem jogando em casa;
- ao observar o desempenho de um time ao longo de um campeonato, deve-se considerar os times contra os quais ele jogou;
- a habilidade atual de um time está mais intimamente atrelada à sua atuação recente do que a partidas mais antigas.

Outra observação a ser levada em conta é o fato de que, quando os times não possuem habilidades técnicas semelhantes, a diferença de sucesso das equipes modificará o resultado de um jogo. Assim, um time com maior êxito terá uma maior probabilidade de vencer um time menos bem-sucedido.

3.2 Modelo para Previsão

Quando ocorre uma partida de futebol em um campeonato de pontos corridos são observados 3 resultados: vitória do mandante, empate ou vitória do visitante, sendo intuitivo afirmar que a soma das probabilidades desses resultados deve ser igual a 1.

Com base nos n jogos anteriores a uma determinada partida, podemos definir a força de um time A , que tenha participado de j_A partidas, sendo $j_A < n$, como sendo a frequência que ele obteve cada um dos 3 resultados possíveis: vitória (V_A), empate (E_A) e derrota (D_A). Observe que $V_A + E_A + D_A = 1$, sendo esses valores calculados conforme abaixo:

$$\begin{aligned} V_A &= \frac{\text{número de vitórias de } A}{j_A}, \\ E_A &= \frac{\text{número de empates de } A}{j_A}, \\ D_A &= \frac{\text{número de derrotas de } A}{j_A}. \end{aligned} \quad (3.1)$$

Baseado na construção acima, [Lima et al. \(2012\)](#) sugeriram a criação de um vetor força para cada equipe a partir dos seus desempenhos nas rodadas anteriores. Dessa maneira, o vetor força de uma equipe A (genérica) pode ser descrito como

$$F_A = \left\{ (V_A, E_A, D_A) \in \mathbb{R}^3 : V_A + E_A + D_A = 1, V_A \geq 0, E_A \geq 0, D_A \geq 0 \right\}, \quad (3.2)$$

com as coordenadas descritas em (3.1). Observe que a obtenção do vetor F_A é simples, dada sua construção, tornando possível o cálculo sem auxílio de nenhum material externo, como computadores, por exemplo.

No que segue, consideraremos um conjunto de dados composto por n partidas e as equipes numeradas de 1 a T . Para a partida i , suponha VM_i, E_i e $DM_i \in \{0,1\}$, com $VM_i + E_i + DM_i = 1$, os resultados das partidas, sendo VM_i a vitória do mandante (m_i), E_i o empate e DM_i a derrota do mandante (v_i), $i = 1, 2, \dots, n$. A [Tabela 3.1](#) ilustra a estrutura dos confrontos do campeonato.

Um problema em utilizar apenas o vetor força para calcular as probabilidades de resultado para uma equipe é o fato de que, neste cálculo, desconsidera-se a influência do adversário. Dessa forma, a probabilidade de vitória da equipe mandante poderá ser diferente da probabilidade de derrota de seu adversário, o que seria um absurdo.

Para determinar as probabilidades dos resultados entre os times A e B ($P_{A \wedge B}$), devemos

Tabela 3.1: Estrutura dos Confrontos.

Partida	Mandante	Visitante	Resultados Possíveis		
1	m_1	v_1	VM_1	E_1	DM_1
2	m_2	v_2	VM_2	E_2	DM_2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	m_n	v_n	VM_n	E_n	DM_n

considerar o vetor força de ambas as equipes (F_A e F_B), conforme sugerido no início desse capítulo, no primeiro item da lista exibido na página 27. Assim, do ponto de vista do mandante, a probabilidade de vitória será dada em função das frequências observadas de vitória do mandante (V_A) e derrota do visitante (D_B); a probabilidade de empate considera as frequências desses resultados para ambas as equipes (E_A e E_B); e a probabilidade de derrota é calculada em função das frequências observadas de derrota do mandante (D_A) e vitória do visitante (V_B).

Dessa forma, temos elaborado nosso primeiro modelo, o qual denominaremos por modelo de Previsão no Futebol Simples (PFS) e será dado por:

$$P_{A \wedge B} = \left(\frac{V_A + D_B}{2}, \frac{E_A + E_B}{2}, \frac{D_A + V_B}{2} \right) = (P_{V_{A \wedge B}}, P_{E_{A \wedge B}}, P_{D_{A \wedge B}}) \quad (3.3)$$

em que V_A, E_A, D_A, V_B, E_B e D_B são dados por (3.1).

A premissa da qual o modelo PFS parte é a de que, se um time vence uma partida, aumenta sua probabilidade de vencer a partida seguinte. O mesmo princípio vale para o empate e a derrota. Desse modo, garante-se uma maior probabilidade de vitória para as equipes mais eficientes e maior probabilidade de derrota para os times que possuíram uma campanha mais fraca. A cada rodada os parâmetros dos modelos são atualizados com o intuito de atingir uma boa acurácia nas probabilidades dos resultados das próximas disputas.

Note que, por construção,

$$\begin{aligned} 0 &\leq V_A, V_B, E_A, E_B, D_A, D_B \leq 1 \implies \\ 0 &\leq V_A + D_B \leq 2 \implies \\ 0 &\leq \frac{V_A + D_B}{2} \leq 1 \implies \\ 0 &\leq P_{V_{A \wedge B}} \leq 1. \end{aligned} \quad (3.4)$$

De forma análoga, temos que $0 \leq P_{E_{A \wedge B}}, P_{D_{A \wedge B}} \leq 1$.

Além disso, o espaço amostral é dado por $\Omega = \{V, E, D\}$, sendo $V \cap E = V \cap D = E \cap D = V \cap E \cap D = \emptyset$. Dessa forma, considerando $\mathbb{P}_P(\mathbb{E})$ o cálculo de probabilidade de um evento \mathbb{E} baseado no modelo PFS, temos que

$$\begin{aligned}
\mathbb{P}_P(\Omega) &= \mathbb{P}_P(V \cup E \cup D) = \mathbb{P}_P(V) + \mathbb{P}_P(E) + \mathbb{P}_P(D) \\
&= P_{V_{A \wedge B}} + P_{E_{A \wedge B}} + P_{D_{A \wedge B}} \\
&= \frac{V_A + D_B}{2} + \frac{E_A + E_B}{2} + \frac{D_A + V_B}{2} \\
&= \frac{V_A + D_B + E_A + E_B + D_A + V_B}{2} \\
&= \frac{(V_A + E_A + D_A) + (V_B + E_B + D_B)}{2} \\
&= \frac{1 + 1}{2} = 1.
\end{aligned} \tag{3.5}$$

3.3 Construção de um Novo Modelo

Quando presenciamos um evento esportivo, é possível notar que a torcida exerce certa interferência sobre o curso da disputa, podendo, inclusive, condicionar o juiz da partida a aplicar penalidades. [Nevill et al. \(2002\)](#) analisaram os efeitos que a presença ou ausência de ruído da torcida poderiam ter sobre a tomada de decisão dos árbitros, revelando uma grande influência e maior incerteza desses profissionais quando pressionados.

Inicialmente, [Pollard \(1986\)](#) e, mais tarde, [Clarke e Norman \(1995\)](#) verificaram que as equipes que jogam em casa possuem uma quantidade maior de vitória que os visitantes, sendo que, no futebol, essas vantagens são mais relevantes nas principais divisões. A ideia de considerar uma vantagem *a priori* para o time mandante expressa uma forma direta e menos subjetiva de ponderar a influência da torcida.

É nesse sentido que propomos um novo modelo, original, desenvolvido nessa dissertação. Fundamentado no modelo PFS, dado em (3.3), e nas especificações apresentadas pelos autores acima, partiremos da hipótese de que os times que possuem o mando de campo devem ter algum incremento nas suas probabilidades de sucesso, visando inferir, com mais precisão, os resultados de cada partida. Essa vantagem de jogar em casa será denominada, no presente trabalho, como “efeito casa”.

Com base no que foi apresentado anteriormente, o cálculo das probabilidades de vitória, empate e derrota entre duas equipes A e B , em uma determinada partida, ainda levará em conta os seus vetores força F_A e F_B , conforme descritos em (3.2). Desse modo, concentrar-nos-emos apenas em como calcular e como incluir o efeito casa (EC) em nossas previsões.

Para o cálculo desse fator, sugerimos uma abordagem semelhante ao descrito em (3.1). Assim, considerando n jogos, podemos definir EC como sendo a frequência em que os times mandantes (m) obtiveram em cada um dos 3 resultados possíveis em um jogo: vitória (V_m), empate (E_m) e derrota (D_m). Novamente, temos que a soma desses elementos será 1, ou seja, $V_m + E_m + D_m = 1$, os quais são obtidos conforme segue:

$$V_m = \frac{\text{número de vitórias dos mandantes}}{n}, \quad (3.6)$$

$$E_m = \frac{\text{número de empates dos mandantes}}{n}, \quad (3.7)$$

$$D_m = \frac{\text{número de derrotas dos mandantes}}{n}. \quad (3.8)$$

Conseqüentemente, o vetor força do efeito casa será dado como uma função desses elementos, de acordo com o exibido abaixo:

$$EC = \{(V_m, E_m, D_m) \in \mathbb{R}^3 : V_m + E_m + D_m = 1, V_m \geq 0, E_m \geq 0, D_m \geq 0\}. \quad (3.9)$$

Do modo em que foi conduzida essa construção, é intuitiva a ideia de que o vetor força do EC será incluído nos cálculos das probabilidades das partidas. De forma mais aprofundada, para um time A e B mandante e visitante, respectivamente, além das frequências observadas de vitória do mandante (V_A) e derrota do visitante (D_B), a probabilidade de vitória da equipe A também levará em consideração a componente V_m (eq. (3.6)). Analogamente, além dos itens já apresentados, para a probabilidade de empate, será acrescida a frequência de empates (eq. (3.7)) e, finalmente, para a probabilidade de vitória do visitante (time B) também será considerado o D_m (eq. (3.8)).

Agora, devemos pensar em como adicionar o EC nos cálculos de $P_{V_{A \wedge B}}$, $P_{E_{A \wedge B}}$ e $P_{D_{A \wedge B}}$ definidos em (3.3). A maneira mais trivial seria uma soma simples. Desse modo,

$$P_{V_{A \wedge B}}^* = P_{V_{A \wedge B}} + V_m = \frac{V_A + D_B}{2} + V_m = \frac{V_A + D_B + 2V_m}{2}.$$

Porém, o efeito casa teria o dobro do peso das demais componentes, fugindo da nossa proposta inicial. Logo, para retirar o peso desses elementos, considere

$$P_{V_{A \wedge B}}^* = P_{V_{A \wedge B}} + \frac{V_m}{2}, P_{E_{A \wedge B}}^* = P_{E_{A \wedge B}} + \frac{E_m}{2} \text{ e } P_{D_{A \wedge B}}^* = P_{D_{A \wedge B}} + \frac{D_m}{2}.$$

Assim, obtemos modelo P^* , dado por

$$\begin{aligned} P_{A \wedge B}^* &= (P_{V_{A \wedge B}}^*, P_{E_{A \wedge B}}^*, P_{D_{A \wedge B}}^*) \\ &= \left(\frac{V_A + D_B + V_m}{2}, \frac{E_A + E_B + E_m}{2}, \frac{D_A + V_B + D_m}{2} \right). \end{aligned} \quad (3.10)$$

Nosso espaço amostral Ω mantêm-se o mesmo, ou seja, $\Omega = \{V, E, D\}$, sendo $V \cap E = V \cap D = E \cap D = V \cap E \cap D = \emptyset$. Dessa forma, considerando $\mathbb{P}_{P^*}(\mathbb{E})$ o cálculo de probabilidade de um evento \mathbb{E} baseado no modelo P^* , temos que

$$\begin{aligned}
\mathbb{P}_{P^*}(\Omega) &= \mathbb{P}_{P^*}(V \cup E \cup D) = \mathbb{P}_{P^*}(V) + \mathbb{P}_{P^*}(E) + \mathbb{P}_{P^*}(D) \\
&= P_{V_{A \wedge B}}^* + P_{E_{A \wedge B}}^* + P_{D_{A \wedge B}}^* \\
&= \frac{V_A + D_B + V_m}{2} + \frac{E_A + E_B + E_m}{2} + \frac{D_A + V_B + D_m}{2} \\
&= \frac{V_A + D_B + V_m + E_A + E_B + E_m + D_A + V_B + D_m}{2} \\
&= \frac{(V_A + E_A + D_A) + (V_B + E_B + D_B) + (V_m + E_m + D_m)}{2} \\
&= \frac{1 + 1 + 1}{2} = \frac{3}{2} \neq 1,
\end{aligned} \tag{3.11}$$

o que é um absurdo, pois pelo axioma 2 da probabilidade, $\mathbb{P}(\Omega) = 1$.

Para contornarmos essa restrição, basta um pequeno ajuste: vamos multiplicar nosso modelo pelo inverso do valor obtido em (3.11). Assim, nosso modelo é dado conforme segue:

$$\begin{aligned}
P'_{A \wedge B} &= (\mathbb{P}_{P^*}(\Omega))^{-1} \times (P_{A \wedge B}^*) \\
&= \left(\frac{3}{2}\right)^{-1} \left(\frac{V_A + D_B + V_m}{2}, \frac{E_A + E_B + E_m}{2}, \frac{D_A + V_B + D_m}{2}\right) \\
&= \left(\frac{2}{3}\right) \left(\frac{V_A + D_B + V_m}{2}, \frac{E_A + E_B + E_m}{2}, \frac{D_A + V_B + D_m}{2}\right) \\
&= \left(\frac{2}{3}\right) \left(\frac{V_A + D_B + V_m}{2}, \frac{E_A + E_B + E_m}{2}, \frac{D_A + V_B + D_m}{2}\right) \\
&= \left(\frac{V_A + D_B + V_m}{3}, \frac{E_A + E_B + E_m}{3}, \frac{D_A + V_B + D_m}{3}\right) \\
&= (P'_{V_{A \wedge B}}, P'_{E_{A \wedge B}}, P'_{D_{A \wedge B}}).
\end{aligned} \tag{3.12}$$

Com isso, finalizamos a construção do modelo proposto, o qual, em sumário, é representado por

$$P'_{A \wedge B} = \left(\frac{V_A + D_B + V_m}{3}, \frac{E_A + E_B + E_m}{3}, \frac{D_A + V_B + D_m}{3}\right), \tag{3.13}$$

que denominaremos por modelo de Previsão no Futebol com Efeito Casa (PFC) e será utilizado no decorrer deste trabalho, tanto como uma alternativa para previsão esportiva quanto para efeito de comparação com o modelo apresentado inicialmente.

Além da interpretação individual para os parâmetros, já apresentados para o modelo PFS, os parâmetros do modelo PFC também serão influenciados pela performance dos demais times na condição de mandante, em que o seu resultado aumentará a probabilidade dos próximos mandantes obterem o mesmo resultado que os anteriores, ou seja, o resultado do mandante de uma disputa qualquer do campeonato trará influência para os demais

mandantes.

3.4 Campeonato Fictício

Para ilustrar os métodos deste Capítulo, introduzimos um campeonato fictício de pontos corridos entre as equipes Corinthians, Palmeiras, Santos e São Paulo, com os resultados dos jogos apresentados na [Tabela 3.2](#) abaixo.

Tabela 3.2: Jogos do Campeonato Fictício.

Jogo	Mandante	Gols	Visitante	Local
1	Corinthians	1 0	Palmeiras	Arena Corinthians
2	São Paulo	1 0	Santos	Estádio Cícero Pompeu de Toledo
3	Palmeiras	3 1	São Paulo	Allianz Parque
4	Corinthians	1 1	Santos	Arena Corinthians
5	Santos	1 1	Palmeiras	Estádio Urbano Caldeira
6	São Paulo	3 1	Corinthians	Estádio Cícero Pompeu de Toledo
7	Palmeiras	1 0	Corinthians	Allianz Parque
8	Santos	0 0	São Paulo	Estádio Urbano Caldeira
9	São Paulo	0 2	Palmeiras	Estádio Cícero Pompeu de Toledo
10	Santos	1 0	Corinthians	Estádio Urbano Caldeira
11	Palmeiras	3 2	Santos	Allianz Parque
12	Corinthians	1 1	São Paulo	Arena Corinthians

A [Tabela 3.3](#) apresenta a classificação final dos times do campeonato fictício, os pontos ganhos (PG), os números de jogos (J), vitórias (V), empates (E), derrotas (D), gols pró (gols realizados - GP), gols contra (gols sofridos - GC) e saldo de gols (SG).

Tabela 3.3: Classificação final do Campeonato Fictício.

Posição	Time	PG	J	V	E	D	GP	GC	SG
1	Palmeiras	13	6	4	1	1	10	5	5
2	São Paulo	8	6	2	2	2	6	7	-1
3	Santos	6	6	1	3	2	5	6	-1
4	Corinthians	5	6	1	2	3	4	7	-3

Suponha agora, que gostaríamos de calcular as probabilidades dos resultados de jogos futuros entre os times participantes desse campeonato, baseando-nos nas 12 disputas exibidas na [Tabela 3.2](#). Para tal, devemos calcular o vetor força F_T (eq. (3.2)) para cada um dos 4 times: Palmeiras (P), São Paulo (SP), Santos (S) e Corinthians (C).

Conforme observado na [Tabela 3.3](#), o time do Palmeiras, que foi a equipe mais bem classificada, participou de 6 conflitos, conquistando 4 vitórias, além de 1 empate e 1 derrota.

Assim sendo,

$$\begin{aligned} V_P &= \frac{4}{6} \approx 0,66667, \\ E_P &= \frac{1}{6} \approx 0,16667, \\ D_P &= \frac{1}{6} \approx 0,16667 \end{aligned} \quad (3.14)$$

e, conseqüentemente,

$$F_P = (V_P, E_P, D_P) = (0,66667, 0,16667, 0,16667). \quad (3.15)$$

Considerando os resultados prévios das demais equipes e, de maneira análoga ao realizado em (3.14), temos os vetores força para os outros times do campeonato em questão:

$$F_{SP} = (V_{SP}, E_{SP}, D_{SP}) = (0,33333, 0,33333, 0,33333); \quad (3.16)$$

$$F_S = (V_S, E_S, D_S) = (0,16667, 0,50000, 0,33333); \quad (3.17)$$

$$F_C = (V_C, E_C, D_C) = (0,16667, 0,33333, 0,50000). \quad (3.18)$$

Agora, para realizar o cálculo do efeito casa desse campeonato, se faz necessário a utilização das informações dos resultados por confronto, pois não é suficiente os desfechos resumidos por time, como foi para os cálculos dos vetores força. De acordo com a [Tabela 3.2](#), das 12 disputas, as equipes mandantes levaram a melhor em 7 (jogos 1, 2, 3, 6, 7, 10 e 11), o número de empates foi 4 partidas (jogos 4, 5, 8 e 12) e o êxito dos visitantes é observado em apenas um confronto (jogo 9). Dessa forma,

$$\begin{aligned} V_m &= \frac{7}{12} \approx 0,58333, \\ E_m &= \frac{4}{12} \approx 0,33333, \\ D_m &= \frac{1}{12} \approx 0,08333 \end{aligned} \quad (3.19)$$

e, conseqüentemente para o campeonato fictício, o vetor do efeito casa é dado por

$$EC = (V_m, E_m, D_m) = (0,58333, 0,33333, 0,08333). \quad (3.20)$$

Dando seqüência, possuímos todas as ferramentas necessárias para o cálculo de vitória, empate e derrota de uma partida entre quaisquer dois dos quatros times do nosso campeonato para ambos os modelos apresentados nas seções [Seção 3.2](#) e [Seção 3.3](#).

De forma ilustrativa, calcularemos os resultados para um confronto futuro entre a equipe do Corinthians e do Palmeiras, sendo eles mandante e visitante, respectivamente. Para o modelo PFS, devemos considerar apenas os vetores F_C (eq. (3.18)) e F_P (eq. (3.15)).

Assim,

$$\begin{aligned}
 P_{C \wedge P} &= \left(\frac{V_C + D_P}{2}, \frac{E_C + E_P}{2}, \frac{D_C + V_P}{2} \right) \\
 &= \left(\frac{0,33333}{2}, \frac{0,50000}{2}, \frac{1,16667}{2} \right) \\
 &= (0,16667, 0,25000, 0,58333) \\
 &= (P_{V_{C \wedge P}}, P_{E_{C \wedge P}}, P_{D_{C \wedge P}}).
 \end{aligned} \tag{3.21}$$

Com isso, concluímos que, com base no modelo PFS, essa eventual partida resultaria na vitória do Corinthians, empate e vitória do Palmeiras com probabilidade de 16,7%, 25% e 58,3%, respectivamente.

No caso de uma disputa entre Corinthians e do Palmeiras, porém com o Palmeiras sendo o mandante, os resultados serão:

$$\begin{aligned}
 P_{P \wedge C} &= \left(\frac{V_P + D_C}{2}, \frac{E_P + E_C}{2}, \frac{D_P + V_C}{2} \right) \\
 &= \left(\frac{1,16667}{2}, \frac{0,50000}{2}, \frac{0,33333}{2} \right) \\
 &= (0,58333, 0,25000, 0,16667) \\
 &= (P_{V_{P \wedge C}}, P_{E_{P \wedge C}}, P_{D_{P \wedge C}}).
 \end{aligned} \tag{3.22}$$

Como esse modelo não penaliza as equipes visitantes, facilmente pode-se verificar que obteremos exatamente as mesmas probabilidades, ou seja, $P_{V_{C \wedge P}} = P_{D_{P \wedge C}}$, $P_{E_{C \wedge P}} = P_{E_{P \wedge C}}$ e $P_{D_{C \wedge P}} = P_{V_{P \wedge C}}$.

Em vista disso, voltando à [Tabela 3.2](#), como a ordem (mandante ou visitante) dos times na partida não é relevante para o modelo, o número total de disputas (de $p = 2$ equipes) diferentes entre os $n = 4$ times do campeonato fictício é dado por

$$\mathcal{C}_{n,p} = \binom{n}{p} = \frac{n!}{p!(n-p)!} = \frac{4!}{2!(4-2)!} = \frac{4 \times 3 \times 2!}{2 \times 2!} = 6. \tag{3.23}$$

Na [Tabela 3.4](#) temos as probabilidades de vitória ($P_{V_{m \wedge v}}$), empate ($P_{E_{m \wedge v}}$) e derrota ($P_{D_{m \wedge v}}$) do mandante (m) contra o visitante (v), para cada disputa (i) possível, conforme [\(3.23\)](#).

Nota-se que a maior probabilidade de um resultado é a vitória do Palmeiras em cima da equipe do Corinthians (jogo 1, 58,3%), condizendo com o esperado, dado que o Palmeiras ostentou a melhor campanha (4 vitórias em 6 jogos) e o Corinthians obteve insucesso em metade de suas partidas, sendo esse o maior número de derrotas do campeonato (3 derrotas em 6 jogos). Além disso, as menores probabilidades observadas (16,7% nas partidas 1 e 5) são das derrotas do Palmeiras, que perdeu apenas 1 confronto, pelas equipes do Santos e do Corinthians, que apresentaram unicamente 1 vitória cada. Novamente, essas probabilidades não fogem do expectável.

Tabela 3.4: Probabilidade dos jogos para o modelo PFS.

Jogo (i)	Mandante (m)	Visitante (v)	$P_{V_{m \wedge v}}$	$P_{E_{m \wedge v}}$	$P_{D_{m \wedge v}}$
1	Corinthians	Palmeiras	0,16667	0,25000	0,58333
2	São Paulo	Santos	0,33333	0,41667	0,25000
3	Palmeiras	São Paulo	0,50000	0,25000	0,25000
4	Corinthians	Santos	0,25000	0,41667	0,33333
5	Santos	Palmeiras	0,16667	0,33333	0,50000
6	São Paulo	Corinthians	0,41667	0,33333	0,25000

Agora vejamos o que acontece com o modelo PFC. Iniciemos, mais uma vez, por um confronto entre a equipe do Corinthians e do Palmeiras, sendo eles mandante e visitante, respectivamente. Dessa vez, além dos vetores F_C (eq. (3.18)) e F_P (eq. (3.15)), também será considerado o vetor EC (eq. (3.20)). Por conseguinte,

$$\begin{aligned}
P'_{C \wedge P} &= \left(\frac{V_C + D_P + V_m}{3}, \frac{E_C + E_P + E_m}{3}, \frac{D_C + V_P + D_m}{3} \right) \\
&= \left(\frac{0,91667}{3}, \frac{0,83333}{3}, \frac{1,25000}{3} \right) \\
&= (0,30556, 0,27778, 0,41667) \\
&= (P'_{V_{C \wedge P}}, P'_{E_{C \wedge P}}, P'_{D_{C \wedge P}}).
\end{aligned} \tag{3.24}$$

Com isso, concluímos que, com base no modelo PFC, essa eventual partida resultaria na vitória do Corinthians, empate e vitória do Palmeiras com probabilidade de 30,6%, 27,8% e 41,7%, respectivamente.

No caso de uma disputa entre essas equipes, porém com o mando de campo invertido, os resultados serão:

$$\begin{aligned}
P'_{P \wedge C} &= \left(\frac{V_P + D_C + V_m}{3}, \frac{E_P + E_C + E_m}{3}, \frac{D_P + V_C + D_m}{3} \right) \\
&= \left(\frac{1,75000}{3}, \frac{0,83334}{3}, \frac{0,41667}{3} \right) \\
&= (0,58333, 0,27778, 0,13889) \\
&= (P'_{V_{P \wedge C}}, P'_{E_{P \wedge C}}, P'_{D_{P \wedge C}}).
\end{aligned} \tag{3.25}$$

Verifica-se que a probabilidade de triunfo do Palmeiras aumentou para 58,3%, ao passo que o sucedimento do Corinthians decaiu para 13,9%, ou seja, como mandantes, as chances de sucesso das equipes foram maiores, salientando a vantagem de se jogar em casa. Logo, matematicamente, temos que $P'_{V_{P \wedge C}} > P'_{D_{C \wedge P}}$ e $P'_{V_{C \wedge P}} > P'_{D_{P \wedge C}}$. De forma geral, para quaisquer dois times A e B ,

$$V_m > D_M \iff (V_m + V_A + D_B) > (D_m + V_A + D_B) \iff P'_{V_{A \wedge B}} > P'_{D_{B \wedge A}}.$$

Em sequência, como a ordem (mandante ou visitante) dos times na partida é primordial para o modelo, o número total de disputas (de $p = 2$ equipes) diferentes entre os $n = 4$ times do campeonato fictício é dado por

$$\mathcal{A}_{n,p} = \frac{n!}{(n-p)!} = \frac{4!}{(4-2)!} = \frac{4 \times 3 \times 2!}{2!} = 12. \quad (3.26)$$

Na [Tabela 3.5](#) temos as probabilidades de vitória ($P'_{V_{m\wedge v}}$), empate ($P'_{E_{m\wedge v}}$) e derrota ($P'_{D_{m\wedge v}}$) do mandante (m) contra o visitante (v), para cada disputa (i) possível, conforme (3.26). As partidas estão dispostas de modo que, o “jogo de volta” (mando de campo alternando no confronto entre as mesmas equipes) de uma partida k , em que $0 < k \leq 6$, será a partida $k + 6$.

Tabela 3.5: Probabilidade dos jogos para o modelo PFC.

Jogo (i)	Mandante (m)	Visitante (v)	$P'_{V_{m\wedge v}}$	$P'_{E_{m\wedge v}}$	$P'_{D_{m\wedge v}}$
1	Corinthians	Palmeiras	0,30556	0,27778	0,41667
2	São Paulo	Santos	0,41667	0,38889	0,19444
3	Palmeiras	São Paulo	0,52778	0,27778	0,19444
4	Corinthians	Santos	0,36111	0,38889	0,25000
5	Santos	Palmeiras	0,30556	0,33333	0,36111
6	São Paulo	Corinthians	0,47222	0,33333	0,19444
7	Palmeiras	Corinthians	0,58333	0,27778	0,13889
8	Santos	São Paulo	0,36111	0,38889	0,25000
9	São Paulo	Palmeiras	0,36111	0,27778	0,36111
10	Santos	Corinthians	0,41667	0,38889	0,19444
11	Palmeiras	Santos	0,52778	0,33333	0,13889
12	Corinthians	São Paulo	0,36111	0,33333	0,30556

Nota-se que as maiores probabilidades de resultado são as vitórias do Palmeiras como mandante sobre os demais times (58,3% contra o Corinthians no jogo 7 e 52,8% contra o Santos e o São Paulo nos jogos 11 e 3, respectivamente). Já as menores, são exatamente a derrota dessa equipe, com o privilégio do mando de campo, para o times do Corinthians (jogo 7, 13,9%) e do Santos (jogo 11, 13,9%). As equipes visitantes só lograram a maior probabilidade em duas partidas (41,7% no jogo 1, sendo essa a maior probabilidade de vitória dos visitantes, e 36,1% no jogo 5), no qual o Palmeiras foi o protagonista desse feito, consolidando sua soberania nesse campeonato. O empate, como o resultado mais provável, também só pôde ser observado em duas partidas (38,9% nos jogo 4 e 8). Além disso, a disputa entre o São Paulo (mandante) e Palmeiras (visitante) foi a única que retratou a mesma probabilidade de vitória para ambas as equipes (jogo 9, 36,1%).

Capítulo 4

Aplicação para Dados Reais e um Estudo de Simulação

Neste capítulo, iniciaremos com a exibição do Campeonato Brasileiro de 2018 - Série A e um retrospecto dessa competição, com a finalidade de realizar análises de resultados para dados reais. Além de apresentar algumas medidas de qualidade de ajuste e de qualidade preditiva encontradas na literatura, será verificada a eficácia dos modelos do [Capítulo 3](#), baseado em algumas rodadas para o ajuste, na previsão dos resultados do torneio em questão. Por fim, é introduzida a ideia de simulação, com um estudo apresentado e discutido.

4.1 Dados da Série A do Campeonato Brasileiro de 2018

A Série A do Campeonato Brasileiro de Futebol de 2018, oficialmente Brasileirão Assaí 2018 – Série A por motivos de patrocínio, foi a 62ª edição da principal divisão do futebol brasileiro. A disputa teve o mesmo regulamento dos anos anteriores, quando foi implementado o sistema de pontos corridos. Houve pausa durante a Copa do Mundo de 2018, realizada entre 14 de junho e 15 de julho, na Rússia.

Esta foi a edição com mais clubes nordestinos na história dos pontos corridos, com quatro representantes: Bahia, Ceará, Sport e Vitória. Desde 2001, quando o campeonato ainda era disputado em sistema misto e contava com 28 participantes, o Nordeste não tinha quatro clubes na elite. Por sua vez, o Sudeste foi a região com mais representantes, onze no total, enquanto o Sul contou com cinco times na competição.

O título foi definido na penúltima rodada: o Palmeiras conquistou o seu décimo título brasileiro, após vencer o Vasco da Gama em São Januário, por 1–0, isolando-se no posto de maior campeão nacional do futebol brasileiro. Além do campeão, Flamengo, Internacional e Grêmio também se classificaram diretamente à fase de grupos da Copa Libertadores

da América de 2019, além do Cruzeiro, campeão da Copa do Brasil. Após o término do campeonato, o Atlético Paranaense conquistou a Copa Sul-Americana e também se garantiu na fase de grupos da Libertadores. Já São Paulo e Atlético Mineiro, quinto e sexto colocados respectivamente, se classificaram para a segunda fase do torneio continental.

Na parte de baixo da tabela, o Paraná foi a primeira equipe a ter o rebaixamento para a Série B de 2019 decretado, na 32ª rodada. A equipe paranaense empatou com o Vitória em Curitiba, e foi prejudicada com resultado positivo do Sport (1–0) diante do Ceará. O próprio Vitória também foi rebaixado, mesmo depois de entrar em campo, onde empatou sem gols contra o Grêmio, prejudicado por outro resultado de 0–0, entre Corinthians e Chapecoense. América Mineiro e Sport completaram a relação de rebaixados na última rodada: a equipe mineira foi derrotada pelo Fluminense (1–0), no Maracanã, em confronto direto contra a queda; já a equipe pernambucana foi rebaixada mesmo após vencer o Santos (2–1), no Recife.

Tabela 4.1: Classificação final da Série A do Campeonato Brasileiro de 2018.

Posição	Time	PG	J	V	E	D	GP	GC	SG
1	Palmeiras	80	38	23	11	4	64	26	38
2	Flamengo	72	38	21	9	8	59	29	30
3	Internacional	69	38	19	12	7	51	29	22
4	Grêmio	66	38	18	12	8	48	27	21
5	São Paulo	63	38	16	15	7	46	34	12
6	Atlético-MG	59	38	17	8	13	56	43	13
7	Atlético-PR	57	38	16	9	13	54	37	17
8	Cruzeiro	53	38	14	11	13	34	34	0
9	Botafogo	51	38	13	12	13	38	46	-8
10	Santos	50	38	13	11	14	46	40	6
11	Bahia	48	38	12	12	14	39	41	-2
12	Fluminense	45	38	12	9	17	32	46	-14
13	Corinthians	44	38	11	11	16	34	35	-1
14	Chapecoense	44	38	11	11	16	34	50	-16
15	Ceará	44	38	10	14	14	32	38	-6
16	Vasco	43	38	10	13	15	41	48	-7
17	Sport Recife	42	38	11	9	18	35	57	-22
18	América-MG	40	38	10	10	18	30	47	-17
19	Vitória	37	38	9	10	19	36	63	-27
20	Paraná	23	38	4	11	23	18	57	-39

A [Tabela 4.1](#) apresenta a classificação final da Série A do Campeonato Brasileiro de 2018, os times, os pontos ganhos, os números de jogos, vitórias, empates, derrotas, gols pró (gols realizados), gols contra (gols sofridos) e saldo de gols. Durante os 380 jogos da competição, houve 202 vitórias dos times mandantes (53,16% do total), 110 empates (28,95% do total) e 68 vitórias dos visitantes (17,89% do total).

4.1.1 Formato da competição e posições relevantes

Vinte clubes participam do Campeonato Brasileiro. Durante o decorrer da temporada, cada clube joga duas vezes contra os outros (em um sistema de pontos corridos), uma vez em seu estádio e a outra no de seu adversário, em um total de 38 jogos. As equipes recebem três pontos por vitória e um por empate. Não são atribuídos pontos para derrotas.

As equipes são classificadas pelo total de pontos, depois pelo número de vitórias e, em seguida, pelo saldo de gols. Em caso de empate entre dois ou mais clubes na pontuação, os critérios de desempate são os seguintes: maior número de vitórias; maior saldo de gols; maior número de gols pró (marcados); confronto direto; menor número de cartões vermelhos recebidos; menor número de cartões amarelos recebidos; sorteio.

Além do título de Campeão do torneio para a equipe mais bem classificada, algumas colocações também oferecem acesso a outros torneios e rebaixamento para a série B do mesmo campeonato.

A partir da temporada de 2016, os seis melhores times do Brasileirão se qualificam para a Copa Libertadores, com os quatro melhores times entrando diretamente na fase de grupos. As equipes do sétimo ao décimo segundo lugar se classificam para a Copa Sul-Americana. Caso algum clube da zona de classificação já estiver classificado, por outros meios, para a Copa Libertadores, a vaga segue para a próxima equipe mais bem colocada no campeonato.

Ademais, as quatro equipes que amarguraram as últimas colocações são rebaixadas para a Série B do Campeonato Brasileiro, sendo elas substituídas, na edição seguinte, pelos quatro clubes mais bem classificados dessa outra série.

4.2 Método Utilizado

Para que previsões de partidas sejam feitas e para ajuste dos modelos aos dados da Série A do Campeonato Brasileiro de 2018 é necessário que jogos anteriores tenham ocorrido. Tendo isso em vista, dados de no mínimo 19 rodadas (primeiro turno) serão utilizados, pois todos os times enfrentaram-se ao menos uma vez durante o campeonato, considerando uma estrutura de ajuste-previsão. Dessa forma, com ajuste das n primeiras rodadas, pode-se fazer previsão para qualquer rodada m tal que $n < m \leq 38$, em especial para partidas da rodada seguinte, ou seja, $m = n + 1$ tal que $n < 38$.

A [Tabela 4.2](#) apresenta a estrutura do ajuste-previsão a ser utilizada, em que RC denota as rodadas consideradas para o ajuste, e RP denota as rodadas a serem previstas.

Tabela 4.2: Esquema das rodadas utilizadas para ajuste dos modelos e previsões.

RC	RP	RC	RP	RC	RP
1 ^a - 19 ^a	20 ^a - 38 ^a	1 ^a - 26 ^a	27 ^a - 38 ^a	1 ^a - 32 ^a	33 ^a - 38 ^a
1 ^a - 20 ^a	21 ^a - 38 ^a	1 ^a - 27 ^a	28 ^a - 38 ^a	1 ^a - 33 ^a	34 ^a - 38 ^a
1 ^a - 21 ^a	22 ^a - 38 ^a	1 ^a - 28 ^a	29 ^a - 38 ^a	1 ^a - 34 ^a	35 ^a - 38 ^a
1 ^a - 22 ^a	23 ^a - 38 ^a	1 ^a - 29 ^a	30 ^a - 38 ^a	1 ^a - 35 ^a	36 ^a - 38 ^a
1 ^a - 23 ^a	24 ^a - 38 ^a	1 ^a - 30 ^a	31 ^a - 38 ^a	1 ^a - 36 ^a	37 ^a - 38 ^a
1 ^a - 24 ^a	25 ^a - 38 ^a	1 ^a - 31 ^a	32 ^a - 38 ^a	1 ^a - 37 ^a	38 ^a
1 ^a - 25 ^a	26 ^a - 38 ^a				

4.3 Medidas de Qualidade Preditiva

Para a construção de um modelo, diversas variáveis podem ser consideradas, como por exemplo no futebol: idade e habilidade dos jogadores, experiência do técnico, número de título do clube, entre outros. Desse modo, o modelo pode ser inflado com várias covariáveis subjetivas, as quais não necessariamente agregam boas informações. É nesse sentido que existem métodos para avaliar a relevância do excesso de parâmetros para um determinado modelo.

Na literatura, os critérios mais utilizados para seleção de modelos, após as estimativas dos parâmetros, são os critérios AIC (Akaike, 1974) e BIC (Schwarz et al., 1978). Definimos $AIC = -2 \log L(\theta) + 2d$ e $BIC = -2 \log L(\theta) + d \log N$; em que $L(\theta)$ é a função de máxima verossimilhança (Akaike, 1998, para mais informações), d é a dimensão do vetor de parâmetros e N é o tamanho da amostra. Quanto menor o valor do AIC e BIC, melhor o ajuste do modelo.

Os critérios supracitados devem ser utilizados quando se tem diversos modelos e deseja-se selecionar aqueles que possuem um ajuste mais eficiente aos dados. Como nesse estudo apenas adicionamos o efeito casa em um modelo já existente, podemos focar na comparação do ponto de vista preditivo. A seguir, apresentamos algumas medidas de qualidade preditiva que serão utilizadas.

4.3.1 Medida de De Finetti

Um método de verificação da qualidade de previsões bastante difundido na literatura é a medida de De Finetti (De Finetti, 1972), que consiste na consideração de um simplexo contido no \mathbb{R}^3 como representação geométrica do conjunto das possíveis previsões probabilísticas. Assim, os vértices desse simplexo correspondem às ocorrências dos resultados, e os demais pontos, a todas as outras possíveis previsões. Dessa forma,

$$S = \left\{ (PV, PE, PD) \in \mathbb{R}^3 : PV + PE + PD = 1, PV \geq 0, PE \geq 0, PD \geq 0 \right\},$$

em que PV , PE e PD denotam, respectivamente, as probabilidades de vitória, empate e derrota do time mandante.

A medida de distância de De Finetti corresponde à distância euclidiana quadrática entre o ponto correspondente à probabilidade prevista e o vértice correspondente ao resultado efetivamente observado. Pode-se construir, para mais de uma previsão, um índice dado pela média aritmética das distâncias de De Finetti, chamado “Medida de De Finetti”.

No futebol, associam-se os vértices $(1,0,0)$, $(0,1,0)$ e $(0,0,1)$ para vitória, empate e derrota do time mandante, respectivamente. Para uma determinada partida, associa-se o ponto $(PV, PE, PD) \in S$ ao seu vetor de probabilidades.

Dessa maneira, a distância de De Finetti será: $(PV - 1)^2 + (PE - 0)^2 + (PD - 0)^2$, quando a equipe mandante vencer, $(PV - 0)^2 + (PE - 1)^2 + (PD - 0)^2$, se ocorrer empate, e $(PV - 0)^2 + (PE - 0)^2 + (PD - 1)^2$, quando a equipe visitante ganhar.

Pode-se considerar como referência para comparação as probabilidades de vitória, empate e derrota equiprováveis ($PV = PE = PD = 1/3$); nesse caso, tem-se que a distância de De Finetti é igual a $(1/3 - 1)^2 + (1/3 - 0)^2 + (1/3 - 0)^2 = 2/3$. Portanto, observa-se que métodos com boa qualidade de previsão possuem uma medida de De Finetti inferior a $2/3$. Note que a média aritmética dessas medidas pode ser utilizada como critério para avaliar a qualidade preditiva dos modelos ajustados. Por exemplo, em uma rodada com 10 jogos, calculá-se a distância de De Finetti de cada partida e depois a média aritmética dos valores (veja [Tabela 4.4](#)).

4.3.2 Erro Quadrático Médio

O Erro Quadrático Médio é amplamente utilizado na teoria Estatística. Essa métrica é obtida basicamente pela média dos quadrados das diferenças entre os valores previstos e os valores reais, de forma que, quanto menor esse valor, melhor a previsão.

Uma desvantagem em nossos modelos é que eles só nos trazem informações sobre a probabilidade de cada resultado e não o número de gols, por exemplo, como outros modelos encontrados na literatura ([Lee, 1997](#); [Maher, 1982](#)). Devido a tal fato, só é possível uma comparação entre o erro quadrático médio dos pontos obtidos. Esse valor será denotado por EQM .

4.4 Análise da Qualidade Preditiva

Nesta seção, mostraremos a análise da qualidade preditiva de nossos modelos para cada partida, com base em confrontos previamente realizados para os ajustes dos modelos. Além da visão por partida, também será analisada a média das distâncias de De Finetti para verificar a qualidade preditiva por rodada. Será denotado como “Satisfatório” (S) quando a distância ou medida de De Finetti for menor que $2/3$ e, caso contrário, “Insatisfatório”

(IS).

A partir de 19 rodadas consideradas para os ajustes, a [Tabela 4.3](#) apresenta o número de vezes que os modelos obtiveram a classificação Satisfatório e Insatisfatório para cada uma das 190 partidas a serem previstas. Observe que, no modelo PFS, apenas 85 jogos receberam o status de insatisfatório, refletindo em uma frequência de 55,26% de bons resultados. Já o modelo PFC possui um melhor desempenho em relação ao outro modelo, por conseguir em 110 confrontos uma distância de De Finetti satisfatória, ou seja, 57,89% das partidas.

Tabela 4.3: Classificação das distâncias de De Finetti para 19 rodadas de ajuste.

Classificação	PFS	PFC
Satisfatório	105	110
Insatisfatório	85	80
Frequência	55,26%	57,89%

A título de exibir outra maneira de aferir a qualidade preditiva, a análise da medida de De Finetti foi realizada por rodada, cujos resultados estão resumidos na [Tabela 4.4](#), que apresenta a rodada a ser prevista, a medida de De Finetti, a classificação, em relação às 10 partidas de cada rodada, a média geral e o percentual de classificações S, para os modelos PFS e PFC.

Tabela 4.4: Medida de De Finetti para cada rodada com base em um ajuste de 19 rodadas.

Rodada	PFS		PFC	
	Média	Classificação	Média	Classificação
20	0,64377	S	0,63332	S
21	0,68947	IS	0,58830	S
22	0,72327	IS	0,69636	IS
23	0,66579	S	0,60398	S
24	0,57978	S	0,54564	S
25	0,69820	IS	0,62113	S
26	0,61163	S	0,60675	S
27	0,62936	S	0,59937	S
28	0,67230	IS	0,66792	IS
29	0,65900	S	0,59645	S
30	0,67355	IS	0,65163	S
31	0,68102	IS	0,62487	S
32	0,62784	S	0,62018	S
33	0,75263	IS	0,71745	IS
34	0,57355	S	0,57729	S
35	0,59183	S	0,56728	S
36	0,63906	S	0,56592	S
37	0,62493	S	0,62066	S
38	0,67313	IS	0,61584	S
Média Geral	0,65316	57,89%	0,61686	84,21%

Note que, o modelo PFS apresentou medida de De Finetti satisfatória em 11 rodadas, representando 57,89% de um total de 19 rodadas. Já o modelo PFC, obteve resultado satisfatório em 84,21% dos casos, ou seja, apenas 3 rodadas (22, 28 e 33) foram insatisfatórias, sobressaindo-se em relação ao outro modelo. Ademais, as médias das medidas de De Finetti foram 0,65316 e 0,61686 para os modelos PFS e PFC, respectivamente.

De maneira análoga, as análises foram também reproduzidas para 28 rodadas de ajuste, a fim de averiguar a eficácia dos modelos com o campeonato mais próximo de sua conclusão.

Agora, das 100 partidas a serem previstas, o modelo PFS apresentou 56 jogos com distância de De Finetti satisfatória (frequência de 56%) e o modelo PFC obteve uma performance melhor, com 59 das disputas satisfatórias (59% dos confrontos), conforme a [Tabela 4.5](#).

Tabela 4.5: Classificação das distâncias de De Finetti para 28 rodadas de ajuste.

Classificação	PFS	PFC
Satisfatório	56	59
Insatisfatório	44	41
Frequência	56%	59%

A [Tabela 4.6](#) mostra as 10 rodadas a serem previstas, a medida de De Finetti, a classificação, a média geral e o percentual de classificações S, para os modelos PFS e PFC. Observe que ambos os modelos obtiveram 90% de rodadas satisfatórias, sendo a 33^a rodada a única com medida de De Finetti superior a 2/3. Além disso, a média das medidas de De Finetti foram 0,64246 e 0,61055 para os modelos PFS e PFC, respectivamente.

Tabela 4.6: Medida de De Finetti para cada rodada com base em um ajuste de 19 rodadas.

Rodada	PFS		PFC	
	Média	Classificação	Média	Classificação
29	0,61550	S	0,56822	S
30	0,65797	S	0,64182	S
31	0,64962	S	0,60616	S
32	0,63323	S	0,62229	S
33	0,75829	IS	0,72510	IS
34	0,59764	S	0,59044	S
35	0,61645	S	0,57857	S
36	0,61276	S	0,54859	S
37	0,63093	S	0,62536	S
38	0,65217	S	0,59892	S
Média Geral	0,64246	90%	0,61055	90%

A [Tabela 4.7](#) apresenta o mandante, o visitante, o resultado e a distância de De Finetti, considerando os cenários abordados nessa seção, dos 10 jogos da 33^a rodada, a qual obteve classificação insatisfatória, conforme as [Tabelas 4.4](#) e [4.6](#). Tal fato pode ter sido ocasionado,

principalmente, pela vitória inesperada do time do Paraná, que era o último colocado do campeonato, sobre o América-MG e pelos empates do Atlético-MG contra o Palmeiras e do Fluminense contra o Sport Recife. Os empates referidos apontam uma distância de De Finetti maior do que 0,8. Além disso, ao menos 70% das partidas desta rodada possuem distância de De Finetti superior a 2/3.

Tabela 4.7: Distâncias de De Finetti para a 33ª rodada.

Mandante	Visitante	Resultado	19 Rodadas		28 Rodadas	
			PFS	PFC	PFS	PFC
Corinthians	São Paulo	E	0,84903	0,79724	0,71365	0,71056
América-MG	Paraná	DM	0,72438	0,81496	0,81696	0,88799
Botafogo	Flamengo	VM	0,89058	0,67418	0,85395	0,64939
Atlético-PR	Cruzeiro	VM	0,70776	0,56947	0,65880	0,53595
Vitória	Bahia	E	0,76593	0,75753	0,80548	0,78913
Grêmio	Vasco	VM	0,42105	0,39631	0,46747	0,42269
Ceará	Internacional	E	0,77008	0,72946	0,75446	0,73573
Atlético-MG	Palmeiras	E	0,87396	0,83823	0,85395	0,81214
Fluminense	Sport Recife	E	0,81579	0,80573	0,89222	0,85840
Santos	Chapecoense	DM	0,70776	0,79133	0,76594	0,84899
Média Geral			0,75263	0,71745	0,75829	0,72510

Em suma, o modelo PFC ostentou o melhor resultado nas conjunturas abordadas, apresentando superioridade em mais de 70% das distâncias de De Finetti em relação ao modelo PFS. Mesmo assim, o modelo PFS obteve uma performance satisfatória sempre maior que 55% dos casos. Desse modo, conclui-se que, conforme o esperado, as probabilidades de resultados obtidas através de ambos os modelos foram mais efetivas do que probabilidades equiprováveis.

4.5 Simulação

A representação de situações do mundo real em computadores é denominada simulação. A simulação consiste na utilização de técnicas matemáticas ou especificações formalizadas, com o propósito de estudar comportamentos e, a partir disso, antecipar algum evento específico ou realizar previsões, por exemplo. O livro *Simulation modeling and analysis* (Law e Kelton, 1991) apresenta diversos métodos e aplicações para simulações.

O princípio da simulação é a geração de números aleatórios, sendo que não existe um software capaz de criar um número totalmente aleatório, caracterizando um abuso de linguagem comumente observado. Como forma de suprir essa lacuna, faz-se necessário utilizar métodos aritméticos aptos a gerar números pseudo-aleatórios, que se aproximam de algumas propriedades dos números aleatórios. Um gerador de número pseudo-aleatório é um algoritmo que concebe uma sequência de números, a partir de um número inicial chamado de semente, os quais são aproximadamente independentes um dos outros. [Vieira](#)

[et al. \(2004\)](#) fizeram um estudo entre três geradores de números pseudo-aleatórios a fim de verificar suas eficiências computacionais.

Diversos softwares e linguagens de programação oferecem a ferramenta de geração de números aleatórios (pseudo-aleatório), o que nos exige de uma preocupação. Um exemplo é o Excel, um software pago e desenvolvido pela Microsoft, o qual é amplamente utilizado devido à sua facilidade e popularidade. [Mélard \(2014\)](#) realizou um estudo sobre a geração desses números através da função RAND (ALEATÓRIO, na versão pt-BR do Excel) e concluiu que, embora não muito satisfatório, a versão 2010 apresentou melhorias significativas em relação às versões anteriores. Em contrapartida, uma linguagem de programação bastante difundida pela comunidade acadêmica estatística é o R ([R Development Core Team, 2011](#)), que é capaz de gerar números aleatórios baseados em diversas distribuições aleatórias, contando com o recurso de escolha da semente geradora de aleatórios, permitindo, assim, a repetição de um experimento. Além da vantagem de ser um software livre, ele conta com a facilidade de instalar bibliotecas e pacotes pré-programados e testados pelos próprios usuários e desenvolvedores.

Para o propósito desse trabalho, utilizaremos o software Excel por entendermos que a sua função de geração de números aleatórios será suficiente para atingir os objetivos aqui propostos e, também, pelo fato de outras linguagens de programação, como o R, demandarem um conhecimento mais aprofundado de técnicas e lógica de programação, que não são ensinados em salas de aula, assim distanciando-se do escopo dessa pesquisa.

Com o objetivo de previsão dos resultados do campeonato, em especial, prever a probabilidade de cada time ser campeão, faremos o uso do artifício da simulação. Para isso, novamente, é necessário que partidas tenham sido disputadas para que sejam efetuados os ajustes dos modelos e, assim, prever os demais confrontos. Para obter maior precisão, as simulações devem ser realizadas inúmeras vezes, pois à medida que aumentamos o número de repetições do experimento, a média de pontos, consecutivamente a classificação dos times, varia cada vez menos até estacionarem, conforme a Lei dos Grandes Números, apresentada na [Seção 2.6](#).

4.5.1 Resultados

Para todo estudo de simulação deste trabalho, 10.000 campeonatos foram gerados, com parte das rodadas consideradas para ajuste dos modelos e as demais simuladas de acordo com as probabilidades dos resultados (Vitória, Empate e Derrota). Cada campeonato representa uma simulação, sendo que, em cada uma delas, os times apresentam uma classificação e pontuação final simulada. Assim, tomamos a média aritmética desses pontos e comparamos com a pontuação real do campeonato, de acordo com o modelo desejado. Para maiores detalhes sobre a simulação desenvolvida, consulte o [Capítulo 6](#).

Para facilitar a compreensão dos resultados obtidos nas simulações será exibida a tabela

da Série A do campeonato Brasileiro de 2018 com base em 19 rodadas, adicionando a posição final classificatória de cada time, para fins comparativos. A [Tabela 4.8](#) apresenta a posição final, posição atual, os times, os pontos ganhos, os números de jogos, vitórias, empates, derrotas, gols pró (gols realizados), gols contra (gols sofridos) e saldo de gols. Durante esses 190 jogos da competição, houve 99 vitórias dos times mandantes (52,11% do total), 55 empates (28,95% do total) e 36 vitórias dos visitantes (18,95% do total).

Tabela 4.8: Classificação da Série A do Campeonato Brasileiro de 2018 com 19 rodadas.

PF	PA	Time	PG	J	V	E	D	GP	GC	SG
1	6	Palmeiras	33	19	9	6	4	29	15	14
2	3	Flamengo	37	19	11	4	4	29	15	14
3	2	Internacional	38	19	11	5	3	27	12	15
4	4	Grêmio	36	19	10	6	3	23	8	15
5	1	São Paulo	41	19	12	5	2	32	16	16
6	5	Atlético-MG	33	19	10	3	6	36	26	10
7	14	Atlético-PR	21	19	5	6	8	21	19	2
8	8	Cruzeiro	26	19	7	5	7	15	16	-1
9	12	Botafogo	22	19	5	7	7	18	25	-7
10	15	Santos	21	19	5	6	8	21	23	-2
11	9	Bahia	25	19	6	7	6	22	23	-1
12	10	Fluminense	23	19	6	5	8	19	24	-5
13	7	Corinthians	26	19	7	5	7	22	16	6
14	16	Chapecoense	21	19	4	9	6	19	26	-7
15	19	Ceará	16	19	3	7	9	11	22	-11
16	13	Vasco	21	19	5	6	8	23	29	-6
17	17	Sport Recife	20	19	5	5	9	19	30	-11
18	11	América-MG	22	19	6	4	9	18	24	-6
19	18	Vitória	19	19	5	4	10	20	39	-19
20	20	Paraná	14	19	3	5	11	9	25	-16

Observe que, nessa rodada, o Palmeiras, campeão do torneio, ocupava a 6^a posição, enquanto o São Paulo, que terminou o campeonato na 5^a colocação, liderava a competição. Além disso, outra mudança notória está na classificação dos times Atlético-PR e América-MG que, ao final do Brasileirão, subiu e desceu 7 posições, respectivamente.

A [Tabela 4.9](#) exibe a pontuação final de cada time e a média final dos pontos das simulações realizadas para os dois modelos, considerando 19 rodadas para os ajustes. Além disso, na parte inferior da tabela pode-se visualizar a raiz quadrada dos EQM. Observe que o modelo PFS obteve um valor maior (7,02) do que o modelo PFC (7,01), em outras palavras, o modelo PFC conseguiu aproximar-se um pouco melhor das pontuações reais de cada time.

Os maiores erros previstos foram sobre a pontuação da equipe do Palmeiras (mais de 17 pontos em ambos os modelos). Sugere-se que tal fato tenha ocorrido devido à ascendência de 5 posições desse clube na segunda metade do campeonato, em que ele conquistou 47 pontos, 12 a mais que o Flamengo, time com a segunda melhor campanha nessa fase do Brasileirão, e 6 pontos a mais que o São Paulo na primeira fase do torneio.

Tabela 4.9: Média de pontos considerando 19 rodadas para os ajustes, após 10.000 simulações.

Time	Pontos	PFS	PFC
Palmeiras	80	62,53	61,55
Flamengo	72	68,69	66,49
Internacional	69	70,16	68,19
Grêmio	66	67,11	65,15
São Paulo	63	74,78	72,17
Atlético-MG	59	62,53	61,00
Atlético-PR	57	44,23	45,29
Cruzeiro	53	51,86	52,00
Botafogo	51	45,84	46,62
Santos	50	44,21	44,98
Bahia	48	50,39	50,27
Fluminense	45	47,32	47,96
Corinthians	44	51,88	51,69
Ceará	44	36,66	38,20
Chapecoense	44	44,37	45,32
Vasco	43	44,25	44,93
Sport Recife	42	42,65	43,97
América-MG	40	45,81	46,26
Vitória	37	41,20	42,55
Paraná	23	33,48	35,37
$\sqrt{\text{EQM}}$		7,02	7,01

A análise das posições relevantes, descritas na [Subseção 4.1.1](#), nos propõe uma outra visão de verificar a eficácia das simulações dos modelos. A soma das frequências em que as simulações acertaram essas classificações pode ser utilizada como métrica, para fins comparativos, e será denominada SF, sendo que quanto maior for esse o valor, melhor é a simulação do modelo.

A [Tabela 4.10](#) exhibe as frequências relativas (em %) em que cada time permaneceu em uma das posições de interesse nas simulações para os dois modelos, sendo elas: Campeão (C), Libertadores (L) e Rebaixamento (R), em que Libertadores refere-se apenas aos times classificados diretamente para a fase de grupos desse torneio. Para efeito de facilitar o estudo, não serão considerados os clubes previamente classificados para a Copa Libertadores.

Como o São Paulo, que era o melhor colocado no primeiro turno, não repetiu a boa campanha para o segundo turno, os modelos não conseguiram capturar, de forma eficiente, que o clube terminara na 5ª posição. Com isso, os modelos conferiram a essa equipe uma posição de primeiro colocado em 54,05% e 50,08% das simulações e entre os 4 primeiros classificados das simulações em 96,05% e 94,22% pelos modelos PFS e PFC, respectivamente.

Em relação aos demais resultados observados, ambos os modelos atribuíram poucas simulações com o Palmeiras sendo o campeão, porém o modelo PFC sobressaiu-se (2,77%

Tabela 4.10: Resumo da classificação de cada time (em %) com as 19 primeiras rodadas para os ajustes, após 10.000 simulações.

Time	PFS			PFC		
	C	L	R	C	L	R
Palmeiras	1,63	33,02	0	2,77	38,01	0,01
Flamengo	14,64	77,47	0	13,66	74,62	0
Internacional	19,47	84,61	0	21,34	82,31	0
Grêmio	7,62	67,22	0	8,89	64,73	0
São Paulo	54,05	96,05	0	50,08	94,22	0
Atlético-MG	2,55	36,94	0,01	3,17	37,34	0,05
Atlético-PR	0	0,03	24,92	0	0,11	24,77
Cruzeiro	0,03	1,63	2,07	0,03	2,87	2,99
Botafogo	0	0,06	16,60	0	0,26	18,07
Santos	0	0,08	25,10	0	0,09	26,60
Bahia	0	0,65	4,02	0,02	1,37	6,01
Fluminense	0	0,26	11,08	0	0,60	11,69
Corinthians	0,01	1,79	2,38	0,04	2,81	3,24
Ceará	0	0	78,80	0	0	74,34
Chapecoense	0	0,01	25,43	0	0,08	25,88
Vasco	0	0,05	25,45	0	0,14	27,32
Sport Recife	0	0	34,96	0	0,09	32,47
América-MG	0	0,11	16,23	0	0,26	18,81
Vitória	0	0,02	43,46	0	0,09	41,69
Paraná	0	0	89,49	0	0	86,06

das simulações) em comparação ao modelo PFS (1,63% das simulações). Já para a Libertadores e para o rebaixamento, o modelo PFS apresentou uma maior taxa de acerto para três das quatro equipes, se comparado ao modelo PFC. Em suma, a SF foi de 448,09 e 441,47 para os modelos PFS e PFC, respectivamente.

Agora, vamos considerar 28 rodadas para os ajustes dos modelos. Observe que restaram apenas 100 jogos a serem simulados; conseqüentemente, há uma redução na variação dos valores simulados e uma expectativa de melhora na relação entre os resultados previstos e observados. Com base nessas rodadas, será exibida a tabela da Série A do campeonato Brasileiro de 2018, adicionando, novamente, a posição final classificatória de cada time, para efeito de comparação. A [Tabela 4.11](#) apresenta a posição final, posição atual, os times, os pontos ganhos, os números de jogos, vitórias, empates, derrotas, gols pró (gols realizados), gols contra (gols sofridos) e saldo de gols. Durante essas 280 partidas ocorridas, houve 148 vitórias dos times mandantes (52,86% do total), 82 empates (29,29% do total) e 50 vitórias dos visitantes (17,86% do total).

Nesse momento do campeonato, o São Paulo já havia deixado o topo da tabela classificatória, caindo para a quarta posição, com o Palmeiras assumindo esse posto. A primeira colocação do Palmeiras contava com apenas 3 pontos de vantagem sobre o Internacional, segundo colocado, e distava em 5 pontos da quinta colocação, que não garante a vaga para fase de grupos da Copa Libertadores. Já na parte inferior da tabela, a luta contra o rebaixa-

Tabela 4.11: Classificação da Série A do Campeonato Brasileiro de 2018 com 28 rodadas.

PF	PA	Time	PG	J	V	E	D	GP	GC	SG
1	1	Palmeiras	56	28	16	8	4	43	18	25
2	3	Flamengo	52	28	15	7	6	41	22	19
3	2	Internacional	53	28	15	8	5	36	19	17
4	5	Grêmio	51	28	14	9	5	37	16	21
5	4	São Paulo	52	28	14	10	4	39	24	15
6	6	Atlético-MG	45	28	13	6	9	47	34	13
7	10	Atlético-PR	36	28	10	6	12	37	29	8
8	9	Cruzeiro	37	28	9	10	9	22	25	-3
9	12	Botafogo	34	28	8	10	10	29	39	-10
10	7	Santos	39	28	10	9	9	32	26	6
11	17	Bahia	31	28	7	10	11	28	34	-6
12	8	Fluminense	37	28	10	7	11	30	32	-2
13	11	Corinthians	35	28	9	8	11	28	26	2
14	16	Chapecoense	31	28	7	10	11	29	41	-12
15	13	Ceará	33	28	8	9	11	24	29	-5
16	15	Vasco	31	28	7	10	11	32	40	-8
17	19	Sport Recife	27	28	7	6	15	25	45	-20
18	14	América-MG	32	28	8	8	12	25	34	-9
19	18	Vitória	29	28	8	5	15	27	49	-22
20	20	Paraná	17	28	3	8	17	12	41	-29

mento estava ainda mais acirrada, pois a mesma diferença de 5 pontos era observada entre o time da décima posição e os das zona de rebaixamento. Tais fatos evidenciam a grande concorrência que o campeonato sofria em ambos os extremos da tabela de classificação.

A [Tabela 4.12](#) exibe, novamente, a pontuação final de cada time e a média final de pontos para os dois modelos, considerando 28 rodadas para os ajustes. Na parte inferior da tabela pode-se visualizar a raiz quadrada dos EQM. Note que, conforme esperado, houve uma redução nos erros cometidos pelos modelos em comparação à [Tabela 4.9](#), em que os erros médios dos pontos foram reduzidos de 7,02 para 3,87 e 7,01 para 3,99 para os modelos PFS e PFC, respectivamente.

Com base em 28 rodadas, a [Tabela 4.13](#) apresenta as frequências relativas (em %) em que cada time permaneceu em uma das posições de interesse nas simulações para os dois modelos, sendo elas: Campeão (C), Libertadores (L) e Rebaixamento (R). Novamente, com o intuito de facilitar o estudo, não serão considerados os clubes previamente classificados para a Copa Libertadores.

Apesar de a equipe do Paraná só consolidar sua condição de matematicamente rebaixado na 32ª rodada, o modelo PFS capturou esse desfecho com maestria, exibindo 100% das simulações com o time entre os 4 piores colocados. Nesse mesmo contexto, o modelo PFC explorou a ínfima chance da não conclusão desse desenredo que, mesmo assim, só fora observado em 0,04% das simulações.

Em síntese, nos resultados observados de interesse, o modelo PFS apresentou melhor desempenho na parte inferior da tabela, expondo uma maior frequência de simulações

Tabela 4.12: Média de pontos considerando 28 rodadas para os ajustes, após 10.000 simulações.

Time	Pontos	PFS	PFC
Palmeiras	80	73,08	72,27
Flamengo	72	67,89	67,16
Internacional	69	69,59	68,94
Grêmio	66	66,71	65,66
São Paulo	63	68,16	67,15
Atlético-MG	59	59,66	59,27
Atlético-PR	57	49,09	49,32
Cruzeiro	53	50,97	50,84
Botafogo	51	47,05	47,17
Santos	50	52,39	52,50
Bahia	48	44,17	44,27
Fluminense	45	50,08	50,24
Corinthians	44	47,99	47,84
Ceará	44	45,98	46,14
Chapecoense	44	43,51	43,85
Vasco	43	42,66	43,31
Sport Recife	42	38,34	39,06
América-MG	40	44,12	44,52
Vitória	37	41,02	41,50
Paraná	23	26,35	27,81
$\sqrt{\text{EQM}}$		3,87	3,99

para os quatro times rebaixados, obtendo um SF de 630,82. Embora o PFS tenha obtido maior êxito com qual equipe seria campeã, o modelo PFC mostrou-se mais eficiente, de maneira geral, para simular a parte superior, pois exibiu maior frequência para três das quatro equipes classificadas para a fase de grupo da Copa Libertadores, apresentando SF de 624,88.

Tabela 4.13: Resumos da classificação de cada time (em %) com as 28 primeiras rodadas para os ajustes, após 10.000 simulações.

Time	PFS			PFC		
	C	L	R	C	L	R
Palmeiras	54,44	96,54	0	53,79	96,63	0
Flamengo	10,92	74,88	0	11,86	75,13	0
Internacional	18,64	85,67	0	19,28	86,79	0
Grêmio	5,96	60,90	0	5,50	58,78	0
São Paulo	9,89	72,76	0	9,99	73,11	0
Atlético-MG	0,15	9,04	0	0,14	10,16	0
Atlético-PR	0	0,02	3,14	0	0,03	3,70
Cruzeiro	0	0,03	1,11	0	0,06	1,61
Botafogo	0	0	9,72	0	0	10,41
Santos	0	0,12	0,31	0	0,23	0,40
Bahia	0	0	28,56	0	0	30,05
Fluminense	0	0,04	1,75	0	0,09	1,87
Corinthians	0	0	5,99	0	0	6,98
Ceará	0	0	13,86	0	0	15,85
Chapecoense	0	0	34,64	0	0	34,70
Vasco	0	0	42,53	0	0	40,77
Sport Recife	0	0	77,53	0	0	75,15
América-MG	0	0	26,49	0	0	25,80
Vitória	0	0	54,37	0	0	52,85
Paraná	0	0	100	0	0	99,96

Capítulo 5

Propostas de Atividades em Sala de Aula

Neste capítulo são propostas atividades que podem ser realizadas em sala de aula, com embasamento no conteúdo apresentado neste trabalho, cujos objetivos são desenvolver o senso crítico dos alunos através de discussões, desenvolver o raciocínio lógico com a construção de modelos para o futebol e incitar a percepção da importância da Matemática em nosso cotidiano, além de apresentar, na prática, a Lei dos Grandes Números e introduzir a ideia de simulação.

5.1 Atividade 1 - Modelagem no Futebol

A motivação para a primeira atividade é desenvolver o senso crítico do alunado por meio de debates e contribuir, de forma significativa, para a percepção da importância da Matemática no cotidiano, abordando os conceitos de modelagem de fenômenos, em especial o futebol, através dos modelos apresentados neste trabalho.

O professor deverá explicar a diferença entre modelos determinísticos e probabilísticos com o uso de exemplos, a fim de facilitar o entendimento dos alunos. Além disso, o docente deverá comentar maneiras probabilísticas de fazer modelagem para uma partida de futebol considerando a força dos times ou a chance das equipes marcarem gols, por exemplo.

O próximo passo será formar grupos com até 5 alunos, com intuito de fomentar um debate sobre maneiras de realizar modelagens para o futebol. O professor apresentará o modelo PFS, discutindo seus parâmetros e dinâmicas, propondo um aprimoramento do mesmo a fim de obter melhor eficiência nas previsões. Por fim, deve-se abordar a motivação do efeito casa para o desenvolvimento e construção do modelo PFC junto à turma.

5.2 Atividade 2 - Qualidade Preditiva

O principal objetivo da atividade 2 é mostrar que, como a modelagem pode ser realizada por diversos caminhos, existem testes para avaliar suas qualidades, a fim de verificar o modelo mais adequado e proporcionar uma maneira de inferir sobre os resultados obtidos. Adicionalmente, ela tem o objetivo de esclarecer a diferença entre “probabilidade de ocorrer” e “ocorrência observada”.

A Série A do Campeonato Brasileiro de Futebol ocorre durante o período letivo, podendo ser referência de estudo na sala de aula. No decorrer do torneio os alunos deverão ajustar os parâmetros dos modelos para a rodada atual e calcular as probabilidades de cada resultado (vitória do mandante, empate e derrota do mandante) para os 10 jogos da rodada seguinte.

O professor deverá apresentar a distância e a medida de De Finetti, de forma mais clara e sucinta possível, como uma maneira de averiguar a capacidade preditiva de um modelo comparado com outros modelos ou com probabilidades equiprováveis. Após a ocorrência das partidas, devem-se calcular as distâncias apresentadas para os jogos dessa rodada e comparar com as probabilidades equiprováveis.

Neste momento, o professor poderá questionar se a turma considera vantajosa a utilização do modelo PFS e se observaram um ganho preditivo na construção do modelo PFC, com base nos resultados observados. Além disso, o docente deve aproveitar a oportunidade para esclarecer a diferença entre “probabilidade de ocorrer”, obtida pelos modelos, e “ocorrência observada”, dado após a rodada acontecer, em que não necessariamente a maior probabilidade de um modelo será o resultado constatado, podendo fazer uma alusão com previsões do tempo, em que pode não haver a ocorrência de chuvas, mesmo com probabilidade alta de precipitação.

Para efeito de comparação, a atividade pode ser reproduzida considerando quantidade distinta de dados para os ajustes na previsão da mesma rodada. Como por exemplo, em um primeiro momento, 10 rodadas podem ser consideradas para os ajustes dos modelos, a fim de obter a probabilidade dos resultados da 12ª rodada. Agora, utiliza-se 11 rodadas para obter, novamente, a probabilidade dos resultados da 12ª rodada. Após a realização dos cálculos e ocorrência da rodada prevista, o docente deve verificar junto à turma qual método foi mais eficaz na previsão, através da medida de De Finetti. Vale a explicação de que, a cada observação do campeonato, os parâmetros dos modelos devem ser atualizados para uma melhor previsão.

5.3 Atividade 3 - Aplicação da Lei dos Grandes Números

A motivação da atividade 3 é apresentar, na prática, a Lei dos Grandes Números, proporcionando uma visualização clara, além de introduzir a simulação através do experimento de lançamento de um dado e de uma partida de futebol, com o uso do Software Excel, para uma quantidade de repetições significativamente grande.

A atividade proposta será dividida em três partes, sendo que a primeira é o experimento de lançamento de uma moeda, a qual não necessita do uso de computadores, a segunda consiste em simular, no Software Excel, o lançamento de um dado e a terceira é a simulação de uma partida de futebol, com as probabilidades dos resultados predefinidas, exibindo os pontos médios obtidos pelos dois times.

5.3.1 Lançamento de uma moeda

Para a realização dessa etapa, os estudantes deverão jogar uma moeda 20 vezes e anotar cada resultado (Cara ou Coroa). Depois deve-se calcular a frequência relativa (razão entre frequência absoluta e número de observações).

O docente deverá estender o experimento, considerando as repetições de cada aluno, obtendo, assim, $20 \times n$ repetições, em que n é o número de alunos. Desse modo, deve-se calcular a frequência relativa e observar que esse valor aproxima-se de $1/2$, como esperado pela Lei dos Grandes Números.

5.3.2 Simulação do lançamento de um dado

Essa parte da atividade consiste em simular o lançamento de um dado de 6 faces, utilizando um computador com o Software Excel. Serão simulados 100.000 lançamentos e, em cada um deles, a face observada será registrada.

A seguir são exibidos os comandos necessários para a simulação:

- Célula “A1” recebe “=LIN()” (“=ROW()”, versão em inglês);
- Célula “B1” recebe “=ALEATÓRIOENTRE(1,6)” (“=RANDBETWEEN(1;6)”, versão em inglês);
- Arraste a fórmula de “A1:B1” até “A100000:B100000”;
- Células “I1”, “J1”, “K1”, “L1”, “M1” e “N1” recebem “1”, “10”, “100”, “1000”, “10000” e “100000”, respectivamente;
- Células “H2”, “H3”, “H4”, “H5”, “H6” e “H7” recebem “1”, “2”, “3”, “4”, “5” e “6”, respectivamente;

- Célula “I2” recebe “=CONT.SES(\$B:\$B, \$H2, \$A:\$A, “<=”&I\$1)/I\$1*100” (“=COUNTIFS(\$B:\$B; \$H2; \$A:\$A; “<=”&I\$1)/I\$1*100”, versão em inglês);
- Arraste a fórmula de “I2” até “N2” e então de “I2:N2” até “I7:N7”.

O professor deverá passar as instruções supracitadas aos alunos. A título de ilustrar essa atividade é apresentada uma simulação com os resultados obtidos. A [Tabela 5.1](#) exibe as faces do dado, as repetições, as frequências relativas (em %) e a raiz quadrada do EQM, para uma simulação do lançamento de um dado.

Tabela 5.1: Frequência relativa (em %) observada por quantidade de repetição no lançamento de um dado, via simulação.

Faces	Repetições					
	1	10	100	1.000	10.000	100.000
1	0	10	10	15,6	16,34	16,761
2	100	40	20	16,5	16,49	16,566
3	0	0	15	17,8	16,56	16,444
4	0	20	14	15,1	16,29	16,991
5	0	20	24	17,5	17,36	16,760
6	0	10	17	17,5	16,96	16,478
$\sqrt{\text{EQM}}$	0,913	0,306	0,109	0,025	0,009	0,005

Nota-se que, conforme a Lei dos Grandes Números, à medida que a quantidade de repetições aumenta as frequências aproximam-se de 1/6. Vale ressaltar que a cada simulação novos resultados serão gerados, entretanto as frequências relativas sempre apresentarão o mesmo comportamento.

5.3.3 Simulação de uma partida de futebol

Essa parte da atividade consiste em simular uma partida de futebol com probabilidades fixadas, utilizando um computador com o Software Excel, e observar a média de pontos obtidos pelo time mandante e visitante. Serão simuladas 100.000 partidas entre as mesmas equipes e, em cada um deles, o resultado será registrado, sendo eles: vitória do mandante (VM), empate (E) e derrota do mandante (DM).

A seguir são exibidos os comandos necessários para a simulação:

- Célula “A1” recebe “=LIN()” (“=ROW()”, versão em inglês);
- Célula “B1” recebe “=ALEATÓRIO()” (“=RAND()”, versão em inglês);
- Célula “C1” recebe “=SE(\$B1<=\$F\$1, "VM", SE(\$B1<=(\$F\$1+\$F\$2), "E", "DM"))” (“=IF(\$B1<=\$F\$1; "VM"; IF(\$B1<=(\$F\$1+\$F\$2); "E"; "DM"))”, versão em inglês);

- Arraste a fórmula de “A1:C1” até “A100000:C100000”;
- Células “E1”, “E2” e “E3” recebem “VM”, “E” e “DM”, respectivamente;
- Células “F1”, “F2” e “F3” recebem “0,50”, “0,30” e “=1-F1-F2”, respectivamente;
- Células “I1”, “J1”, “K1”, “L1”, “M1” e “N1” recebem “1”, “10”, “100”, “1000”, “10000” e “100000”, respectivamente;
- Células “H2”, “H3”, “H4”, “H5”, “H6” recebem “VM”, “E”, “DM”, “PM”, “PV”, respectivamente;
- Célula “I2” recebe “=CONT.SES(\$C:\$C, \$H2, \$A:\$A, “<=”&I\$1)/I\$1*100” (“=COUNTIFS(\$B:\$B; \$H2; \$A:\$A; “<=”&I\$1)/I\$1*100”, versão em inglês);
- Arraste a fórmula de “I2” até “I4”;
- Célula “I5” recebe “=(I2*3+I3)/100”;
- Célula “I6” recebe “=(I4*3+I3)/100”;
- Arraste a fórmula de “I2:I6” até “N2:N6”.

O professor deverá passar as instruções supracitadas para os alunos. Para exemplificar essa atividade é apresentada uma simulação para uma partida em que as probabilidades de vitória do mandante (VM), empate (E) e derrota do mandante (DM) é de 50%, 30% e 20%, respectivamente. A [Tabela 5.2](#) exibe resultados da partida, as repetições, as frequências relativas (em %) e a raiz quadrada do EQM entre as frequências observadas e as probabilidades previamente determinadas, além de mostrar os pontos médios obtidos para o mandante (PM) e para o visitante (PV). Observa-se que, à medida que o número de repetições aumenta a raiz quadrada do EQM diminui.

Tabela 5.2: Frequência relativa (em %) observada por quantidade de repetição em uma partida de futebol e pontos médios obtidos, via simulação.

Resultados	Repetições					
	1	10	100	1.000	10.000	100.000
VM	100	70	52	47,9	49,14	50,119
E	0	20	31	30,6	29,83	29,852
DM	0	10	17	21,5	21,03	20,029
$\sqrt{\text{EQM}}$	0,616	0,245	0,037	0,026	0,014	0,002
PM	3	2,300	1,870	1,743	1,773	1,802
PV	0	0,700	1,100	1,133	1,105	1,096

Com essa simulação é possível obter os pontos médios para dois times de uma determinada partida. O professor deve aproveitar esse fato para comentar o princípio da

simulação de um campeonato, que seria adquirido pela repetição desse experimento para todas os jogos ainda não realizados. Para finalizar, o docente deve solicitar para que os alunos alterem as probabilidades pré-fixadas, podendo ser as probabilidades calculadas pelos modelos PFS ou PFC, a fim de fomentar discussões acerca dos novos resultados.

5.4 Atividade 4 - Simulação de um campeonato

A simulação completa de um campeonato de futebol exige conhecimentos mais aprofundados em determinados conteúdos, como lógica de programação, por exemplo, o que constitui uma limitação para a sua realização em sala de aula. Para o professor interessado em desenvolver esse exercício com o alunado é recomendada a construção da simulação pelo docente e distribuição para que os alunos apenas observem as dinâmicas envolvendo esse conteúdo. O [Capítulo 6](#) traz o passo a passo para a elaboração de uma simulação pelo modelo PFS.

Capítulo 6

Considerações Finais

Neste trabalho foram abordados conceitos básicos de modelagem matemática e Probabilidade, foi apresentado o modelo PFS (Lima et al., 2012) e construído o modelo PFC, que foram aplicados aos dados da Série A do Campeonato Brasileiro de 2018. Os critérios de seleção de modelo AIC (Akaike, 1974) e BIC (Schwarz et al., 1978) foram introduzidos e o erro quadrático médio e a medida de De Finetti (De Finetti, 1972) foram utilizadas como métricas para comparar os modelos do ponto de vista preditivo, além de ter sido realizado um estudo de simulação para o torneio em questão, considerando algumas rodadas ocorridas. O conteúdo discorrido neste estudo foi transcrito de uma maneira plausível para ser aplicado para alunos do Ensino Médio.

O desempenho de uma equipe em uma partida de futebol é influenciado por diversos aspectos, sendo o mando de campo um fator relevante, conforme Pollard (1986). O modelo PFS, que não considera esse fato empírico, foi aprimorado por meio da construção do modelo PFC, o qual é original. A fim de verificar se esse modelo apresentou ganho preditivo em relação ao seu precursor, calculou-se a distância de De Finetti para ambos, constatando uma vantagem do modelo PFC em todos os cenários considerados. Para o estudo de simulação, o EQM, entre os pontos médios obtidos pelos times através da simulação do campeonato e os pontos observados, considerando 19 rodadas, foi próximo para ambos os modelos, e com 28 rodadas o modelo PFS sobressaiu-se. Do ponto de vista das posições relevantes, o modelo PFS obteve maior SF nas situações abordadas, entretanto o modelo PFC mostrou-se eficiente em simular determinadas classificações de interesse. Por fim, conclui-se que abordar temas do cotidiano como o futebol, por exemplo, pode ser uma maneira de instigar o interesse dos alunos pela Matemática e as aplicações apresentadas neste estudo podem contribuir como material de apoio aos professores.

A habilidade atual de um time está mais intimamente atrelado à sua atuação recente do que a partidas mais antigas. Dessa forma, propomos como trabalho futuro maneiras de implementar pesos para as partidas mais recentes. A abordagem da inferência estatística com aplicação no futebol também pode ser utilizada, como proposta futura, para o incentivo

do estudo da Teoria Estatística, realizando análises descritivas sobre vários campeonatos e verificando suposições através de testes de hipóteses.

Referências Bibliográficas

- Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723.
- Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. Em *Selected papers of hirotugu akaike*, pgs. 199–213. Springer.
- Bassanezi, R. C. (2009). *Ensino-aprendizagem com modelagem matemática: uma nova estratégia*. Editora Contexto, 3rd edition.
- Boaler, J. (2015). *What's math got to do with it?: How teachers and parents can transform mathematics learning and inspire success*. Penguin.
- Brillinger, D. R. (2008). Modelling game outcomes of the brazilian 2006 series a championship as ordinal-valued. *Brazilian Journal of Probability and Statistics*, 22(2):89–104.
- Clarke, S. R. e Norman, J. M. (1995). Home ground advantage of individual clubs in english soccer. *The Statistician*, pgs. 509–521.
- Crowder, M., Dixon, M., Ledford, A., e Robinson, M. (2002). Dynamic modelling and prediction of english football league matches for betting. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 51(2):157–168.
- De Finetti, B. (1972). Probability, induction, and statistics. *John Wiley: London*.
- Dixon, M. J. e Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(2):265–280.
- Gutstein, E. (2003). Teaching and learning mathematics for social justice in an urban, latino school. *Journal for Research in Mathematics education*, pgs. 37–73.
- James, B. R. (1996). *Probabilidade: um curso em nível intermediário*. Number 519.2.
- Karlis, D. e Ntzoufras, I. (2003). Analysis of sports data by using bivariate poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3):381–393.

- Karlis, D. e Ntzoufras, I. (2009). Bayesian modelling of football outcomes: using the skellam's distribution for the goal difference. *IMA Journal of Management Mathematics*, 20(2):133–145.
- Kocherlakota, S. e Kocherlakota, K. (1992). *Bivariate discrete distributions*. Wiley Online Library.
- Koopman, S. J. e Lit, R. (2013). A dynamic bivariate poisson model for analysing and forecasting match results in the english premier league. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.
- Law, A. M. e Kelton, W. D. (1991). *Simulation modeling and analysis*, volume 2. McGraw-Hill New York.
- Lee, A. J. (1997). Modeling scores in the premier league: is manchester united really the best? *Chance*, 10:15–19.
- Lima, B. N. B. d., Costa, G. N., Nacife, R., Vidal Martins, R., e Guimarães, R. (2012). Probabilidades no esporte. pgs. 39–53.
- Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, 36(3):109–118.
- Mélard, G. (2014). On the accuracy of statistical procedures in microsoft excel 2010. *Computational statistics*, 29(5):1095–1128.
- Meyer, P. L. (1970). Probabilidade: aplicações à estatística. Em *Probabilidade: aplicações à estatística*. Livro Técnico.
- Moroney, M. J. (1956). Facts from figures. Technical report.
- Nevill, A. M., Balmer, N. J., e Williams, A. M. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise*, 3(4):261–272.
- Owen, A. (2011). Dynamic bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA Journal of Management Mathematics*, 22(2):99–113.
- Pollard, R. (1986). Home advantage in soccer: A retrospective analysis. *Journal of Sports Sciences*, 4(3):237–248.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- Rosen, L. (2001). Myth making and moral order in a debate on mathematics education policy. *Policy as practice: Toward a comparative sociocultural analysis of educational policy*, 1:295–316.
- Rue, H. e Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(3):399–418.
- Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.
- Steen, L. A. (1997). *Why numbers count: Quantitative literacy for tomorrow's America*. New York: College Entrance Examination Board.
- Vieira, C. E. C., e Souza, R. d. C., e Ribeiro, C. C. (2004). *Um estudo comparativo entre três geradores de números aleatórios*. PUC.

Apêndice A

Simulação para um Campeonato de Pontos Corridos

Neste apêndice será descrito a elaboração da simulação do modelo PFS para um campeonato de pontos corridos, considerando 20 equipes, com o total de 380 jogos, com base em 190 partidas para o ajuste e com 10.000 campeonatos gerados. Para gerar a simulação utilizando o *software* Excel, vamos segregar o arquivo em 6 planilhas, sendo que, para o correto funcionamento dos comandos, os nomes devem ser os mesmos aqui intitulados. Cada planilha será descrita no formato de tópico, recebendo seus respectivos detalhes.

Vale ressaltar, que essa simulação pode servir de base para outros modelos ou ser modificada para distintas quantidades de partidas para o ajuste.

Ajuste

Nesta planilha serão adicionados os resultados das partidas para o ajuste, exibido a tabela de classificação e calculado o vetor força de cada equipe.

- Células “A1”, “B1”, “C1”, “D1”, “E1”, “F1”, “G1”, recebem “Jogo”, “Mandante”, “Visitante”, “Gols Mandante”, “Gols Visitante”, “PM”, “PV”, respectivamente;
- Preencher as colunas *A*, *B*, *C*, *D* e *E* com os respectivos resultados dos 190 jogos;
- Célula “F2” recebe “=SE(D2-E2>0, 3, IF(D2-E2=0, 1, 0))” (“=IF(D2-E2>0; 3; IF(D2-E2=0; 1; 0))”, versão em inglês);
- Célula “G2” recebe “=SE(D2-E2<0, 3, IF(D2-E2=0, 1, 0))” (“=IF(D2-E2<0; 3; IF(D2-E2=0; 1; 0))”, versão em inglês);
- Arraste a fórmula de “F2:G2” até “F191:G191”;

- Células “J1”, “K1”, “L1”, “M1”, “N1”, “O1”, recebem “Time”, “PT”, “V”, “E”, “D”, “Jogos”, respectivamente;
- Preencher as células “J2:J21” com os 20 times do campeonato;
- Célula “K2” recebe “=SOMASES(\$F:\$F, \$B:\$B, \$J2) + SOMASES(\$G:\$G, \$C:\$C, \$J2)” (“=SUMIFS(\$F:\$F; \$B:\$B; \$J2) + SUMIFS(\$G:\$G; \$C:\$C; \$J2)”, versão em inglês);
- Célula “L2” recebe “=CONT.SES(\$F:\$F, 3, \$B:\$B, \$J2) + CONT.SES(\$G:\$G, 3, \$C:\$C, \$J2)” (“=COUNTIFS(\$F:\$F; 3; \$B:\$B; \$J2) + COUNTIFS(\$G:\$G; 3; \$C:\$C; \$J2)”, versão em inglês);
- Célula “M2” recebe “=CONT.SES(\$F:\$F, 1, \$B:\$B, \$J2) + CONT.SES(\$G:\$G, 1, \$C:\$C, \$J2)” (“=COUNTIFS(\$F:\$F; 1; \$B:\$B; \$J2) + COUNTIFS(\$G:\$G; 1; \$C:\$C; \$J2)”, versão em inglês);
- Célula “N2” recebe “=CONT.SES(\$F:\$F, 0, \$B:\$B, \$J2) + CONT.SES(\$G:\$G, 0, \$C:\$C, \$J2)” (“=COUNTIFS(\$F:\$F; 0; \$B:\$B; \$J2) + COUNTIFS(\$G:\$G; 0; \$C:\$C; \$J2)”, versão em inglês);
- Célula “O2” recebe “=SOMA(L2:N2)” (“=SUM(L2:N2)”, versão em inglês);
- Arraste a fórmula de “K2:O2” até “K21:O21”;
- Células “Q1”, “R1”, “S1”, recebem “Va”, “Ea”, “Da”, respectivamente;
- Célula “Q2” recebe “=L2/\$O2”;
- Arraste a fórmula de “Q2” até “S2” e, então, de “Q2:S2” até “Q21:S21”.

Números Aleatórios

Para evitar problemas com mais de um número aleatório para gerar o resultado de uma partida, o intuito dessa planilha é garantir que cada disputa simulada será gerada a partir de apenas um número aleatório.

- Célula “B7” recebe “=ALEATÓRIO()” (“=RAND()”, versão em inglês);
- Arraste a fórmula de “B7” até “GI7” e, então, de “B7:GI7” até “B10006:GI10006”.

Simulação

Esta planilha deverá conter todas as partidas a serem simuladas e, com base nos números aleatórios gerados anteriormente e nos vetores força de cada equipe, será gerado um resultado para todas as partidas em cada uma das 10.000 simulações.

- Células “A1”, “A2”, “A3”, “A4”, “A5”, “A6” recebem “Jogo”, “Mandante”, “Visitante”, “PV”, “PE”, “PD”, respectivamente;
- Células “B1:G11” recebem o número dos jogos a serem previstos (de 191 a 380);
- Preencher as linhas 2 e 3 com os respectivos mandantes e visitantes dos 190 jogos remanescentes;
- Célula “B4” recebe “=(ÍNDICE(Ajuste!\$Q\$2:\$Q\$21, CORRESP(B\$2, Ajuste! \$J\$2:\$J\$21, 0)) + ÍNDICE(Ajuste! \$S\$2:\$S\$21, CORRESP(B\$3, Ajuste! \$J\$2:\$J\$21, 0)))/2” (“=(INDEX(Ajuste! \$Q\$2:\$Q\$21; MATCH(B\$2; \$J\$2:\$J\$21; 0)) + INDEX(Ajuste!\$S\$2:\$S\$21; MATCH(B\$3; Ajuste!\$J\$2:\$J\$21; 0)))/2”, versão em inglês);
- Célula “B5” recebe “=(ÍNDICE(Ajuste!\$R\$2:\$R\$21, CORRESP(B\$2, Ajuste! \$J\$2:\$J\$21, 0)) + ÍNDICE(Ajuste! \$R\$2:\$R\$21, CORRESP(B\$3, Ajuste! \$J\$2:\$J\$21, 0)))/2” (“=(INDEX(Ajuste!\$R\$2:\$R\$21; MATCH(B\$2; \$J\$2:\$J\$21; 0)) + INDEX(Ajuste!\$R\$2:\$R\$21; MATCH(B\$3; Ajuste!\$J\$2:\$J\$21; 0)))/2”, versão em inglês);
- Célula “B6” recebe “=(ÍNDICE(Ajuste!\$S\$2:\$S\$21, CORRESP(B\$2, Ajuste! \$J\$2:\$J\$21, 0)) + ÍNDICE(Ajuste! \$Q\$2:\$Q\$21, CORRESP(B\$3, Ajuste! \$J\$2:\$J\$21, 0)))/2” (“=(INDEX(Ajuste!\$S\$2:\$S\$21; MATCH(B\$2; \$J\$2:\$J\$21; 0)) + INDEX(Ajuste!\$Q\$2:\$Q\$21; MATCH(B\$3; Ajuste!\$J\$2:\$J\$21; 0)))/2”, versão em inglês);
- Célula “B7” recebe “=SE('Números Aleatórios'!B7 <= B\$4, "VM", SE('Números Aleatórios'!B7 <= B\$4+B\$5, "E", "DM"))” (“=IF('Números Aleatórios'!B7 <= B\$4; "VM"; IF('Números Aleatórios'!B7 <= B\$4+B\$5; "E"; "DM"))”, versão em inglês);
- Arraste a fórmula de “B7” até “GI7” e, então, de “B7:GI7” até “B10006:GI10006”.

Pontos por Campeonato

Para contornar o empecilho de mais de uma equipe possuir exatamente a mesma quantidade de pontos, impossibilitando verificar qual a sua classificação, podemos contabilizar suas vitórias e realizar um “sorteio” como critério de desempate, conforme as regras do campeonato. Desse modo, essa planilha contabiliza a pontuação de cada equipe nos campeonatos simulados, adicionando um peso baseado nos critérios citados.

- Células “A1” recebe “Time”;
- Preencher as células “B1:U1” com os 20 times do campeonato;
- Células “A2:A10001” recebem "Campeonato 1"até "Campeonato 10000";
- Célula “B2” recebe “=(CONT.SES(Simulação!7:7, "VM", Simulação!\$2:\$2, B\$1) + CONT.SES(Simulação!7:7, "DM", Simulação!\$3:\$3, B\$1)) * (3+10^-9) + CONT.SES(Simulação!7:7, "E", Simulação!\$2:\$2, B\$1) + CONT.SES(Simulação!7:7, "E", Simulação!\$3:\$3, B\$1) + ÍNDICE(Ajuste!\$K\$2:\$K\$21, CORRESP(B\$1, Ajuste!\$J\$2:\$J\$21, 0)) + (ÍNDICE(Ajuste!\$L\$2:\$L\$21, CORRESP(B\$1, Ajuste!\$J\$2:\$J\$21, 0)) + ALEATÓRIO()) * 10^-9” (“=(COUNTIFS(Simulação!7:7; "VM"; Simulação!\$2:\$2; B\$1) + COUNTIFS(Simulação!7:7; "DM"; Simulação!\$3:\$3; B\$1)) * (3+10^-9) + COUNTIFS(Simulação!7:7; "E"; Simulação!\$2:\$2; B\$1) + COUNTIFS(Simulação!7:7; "E"; Simulação!\$3:\$3; B\$1) + INDEX(Ajuste!\$K\$2:\$K\$21; MATCH(B\$1; Ajuste!\$J\$2:\$J\$21; 0)) + (INDEX(Ajuste!\$L\$2:\$L\$21; MATCH(B\$1; Ajuste!\$J\$2:\$J\$21; 0)) + RAND()) * 10^-9”, versão em inglês);
- Arraste a fórmula de “B2” até “U2” e, então, de “B2:U2” até “B10001:U10001”.

Posições de Destaque

A ordem dos times inserido na planilha “Pontos por Campeonato” deve se manter a mesma. Com base na pontuação gerada nas simulações, é possível verificar se cada time ficou em alguma posição relevante.

- Células “A1” recebe “Time”;
- Preencher as células “B1:U1” com os 20 times do campeonato;
- Células “A2:A10001” recebem "Campeonato 1"até "Campeonato 10000";
- Célula “B2” recebe “=SE('Pontos Por Campeonato'!B2 = MÁXIMO('Pontos Por Campeonato'!\$B2:\$U2), "Campeão", SE('Pontos Por Campeonato'!B2 >= MAIOR('Pontos Por Campeonato'!\$B2:\$U2, 4), "Libertadores", SE('Pontos Por Campeonato'!B2 <= MENOR('Pontos Por Campeonato'!\$B2:\$U2, 4), "Rebaixado", "-"))” (“=IF('Pontos Por Campeonato'!B2 = MAX('Pontos Por Campeonato'!\$B2:\$U2); "Campeão"; IF('Pontos Por Campeonato'!B2 >= LARGE('Pontos Por Campeonato'!\$B2:\$U2, 4); "Libertadores"; IF('Pontos Por Campeonato'!B2 <= SMALL('Pontos Por Campeonato'!\$B2:\$U2, 4); "Rebaixado"; "-"))”, versão em inglês);
- Arraste a fórmula de “B2” até “U2” e, então, de “B2:U2” até “B10001:U10001”.

Tabela Final

Novamente, a ordem dos times deve se manter a mesma que o inserido anteriormente. Nesta planilha é possível visualizar o resultado final obtido pela simulação. Para gerar novos resultados, basta atualizar (botão “F9”) o arquivo.

- Células “A1”, “A2”, “A3”, “A4”, “A5”, recebem “Time”, “Média de Pontos”, “Campeão”, “Libertadores”, “Rebaixado”, respectivamente;
- Preencher as células “B1:U1” com os 20 times do campeonato;
- Célula “B2” recebe “=MÉDIA('Pontos Por Campeonato'!B:B)” (“=AVERAGE('Pontos Por Campeonato'!B:B)”, versão em inglês);
- Célula “B3” recebe “=CONT.SES('Posições de Destaque'!B:B, "Campeão") / 10000” (“=COUNTIFS('Posições de Destaque'!B:B; "Campeão") / 10000”, versão em inglês);
- Célula “B4” recebe “=CONT.SES('Posições de Destaque'!B:B, "Libertadores") / 10000 + B3” (“=COUNTIFS('Posições de Destaque'!B:B; "Libertadores") / 10000 + B3”, versão em inglês);
- Célula “B5” recebe “=CONT.SES('Posições de Destaque'!B:B, "Rebaixado") / 10000” (“=COUNTIFS('Posições de Destaque'!B:B; "Rebaixado") / 10000”, versão em inglês);
- Arraste a fórmula de “B2:B5” até “U2:U5”.