



Universidade Estadual de Campinas  
Faculdade de Ciências Aplicadas  
Mestrado em Engenharia de Produção  
e de Manufatura



Iago Cambre Añon

Aplicações de Pesquisa Operacional no Futebol

Applications of Operations Research in Football

LIMEIRA-SP

2017

Iago Cambre Añon

Aplicações de Pesquisa Operacional no Futebol

Applications of Operations Research in Football

Dissertação apresentada à Faculdade de Ciências Aplicadas da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia de Produção e de Manufatura, na área de Pesquisa Operacional e Gestão de Processos.

Dissertation presented to the School of Applied Sciences of the University of Campinas in partial fulfillment of the requirements for the degree of Master of Science in Management in the area of Production and Manufacturing Engineering, in the area of Operational Research and Processes Management.

Este exemplar corresponde à versão final da dissertação defendida pelo aluno Iago Cambre Añon, e orientado pelo Prof. Dr. Cristiano Torezzan,

---

*Prof. Dr. Cristiano Torezzan (orientador)*

LIMEIRA-SP

2017



**Agência(s) de fomento e nº(s) de processo(s):** Não se aplica.

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca da Faculdade de Ciências Aplicadas  
Renata Eleuterio da Silva - CRB 8/9281

An78a Añon, Iago Cambre, 1992-  
Aplicações de pesquisa operacional no futebol / Iago Cambre Añon. –  
Limeira, SP : [s.n.], 2017.

Orientador: Cristiano Torezzan.  
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade  
de Ciências Aplicadas.

1. Markov, Cadeias de. 2. Probabilidades. 3. Programação linear. 4. Análise  
envoltória de dados. 5. Pesquisa operacional. 6. Futebol. I. Torezzan,  
Cristiano, 1976-. II. Universidade Estadual de Campinas. Faculdade de  
Ciências Aplicadas. III. Título.

#### Informações para Biblioteca Digital

**Título em outro idioma:** Applications of operations research in football

**Palavras-chave em inglês:**

Markov processes

Probabilities

Linear programming

Data envelopment analysis

Operational research

Soccer

**Área de concentração:** Pesquisa Operacional e Gestão de Processos

**Titulação:** Mestre em Engenharia de Produção e de Manufatura

**Banca examinadora:**

Cristiano Torezzan [Orientador]

Leonardo Tomazeli Duarte

Paulo Roberto Pereira Santiago

**Data de defesa:** 19-05-2017

**Programa de Pós-Graduação:** Engenharia de Produção e de Manufatura

## Folha de aprovação

Autor(a): Iago Cambre Añon

Título: Aplicações de Pesquisa Operacional no Futebol

Natureza: Dissertação

Área de concentração: Engenharia de Produção e de Manufatura /Pesquisa Operacional

Instituição: Faculdade de Ciências Aplicadas – FCA/Unicamp

Data da Defesa: Limeira-SP, 19 de maio de 2017.

### BANCA EXAMINADORA:

-----

Prof. Dr. Cristiano Torezzan (Orientador)

Presidente da Comissão

-----

Prof. Dr. Leonardo Tomazeli Duarte

-----

Prof. Dr. Paulo Roberto Pereira Santiago

A Ata da Defesa com as respectivas assinaturas dos membros da banca encontra-se no processo de vida acadêmica do aluno.

*“Não deixe o ruído das opiniões dos  
outros abafar a sua própria voz interior”*

*Steve Jobs.*

# Agradecimentos

Este momento e trabalho estão ligados a mais uma fase de transição da minha vida. Nesta etapa encerro um ciclo, um ciclo de aprendizagem e de dor, de responsabilidades e de grande imaturidade de minha parte. Um ciclo que contempla grandes conquistas, mas que traz nas derrotas as maiores lições. O que farei a partir desse instante em que acabo minha participação no mestrado ainda não está claro, mas a certeza de que saio mais preparado para encarar qualquer um dos meus sonhos de frente é real. Para finalizar de forma justa esta etapa, passo a agradecer as pessoas que marcaram esta fase.

Os primeiros são meus pais, que em todos os momentos me apoiaram e estiveram presentes, que abriram mão de seus desejos para satisfazer os meus. A vocês deixo aqui que os amo e que peço desculpas por não ter estado muito presente nestes últimos anos e provavelmente ao longo da carreira, mas sei que vocês compreenderam.

Em segundo lugar a minha avó, Dona Maria que, sem dúvida é um dos maiores exemplos de ser humano que tive em minha vida. Com ela aprendi a cumprimentar todas as pessoas, a oferecer ajuda e a tomar café. Fico feliz por até se lembrar de mim e por poder ter tido tempo de dizer o quanto você foi especial.

Aos meus amigos que me apoiaram em todos estes momentos, me aturando e me provocando em momento em que precisei crescer profissional e emocionalmente. Dentre eles, nesta fase, posso destacar alguns, sendo eles: Takahashi, Yamanaka, Rovariz, Caio, Vitor, João Gabriel, Modesto, a Fernanda e o Júlio, dentre outros. Novamente obrigado pelas conversas, reflexões e conselhos.

Durante toda esta trajetória, desde que cheguei na FCA em 2010, sempre tive a sorte de ter por perto bons professores. Agradeço a todos eles. Porém alguns foram além do papel de professor, passaram a ser tornar mestres com que tive a honra de compartilhar trabalhos, ideias e experiências de vida.

Agradeço ao professor Alcides José Scaglia, que, em 2010, apostou em um garoto de 17 anos com um projeto sem grandes perspectivas nas mãos. Me auxiliou a permanecer em Limeira quando não tinha condições financeiras, me deu a chance de liderar diversos projetos, sem nunca ter me tirado a liberdade e que sempre me tratou como um filho. Da mesma forma que falo aos meus pais que espero estar honrando a família, te digo o mesmo, espero estar honrando tudo o que você construiu. Foi até agora e será sempre uma honra trabalhar como você.

Agradeço também ao professor Cristiano Torezzan. Este professor que me ensinou a pensar de forma organizada, que não deixou em nenhum momento com que eu desistisse do mestrado e que em diversas situações ia até minha casa para ver se eu tinha aprendido o conteúdo ou se estava precisando de alguma ajuda. Obrigado por me acolher no momento de transição em que estava e me orientar até agora. Espero em nenhum momento ter te decepcionado durante este caminho. Desculpa pelos dias que não pude me dedicar ao

máximo, desculpe pelos dias que não rendi e obrigado novamente por tudo.

Agradeço ao professor João Paulo Medina, que, apesar da vivência conturbada nestes últimos meses, sempre serei grato pelas conversas e discussões. Em pouco tempo cresci muito com teu auxílio, profissional e pessoalmente. Aliás, muito obrigado por ter me ensinado a respirar.

Agradeço de forma especial a todos os professores que me acompanharam nesta etapa no curso de Engenharia de Produção e Manufatura e a banca pela colaboração com o trabalho aqui apresentado. Graças a vocês, o trabalho evoluiu muito neste espaço de tempo entre a qualificação e a defesa.

Por fim, acredito que na vida muito do que devemos agradecer é pelos momentos que não serão publicados em nenhum artigo e pelas pessoas que aparecem e te fazem sentir especial. Alessandra, obrigado por compartilhar o teu sorriso comigo. Obrigado por estar comigo nestes momentos e de demonstrar companheirismo em cada instante. Hoje sei que sou capaz de amar alguém graças a você.

## Resumo

Este trabalho apresenta algumas aplicações de métodos matemáticos presentes em Pesquisa Operacional no futebol. A dissertação apresenta três estudos: no primeiro utiliza-se um processo de Markov com o objetivo estimar a probabilidade de uma equipe marcar ou sofrer gols, em função do histórico de transição de bola pelas zonas do campo. Para tanto, uma matriz de transição é construída observando-se o deslocamento da bola através de zonas pré-definidas do campo e dois estados absorventes: realizar o gol ou perder a bola. Para validação, apresenta-se dois estudos com base em dados reais coletados de competições oficiais. Com estes dados, podemos verificar que a análise passa a ser completa apenas quando observado o momento defensivo e momento ofensivo do jogo, sendo que, na observação de um momento isolado, os dados podem não refletir o resultado do placar. Na segunda investigação apresenta-se um modelo de programação linear inteira para a designação de árbitros para partidas em campeonatos de futebol. A função objetivo minimiza a distância total percorrida pelos árbitros e as restrições modelam as preferências e normas definidas pelos responsáveis pela escala. O modelo foi testado com dados reais do Campeonato Brasileiro de Futebol da Série A de 2014 e os resultados mostram que, utilizando a proposta deste artigo, seria possível economizar até 38,25% na distância total percorrida pelos árbitros e, ainda, respeitar um limite máximo de duas partidas do mesmo árbitro para cada time e um total entre dezoito e vinte partidas por árbitro no campeonato. O modelo proposto por ser adaptado para incorporar diferentes regras e preferências, ou ainda ser utilizado para escalas de árbitros em outros esportes. No terceiro estudo, utiliza-se Análise Envoltória de Dados (DEA) para analisar as eficiências relativas de clubes e jogadores de futebol. O trabalho foi realizado com base em dados reais, considerando os 20 clubes mais ricos do mundo em 2015 e os 23 melhores jogadores do mundo segundo o ranking FIFA. Além da análise das eficiências, propomos também a formação de ranking das unidades analisadas com base na média dos pesos ótimos atribuídos por cada unidade. Observando os cenários em análise, verifica-se que tanto as equipes quanto os jogadores mais eficientes não são os de maior investimento, o que sugere que os altos valores alocados em salários não estão retornando na mesma escala em desempenho dos clubes ou dos jogadores. Para sua conclusão, no último capítulo é apresentado as considerações sobre as possíveis aplicações dos modelos em diferentes situações práticas.

Palavras-Chaves: Cadeia de Markov; Problema de Designação; Análise Envoltória de Dados; Pesquisa Operacional; Futebol.

# Abstract

This study presents some applications of mathematical methods of Operations Research in soccer. This dissertation is composed of three studies; in the first we use a Markov process to estimate the probability of a team to make or suffer goals, by the historic of ball transition in field zones. Therefore, a transition matrix is constructed by observing the transition of the ball through pre-defined zones of the field and two absorbing states: goal and miss the ball. For validation two studies are presented based on actual data collected from official competitions. In the second study we present a model of integer linear programming for designation of referees to matches of football championships. The objective function minimizes the total distance traveled by the referees, the constraints models preferences and rules defined by the manager. The model was tested with real data from the 2014 Brazilian Serie A Football Championship. Results shows that using the proposed model can save up to 38.25% of the total distance traveled by the referees and it is also possible to respect: a maximum limit of two matches per referee for each team; and a total between eighteen and twenty matches per referee in the championship. The model proposed can be adapted to incorporate different rules or preferences or even to be used in other sports. The third study applies data envelopment analysis (DEA) to quantify clubs and soccer players efficiency. This model was tested with real data from the 20 richest clubs in the world in 2015 and the 23 best players in the world according to the FIFA ranking. In addition to the efficiency analysis we also propose a classification of the objects, based on this analysis we could realize that both, teams and the most efficient players, are not the most highly efficient, suggesting that high investment in salary are not returning on the same scale as club or player performance. For its conclusion, the last chapter presents the considerations about the possible applications of the models in different practical situations.

Key Words: Markov Chain; Designation Problem; Data Envelopment Analysis; Operations Research; Soccer.

# Lista de Figuras

1	Estrutura da Dissertação. . . . .	14
1.1	Grafo de ilustração do exemplo desta seção. . . . .	18
1.2	Divisão das zonas. . . . .	21
1.3	Transição defensiva. . . . .	22
1.4	Transição Ofensiva. . . . .	23
1.5	Tela do <i>software</i> Skout 1.0. . . . .	23
1.6	Exemplo de matriz elaborada no Excel. . . . .	24
1.7	Zonas de início do momento ofensivo da Alemanha. . . . .	25
1.8	Transições da Zona 4 da Alemanha. . . . .	26
1.9	Transições da zona 5 da Alemanha. . . . .	26
1.10	Transições da zona 8 da Alemanha. . . . .	27
1.11	Probabilidade de gols das equipes semi finalistas da Copa do Mundo 2014. . . . .	27
1.12	Zonas de início do momento ofensivo do Barcelona. . . . .	28
1.13	Transições da Zona 4 do Barcelona no momento ofensivo. . . . .	29
1.14	Transições da zona 5 do Barcelona no momento ofensivo. . . . .	29
1.15	Transições da zona 8 do Barcelona no momento ofensivo. . . . .	30
1.16	Zonas de início do momento defensivo do Barcelona. . . . .	30
1.17	Transições da Zona 5 do Barcelona no momento defensivo. . . . .	31
1.18	Transições da zona 4 do Barcelona no momento defensivo. . . . .	31
1.19	Transições da zona 1 do Barcelona no momento defensivo. . . . .	32
1.20	Probabilidade de gols marcados e sofridos dos finalistas da <i>Champions League</i> . . . . .	32
2.1	Número máximo de partidas designadas a um mesmo árbitro no Campeonato Brasileiro de 2014 – Série A. . . . .	38
2.2	Número de partidas apitadas para cada um dos 20 árbitros que mais atuaram no Campeonato Brasileiro de 2014. . . . .	38
2.3	Distância total percorrida (em Km) pelos árbitros no Campeonato Brasileiro de 2014 – Série A em comparação com o resultado da otimização. . . . .	41
2.4	Número total de partidas atribuídas a cada árbitro pelo modelo. . . . .	42
2.5	Número máximo de partidas designadas a um mesmo árbitro para uma mesma equipe. . . . .	42



3.1	Curva de um processo de produção (Mello <i>et al.</i> , 2005). . . . .	46
3.2	Ilustração da eficiência nos modelos BCC e CCR (Kumar; & Gulati, 2008). . . . .	49
3.3	Dados normalizados de eficiência dos 23 melhores jogadores do mundo em 2015. . . . .	53
3.4	Dados normalizados de eficiência dos 20 clubes mais ricos em 2015. . . . .	54
3.5	Resultados CCR e BCC para eficiência dos 20 clubes mais ricos em 2015. . . . .	55
3.6	Ranking de Eficiência por <i>Cross Efficiency</i> dos Clubes. . . . .	55
3.7	Resultados CCR e BCC para eficiência dos 23 melhores jogadores do mundo em 2015. . . . .	56
3.8	Ranking de Eficiência por <i>Cross Efficiency</i> dos jogadores. . . . .	57

# Sumário

<b>1</b>	<b>Processo de Markov para transição de bola em futebol</b>	<b>15</b>
1.1	Introdução . . . . .	15
1.2	Aplicação de processo de Markov no futebol . . . . .	18
1.3	A ideia intuitiva . . . . .	20
1.4	Definição do problema . . . . .	20
1.5	Modelo proposto . . . . .	21
1.6	Resultados . . . . .	24
1.7	Conclusão . . . . .	33
1.8	Implicações práticas . . . . .	34
<b>2</b>	<b>Problema de designação em escala de árbitros de futebol</b>	<b>35</b>
2.1	Otimização de recursos na designação de árbitros no futebol . . . . .	35
2.2	Definição do problema . . . . .	37
2.3	Modelo proposto . . . . .	39
2.4	Resultados . . . . .	40
2.5	Conclusão . . . . .	43
2.6	Implicações práticas . . . . .	43
<b>3</b>	<b>Aplicação de DEA para ranqueamento de eficiência no futebol</b>	<b>45</b>
3.1	Introdução . . . . .	45
3.2	DEA (análise envoltória de dados) . . . . .	47
3.3	Propostas do estudo . . . . .	52
3.4	Resultados . . . . .	54
3.5	Conclusão . . . . .	57
3.6	Implicações práticas . . . . .	58

# Introdução

Pesquisa Operacional (PO) pode ser descrita como uma abordagem científica interdisciplinar quantitativa para a solução de problemas que envolvem gestão de sistemas complexos. A escolha e a implementação de soluções mais eficazes é cada vez mais valorizada nos mais diversos setores de atuação humana. Tais soluções normalmente demandam interações complexas entre pessoas, materiais, processos e aspectos monetários. Nestes cenários, a Pesquisa Operacional tem sido intensamente utilizada como um método sistemático para apoio a decisão em negócios, indústrias e governos.

O surgimento da Pesquisa Operacional coincide com a Segunda Guerra Mundial, quando as primeiras aplicações foram levadas a cabo em problemas militares, principalmente relacionados a aspectos logísticos e estratégicos.

Desde seu surgimento a área de PO evoluiu e, atualmente, tem sido utilizada como auxílio na tomada de decisão em organizações que visam a melhora constante de desempenho, nas mais variadas áreas.

De maneira geral, os problemas estudados na área de PO podem ser divididos em dois grandes grupos: problemas estáticos, cujos componentes são conhecidos a priori e não há aleatoriedade em sua ocorrência. A segunda classe de problemas é denominada de problemas dinâmicos, onde pelo menos um dos parâmetros (ou componentes) não é conhecido de forma exata, mas descrito em termos de distribuição de probabilidades.

Em ambos os casos podemos entender que o foco da PO está relacionado a um melhor uso técnico, econômico, social ou político de recursos, escassos em sua maioria. Este cenário pode ser observado em inúmeros métodos de PO, principalmente em aplicações no âmbito da Programação Linear e otimização em geral.

Segundo Winston (2003) a construção do pensamento em PO envolve geralmente os sete passos a seguir:

1. Formulação do problema;
2. Observação do sistema e coleta de dados;
3. Modelagem matemática do problema;
4. Verificação do modelo e uso para a obtenção de predições;
5. Seleção de uma solução alternativa;
6. Apresentação dos resultados;
7. Implementação e avaliação de recomendações.

Embora a PO tenha apresentado resultados incontestáveis em muitas áreas como engenharia, economia, marketing e administração, é surpreendente que sua aplicação no esporte ainda seja bastante tímida, sobretudo no futebol.

Segundo Ribeiro & Urrutia (2004) a gestão esportiva é uma área muito atraente para aplicações de pesquisa operacional, dado que as competições esportivas envolvem muitas questões econômicas e logísticas. Entre os temas ligados ao esporte inclui-se: a avaliação de atletas e equipes, planejamento de competições específicas ou calendários completos, e gestão de clubes (questões econômicas ou de marketing). Também é notável o grande número de tomadores de decisão existentes, o treinador e a comissão técnica, os diretores e presidentes de clubes ou federações, jogadores, jornalistas e torcedores.

Entre as possíveis aplicações, pode-se notar que as mesmas são vistas em diversas áreas como:

- Análise de desempenho da equipe ou de atletas;
- Análise econômica de clubes, federações ou confederações;
- Organização de competições, que passa da elaboração de um calendário esportivo à escalação dos profissionais responsáveis;
- Auxílio a tomada de decisão de diretores de diversos setores de um clube, como marketing, financeiro ou logístico.

Apesar de cada tomador de decisões (sejam de clubes, federações ou comissões) ter seu próprio procedimento de análise, pode-se identificar que tais decisões têm ficado mais complexas, tornando cada vez mais necessário o uso de ferramentas de apoio à tomada de decisão.

Sendo assim, o campo de tomada de decisão é amplo, o que permite ao pesquisador diversas aplicações dos modelos matemáticos em diferentes cenários. Com este intuito, os capítulos a seguir foram produzidos a fim de que se possa ter uma visão mais ampliada sobre as diversas aplicações de modelos matemáticos presentes na pesquisa operacional no futebol.

A Figura 1 sintetiza a estrutura desta dissertação bem como as principais problemáticas e contribuições.

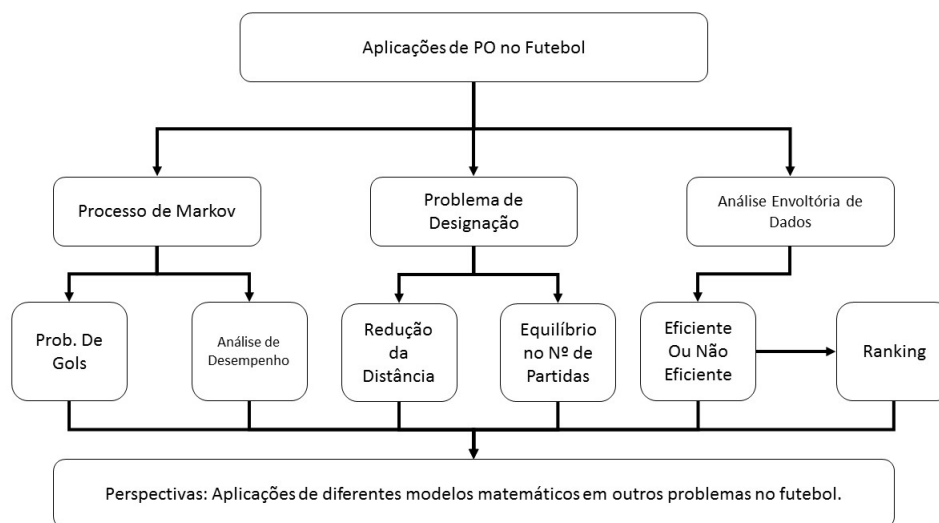


Figura 1: Estrutura da Dissertação.

No Capítulo 1, apresenta-se uma proposta original de aplicação de simulação através do processo de Markov em partidas de futebol. Os estudos apresentados neste capítulo têm como objetivo determinar a probabilidade de realizar ou sofrer gols por parte das equipes analisadas, através de uma matriz de transição composta pela movimentação da bola em campo ou posicionamento dos atletas. Para tal, foram abordados estudos com partidas oficiais da Copa do Mundo de 2014 e *Champions League* de 2014-2015.

No Capítulo 2, realiza-se um estudo sobre o problema de escolha de árbitros profissionais de futebol para apitarem partidas do Campeonato Brasileiro. Para o entendimento do mesmo, foi proposto um modelo de designação que tem como objetivo minimizar a distância total percorrida pelos árbitros durante a competição. Já as restrições modelam as preferências e normas definidas pelos responsáveis pela escala. Este modelo foi testado com dados reais referentes ao Campeonato Brasileiro de 2014. Este capítulo foi publicado em formato de artigo Anon *et al.* (2017).

Já no Capítulo 3, apresenta-se uma proposta de análise de eficiência utilizando o DEA aplicado para clubes e jogadores de futebol. Para sua realização, aplicou-se o modelo em dados reais referentes aos 20 clubes mais ricos do mundo em 2015 e os 23 melhores jogadores do mundo segundo o *ranking* FIFA. Após a obtenção da eficiência, ainda foi possível a obtenção de um *ranking* através das médias de cada peso obtido na análise envoltória de dados.

Por fim, no último capítulo são apresentadas algumas conclusões obtidas através dos estudos realizados nos capítulos anteriores e considerações sobre a possibilidade de aplicação dos modelos abordados em situações práticas.

# Capítulo 1

## Processo de Markov para transição de bola em futebol

Neste capítulo apresentamos uma aplicação original de um processo de Markov em simulação de partidas de futebol. Este estudo tem como objetivo estimar a probabilidade de uma equipe marcar ou sofrer um gol, demonstrar os usos da matriz de transição para o melhor entendimento da partida. Essa probabilidade é dada pela matriz de transição, que é construída observando o histórico de deslocamento da bola por zonas pré-determinadas do campo (transitórias ou absorventes: marcar ou sofrer o gol, perder ou ganhar a posse de bola). Para demonstrar o modelo, apresenta-se dois estudos em competições oficiais.

### 1.1 Introdução

Pode-se definir um processo estocástico como uma sequência de eventos em que o resultado, em qualquer fase, depende de alguma probabilidade.

Dentre o conjunto de modelos matemáticos presentes com esta característica, destaca-se para este estudo o processo de Markov, que é caracterizado pelas seguintes propriedades:

- a. O número de resultados ou estados possíveis é finito;
- b. O resultado em qualquer estado depende apenas do resultado do estado anterior;
- c. As probabilidades são constantes ao longo do tempo.

Se um estado inicial de um sistema for representado por um vetor  $x_0$ , pode-se assumir que existe uma matriz  $M$ , de forma que o estado seguinte é representado pelo vetor  $Mx_0$ . Se seguirmos com as interações, obtemos uma cadeia de vetores de estado:  $\{x_0, Mx_0, M^2x_0, \dots\}$ . Após  $n$  interações, o estado do sistema passa a ser dado por  $x_{n+1} = M^n x_0$ . Esta cadeia de interações é denominada processo de Markov (Shamblin & Stevens, 1979) e a matriz  $M$  representa a matriz de transição entre os estados.

Os vetores formados pelo produto entre as potências da matriz  $M$  pelo vetor  $x_0$  são denominados vetores de estados. Estes vetores de estados podem ser classificados em dois

tipos: absolutos, quando representam quantidades de objetos em um dado estado; ou de probabilidade, quando suas entradas representam o percentual (ou probabilidade) de objetos em um dado estado.

Nesta dissertação, os vetores de estado representarão probabilidades. Nestes casos, a soma de todas as coordenadas de um vetor de estado é igual a 1. O Teorema 1 abaixo apresenta um resultado importante sobre a convergência da sequências de estados de um processo de Markov.

**Teorema 1.** Seja  $M$  a matriz de transição de um processo de Markov tal que  $M^k$  tenha somente entradas positivas para algum  $k$ . Existe um único vetor de probabilidade  $x_s$  tal que  $Mx_s = x_s$ . Além disso  $\lim_{k \rightarrow \infty} M^k x_0 = x_s$  para qualquer vetor de probabilidade de estado inicial  $x_0$ .

O vetor  $x_s$  é denominado de “vetor estacionário”. O vetor estacionário pode ser calculado, observando-se que:

$$Mx_s = x_s \Rightarrow Mx_s - x_s = 0 \Rightarrow (M - I)x_s = 0,$$

Ou seja, o vetor  $x_s$  pertence ao núcleo da matriz  $(M - I)$ . Como ele é único pelo Teorema 1, basta encontrarmos uma solução não nula do sistema  $(M - I)x_s = 0$ . Em MatLab isto pode ser feito com o seguinte comando:  $x_s = \text{null}(M - I)$ . Isto equivale a fazer  $x_s = \lim_{n \rightarrow \infty} M^n x_0$ .

## Classificações de processo de Markov

Define-se o conjunto de resultados possíveis em um processo de Markov como estados. Neste trabalho representar-se-á os estados por  $S = s_1, s_2, \dots, s_n$ . Sendo assim, a formação da matriz de transição se dará pela transição entre estes estados.

Desta forma, pode-se denominar  $m_{ij}$  a probabilidade de transição entre os estados  $s_j$  e  $s_i$ . O conjunto de todas as transições pode ser representado por uma matriz  $M$ :

$$M = [m_{ij}] = \begin{pmatrix} m_{11} & \dots & m_{1n} \\ \vdots & \ddots & \vdots \\ m_{n1} & \dots & m_{nn} \end{pmatrix}$$

Sendo que seus elementos apresentam as seguintes propriedades:

- $m_{ij} \geq 0$ ;
- $\sum_{j \in S} m_{ij} = 1, \forall i \in S = \{1, \dots, h\}$

O  $n$ ésimo estado em uma cadeia de Markov é dado por  $x_n = Mx_{n-1} = M^n x_0$ .

Os estados de um processo de Markov podem ser caracterizados como: alcançável, comunicante ou irredutível. Um estado  $j$  é denominado alcançável a partir de um estado  $i$  se  $m_{ij}^{(t)} > 0$ . Já um estado  $j$  é dito comunicante com o estado  $i$  se os dois estados são

alcançáveis entre si. Se todos os estados do sistema são comunicantes, a cadeia de Markov é dita irredutível.

Além destas classificações, um processo de Markov pode se caracterizar por ser transiente, recorrente ou absorvente. Uma cadeia de Markov se caracteriza como transiente se, entrando nesse estado, o processo pode nunca retornar novamente para este estado, ou seja, um estado  $j$  diz-se transiente se existir um estado  $i$  que seja atingível a partir de  $j$ , mas não sendo  $j$  atingível a partir de  $i$ . Já um estado absorvente é quando  $m_{ij} = 1$ . Sendo assim, quando se entra no estado  $j$ , não se sai mais dele.

Além disso, uma cadeia de Markov é dita periódica, com período  $k > 1$ , se todos os caminhos que saem e entram do estado  $i$  tem o mesmo comprimento ou são múltiplos de  $k$ . Caso contrário, estes estados são aperiódicos.

Podemos também definir se uma cadeia de Markov é ergódica ou não. Se todos os estados forem recorrentes, aperiódicos e comunicantes, a cadeia é dita ergódica. Caso contrário ela não é ergódica. Em uma cadeia ergódica, decorrido um número muito elevado de períodos de tempo  $n$ , a probabilidade de o processo estar no estado  $j$  passa a ser constante, independente do estado inicial.

Caso a cadeia não seja ergódica e apresente um estado absorvente, após um período elevado de tempo  $t$ , os estados transientes tendem a 0, pois a probabilidade de se estar em algum dos estados absorventes é grande, afinal, ao longo das transições, toda vez que uma transição passar para este estado, ela permanecerá nele. Sendo assim, para este tipo de processo de Markov, observar em que tempo  $t$  ocorre o estado absorvente passa a ser de fundamental importância.

A probabilidade de o processo transitar de um dos estados transientes para um estado absorvente num qualquer número de períodos é denominada probabilidade estacionária de absorção.

## Exemplo

Suponha que em um dado mercado de refrigerantes só havia uma empresa, denominada 2 e que acaba de ser inaugurada outra empresa, 1. Após uma pesquisa realizada, notou-se que quem compra o refrigerante 1, volta a comprar o mesmo refrigerante em 90% das vezes. Os outros 10% compram o refrigerante 2. Já quem compra o refrigerante 2, compra o mesmo em 80% das vezes e 20% passam a comprar o refrigerante 1, como o demonstrado na Figura 1.1 a seguir.



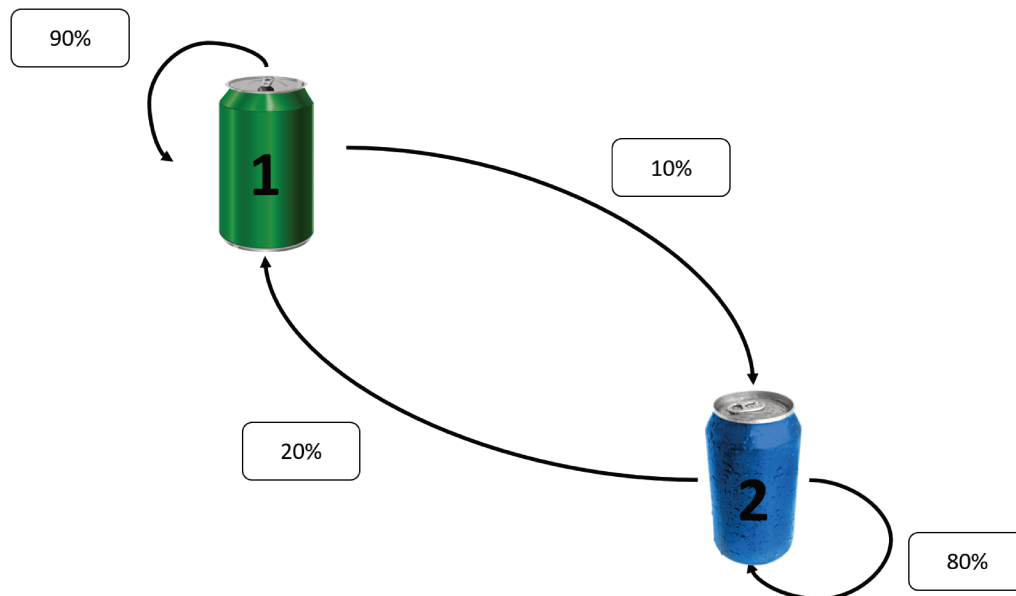


Figura 1.1: Grafo de ilustração do exemplo desta seção.

Este problema pode ser modelado como um processo de Markov com a seguinte matriz de transição:

$$M = \begin{pmatrix} 0,9 & 0,2 \\ 0,1 & 0,8 \end{pmatrix}$$

Suponha que, no estado inicial, 100% dos clientes estavam comprando a marca 2, ou seja,  $x_0 = (0, 1)^t$ . Assim, após duas transições, o vetor de estados passa a ser  $Mx_2 = (0,34, 0,66)$ . Após  $n$  transições, o vetor de estados é dado por  $M^n x_0$ .

Note que a matriz  $M$  satisfaz as condições do Teorema 1, portanto, existe um estado estacionário para este processo, que pode ser obtido calculando  $\lim_{n \rightarrow \infty} M^n x_0$ . Neste caso,  $x_s = (0,67, 0,33)^t$ .

Sendo assim, pelo exemplo, pode-se notar que a empresa 1 após um número suficientemente grande de transições, tende a ter 67% do mercado, mesmo que inicialmente todos os consumidores comprem da empresa 2.

## 1.2 Aplicação de processo de Markov no futebol

No futebol, a possibilidade de prever o resultado da partida antes mesmo de seu início tem atraído interesse por parte de pesquisadores, profissionais do futebol e pessoas interessadas na área como jornalistas e apostadores (Gomes, 2015; Severino, 2013). Porém, modelar um sistema complexo como o futebol é uma tarefa difícil, tendo em vista a diversidade de variáveis que podem ser observadas.

A observação e análise de um jogo de futebol já foram denominadas de diversas maneiras, dentre elas: observação do jogo (*game observation*), análise do jogo (*match analysis*)

e análise notacional (*notational analysis*) (Garganta, 2001). Porém, ainda segundo Garganta (2001), a expressão mais utilizada na literatura é análise do jogo. Esta expressão engloba, em sua definição, diferentes fases do processo: a observação dos acontecimentos, a notação dos dados e a sua interpretação (Franks & Goodman, 1986; Hughes, 1996).

Garganta (2001) aponta que a análise de desempenho de uma equipe pode contribuir para a mesma das seguintes maneiras:

- 1) configurar modelos da atividade dos jogadores e das equipes;
- 2) identificar os traços da atividade cuja presença/ ausência se correlaciona com a eficácia de processos e a obtenção de resultados positivos;
- 3) promover o desenvolvimento de métodos de treinamento que garantam uma maior especificidade e, portanto, superior transferibilidade;
- 4) indiciar tendências evolutivas das diferentes modalidades esportivas.

Além das contribuições oferecidas para as equipes, diversos autores tentaram entender o problema de se analisar partidas através de previsões de resultados. Gamerman & Gamerman (2003) apresentam como base de seu modelo o número de gols marcados e sofridos, além do fator campo. Farias (2008) considera o número de gols ao longo da temporada como base e Alves *et al.* (2008) apontam o nível da equipe e local da partida.

Entre os estudos que apresentaram a utilização de processo de Markov no futebol, podemos destacar Seligman (2011). Neste estudo, há a comparação entre o futebol jogado no campeonato italiano e brasileiro através do resultado do campeonato, a qualidade das equipes e o local onde o jogo foi realizado. Como conclusão, apresentou-se que no campeonato brasileiro, a influência do mando de campo é maior que no italiano e que no a influência da qualidade de equipe prevalece no campeonato italiano. Outro estudo de destaque foi apresentado por Rue & Salvesen (2000), ao tentar prever resultados com base nos gols e no local de jogo, principalmente na vantagem de se jogar em casa.

Este capítulo tem como finalidade apresentar um modelo original baseado no processo de Markov para a determinação da probabilidade de uma equipe realizar e de se sofrer gols. Para a análise ofensiva foram observadas oito zonas de campo, que compõem os estados transientes e dois estados finais que são compostos pelo gol e o final da jogada. Suas transições são mensuradas através das ações realizadas com a bola. Já para a análise defensiva foram observadas as mesmas oito zonas de campo que compõem os estados transientes e dois estados finais que são compostos pela recuperação da posse ou por sofrer o gol. A matriz que se obtêm das transições é composta pelo posicionamento dos jogadores de defesa a cada ação com bola da equipe adversária.

Para este objetivo, serão apresentados dois casos realizados. No primeiro foi analisada a Copa do Mundo FIFA 2014, tendo como foco os quatro semifinalistas, apenas observando as ações ofensivas da equipe. No segundo foram analisadas as seis últimas partidas das equipes do Barcelona e da Juventus antes da final da UEFA *Champions League* 2014-2015, com análises ofensiva e defensiva das equipes.

### 1.3 A ideia intuitiva

Antes de apresentar uma definição formal do modelo, pode-se descrever a ideia-chave intuitivamente. Imagine que você está assistindo seu jogo de futebol favorito e, de repente, todos os jogadores tornam-se invisíveis e apenas a bola pode ser vista movendo pelo campo. Os jogadores continuam lá, vendo uns aos outros e jogando normalmente, apenas você não consegue vê-los. Para facilitar o acompanhamento do jogo, a bola apareça azul quando a posse é do time A e vermelha quando a posse é do time B.

A invisibilidade fictícia dos jogadores permite olhar para o futebol de uma forma bastante peculiar, que torna mais evidente os objetivos táticos do jogo: tocar a bola pelo campo até marcar o gol. Como os jogadores não estão visíveis, torna-se muito importante acompanhar a transição da bola de uma zona para outra do campo. Neste sentido, a chance de um time marcar um gol é tão maior quanto maiores forem as probabilidades da bola transitar de uma zona para a outra sem que o time perca sua posse. Para completar a ideia, resta adicionar dois estados artificiais: o gol e a perda da posse.

Neste contexto podemos modelar o jogo de futebol como um processo estocástico com estados discretos, definidos pelas ações do jogo:  $A = \{\text{passe, chute, roubada de bola, ...}\}$ . Assumindo que a probabilidade da bola estar em uma determinada zona após  $k$  transições depende apenas dos estados anteriores podemos construir um processo de Markov para representar nosso modelo e, com isso, calcular a probabilidade da bola estar em qualquer zona do jogo, inclusive de marcar um gol ou de perder a posse. Uma formulação matemática para este modelo é apresentada na seção 1.4.

### 1.4 Definição do problema

O futebol é um esporte composto por duas situações: ataque e defesa. O ataque é caracterizado pela posse de bola de uma dada equipe. Já no instante em que a equipe não está em posse da bola, denomina-se momento defensivo. Desta forma, podemos entender que o resultado de um bom ataque é a realização do gol. Enquanto o de uma boa defesa é evitar o gol do adversário. Este saldo composto pelo número de gols marcados e sofridos é o índice que baliza o resultado, podendo ser ele vitória, empate ou derrota.

Com este propósito, apresentam-se dois indicadores: ofensivo e defensivo. O primeiro mensura a probabilidade de realização de gols por parte da equipe analisada e o segundo a probabilidade de sofrer gols. Outra aplicação deste estudo é a análise das equipes através da matriz de transição obtida na coleta dos dados. Com ele pode-se identificar questões de ordem comportamental das equipes.

Desta forma, o presente estudo desenvolve um modelo constituído de transições da bola dentro do campo de futebol, que fornece base para a formação do processo de Markov, o que permite chegar à probabilidade de gol dado que a equipe observada se encontre em

posse ou a probabilidade de sofrer gols, dado que a equipe não se encontre com posse de bola.

## 1.5 Modelo proposto

Como dito acima, a análise dividi-se em dois momentos. O primeiro é o momento defensivo, que pode-se definir como os instantes em que a equipe não detêm a posse da bola. O modelo tem sua base na concepção de observar o posicionamento do jogador de defesa a cada ação com a bola realizada pela equipe de ataque. Ou seja, a cada ação observada da equipe que está com a posse da bola, coleta-se o posicionamento do jogador de defesa mais próximo. A partir desses pontos, pode-se observar a transição do posicionamento dos jogadores de defesa dentro do campo. Forma-se assim, o campo em oito zonas de transição e duas de estados absorventes, o sofrer o gol e o recuperar a posse.

O segundo momento é o ofensivo, que é definido pelos instantes em que a equipe se encontra em posse da bola. Para constituir o modelo, utilizou-se de uma abordagem pautada no deslocamento da bola dentro do campo de futebol. Para isto, dividiu-se o campo em oito zonas de transição ou estados transientes e duas zonas finais, ou estados absorventes, formadas pelo gol e perda da posse.

### Campo

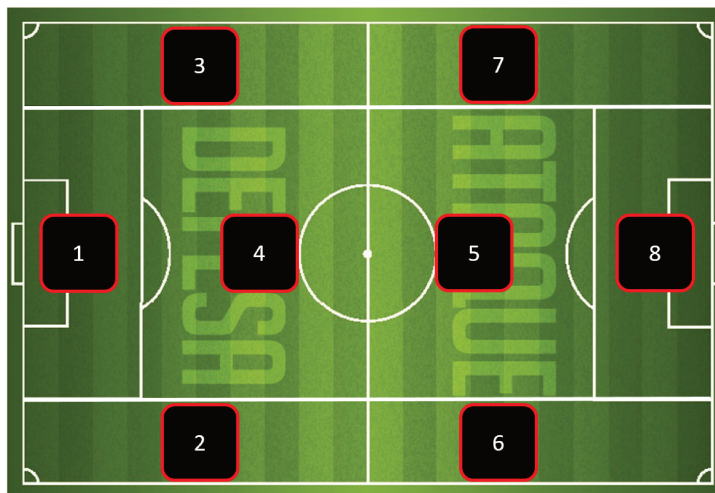


Figura 1.2: Divisão das zonas.

O campo é definido em oito zonas de transição. A divisão do campo levou em consideração as linhas do campo, sendo que as zonas 1 e 8 são compostas pelas grandes áreas, as zonas 2, 3, 6 e 7 são compostas pelas laterais do campo, tendo como referência a linha lateral da área e a divisão da linha do meio campo, zonas 4 e 5 são compostas pela região de meio campo, tendo como referência as linhas frontais da grande área e a linha do meio

campo, como observado na Figura 1.2. Além dessas zonas de transição, temos as zonas que representam os dois estados absorventes. No caso do momento defensivo, os estados absorventes são o sofrer o gol ou a recuperação da posse. No momento ofensivo, gol e perda da posse.

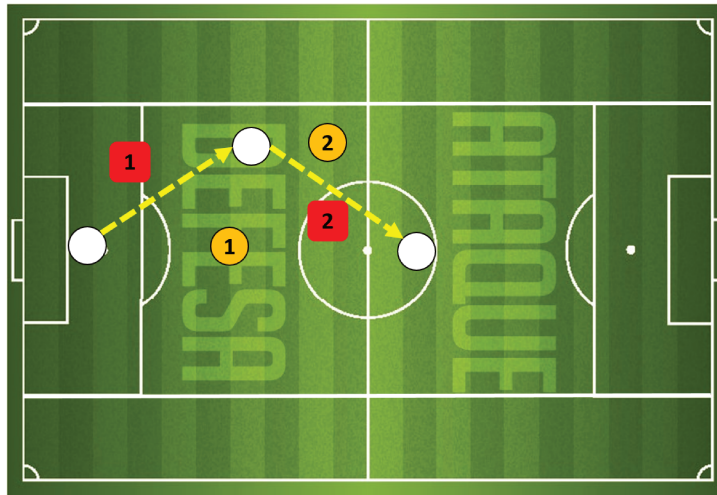


Figura 1.3: Transição defensiva.

Círculo branco: jogadores em posse de bola; círculo laranja: jogadores de defesa; quadrado vermelho: momento da transição; linha tracejada amarela: movimento da bola no campo.

No momento defensivo, compõem-se a matriz de transição com o deslocamento da marcação ao longo das ações ofensivas do adversário. Para este procedimento, marca-se o local onde o jogador de defesa mais próximo se encontra da ação ofensiva do adversário, como demonstrado na Figura 1.3. Nela, observa-se que a linha tracejada amarela é o deslocamento da bola da equipe em posse e que há dois defensores, em destaque no quadrado vermelho. A primeira ação da equipe com posse é representada pelo quadrado vermelho 1, nesta ação, será coletada a informação do posicionamento do jogador de defesa 1, representado pelo bola amarela 1. No segundo momento, a transição do adversário com posse será representada pelo quadrante vermelho 2, nesta ação, será coletada a informação do posicionamento do jogador de defesa 2, representado pela bola laranja 2. Estes deslocamentos podem ter dois fins, o primeiro que é sofrer o gol e o segundo que é finalizar o ataque adversário, seja pela recuperação da posse ou pela paralisação do jogo.

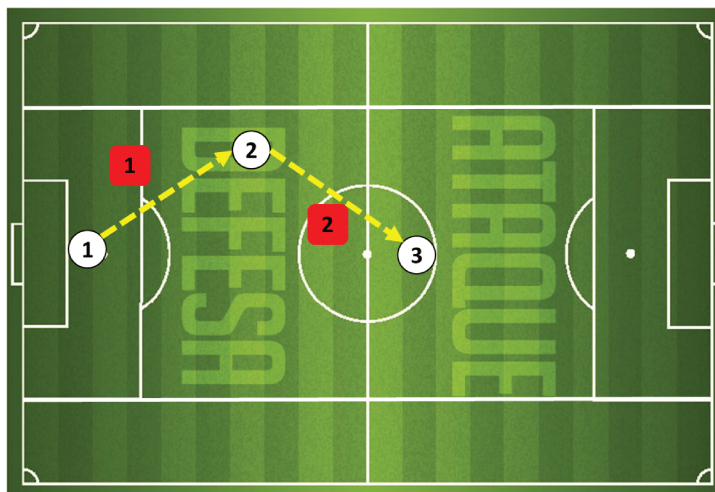


Figura 1.4: Transição Ofensiva.

Círculo branco: jogadores em posse de bola; quadrado vermelho: momento da transição; linha tracejada amarela: movimento da bola no campo.

No momento ofensivo, a cada ação dentro do campo é marcado o local de origem e de destino da bola, sendo este deslocamento o componente da matriz de transição. Este deslocamento só obtém um fim em dois estados, no de perda da posse (seja ela para a equipe adversária ou por paralisação do jogo), ou na de realização do gol. Na Figura 1.4, nota-se que a primeira ação ofensiva é composta pelo deslocamento (quadrado vermelho 1) da bola pela ação do jogador 1 (bola branca 1) para o jogador 2 (bola branca 2). Nesta ação serão coletados o local da bola no momento da ação do jogador 1 e a recepção da bola pelo jogador 2. O mesmo ocorre na troca da posse da bola do jogador 2 para o jogador 3.

## Coleta de dados

Para este estudo, utilizou-se o software de análise de desempenho no futebol “Skout 1.0”, desenvolvido por Bergo *et al.* (1998). Com o software, ilustrado na Figura 1.5, pode-se observar um campo de jogo com valores de X e Y estimados, onde são marcadas as ações técnicas de acordo com o posicionamento da bola.

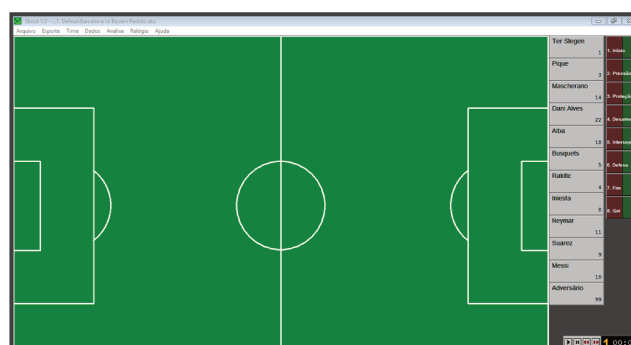


Figura 1.5: Tela do *software* Skout 1.0.

Após a coleta dos dados, a segunda etapa é composta pela tabulação e tratamento dos dados, utilizando-se uma planilha no *software* “Microsoft Excel 2010”. Como o *software* utilizado na coleta exporta os dados em formato txt, houve a necessidade de inserção dos mesmo no “Excel”, que contém fórmulas que convertem os pontos X e Y do campo nas zonas de campo apresentadas na Figura 1.2. Em posse da matriz, exemplificada na Figura 1.6, realiza-se a composição do estado estacionário. Para tal, foi necessária a inserção de uma macro no Excel para a realização do cálculo. Após estas etapas, obtêm-se as zonas de campo e os cálculos necessários no processo de Markov, além da determinação da probabilidade de sofrer o gol ou finalizar o ataque adversário no momento defensivo, e realizar gol ou recuperar a posse no momento ofensivo.

Matriz de Transição										
	1	2	3	4	5	6	7	8	Gol	Fora
1	0,137	0,066	0,074	0,045	0,000	0,000	0,000	0,000	0,000	0,000
2	0,140	0,334	0,013	0,083	0,001	0,025	0,002	0,000	0,000	0,000
3	0,193	0,003	0,272	0,083	0,003	0,000	0,051	0,000	0,000	0,000
4	0,210	0,280	0,282	0,366	0,054	0,065	0,072	0,000	0,000	0,000
5	0,023	0,029	0,028	0,164	0,491	0,301	0,326	0,064	0,000	0,000
6	0,003	0,072	0,003	0,048	0,130	0,397	0,000	0,000	0,000	0,000
7	0,003	0,006	0,056	0,044	0,092	0,004	0,374	0,021	0,000	0,000
8	0,000	0,003	0,003	0,004	0,023	0,012	0,005	0,064	0,000	0,000
Gol	0,000	0,000	0,000	0,000	0,001	0,000	0,000	0,128	1,000	0,000
Fora	0,290	0,207	0,272	0,162	0,206	0,196	0,169	0,723	0,000	1,000

Figura 1.6: Exemplo de matriz elaborada no Excel.

Todos os jogos analisados neste estudo foram gravados por um gravador de DVD com imagens coletadas das emissoras de televisão que transmitiam as partidas em que as equipes do Barcelona e da Juventus estavam presentes, assim como os jogos da Copa do Mundo 2014.

## 1.6 Resultados

Nesta seção serão abordados os resultados dos dois estudos realizados, assim como suas características e diferenças de análise de cada um. Para o primeiro caso, foram observados os jogos de quatro seleções que compõem os semifinalistas da Copa do Mundo de 2014, sendo elas Brasil, Holanda, Argentina e Alemanha, observando somente as ações ofensivas de cada equipe. Neles serão observados os estados estacionários de cada seleção, além de um aprofundamento na análise da equipe campeã, a Alemanha. No segundo caso, foram coletados os 6 últimos jogos das equipes presentes na final da UEFA *Champions League* 2015 antes da final. Neste estudo, foram analisados os dois momentos do jogo, ofensivo e defensivo, sendo apresentados aqui a análise da equipe campeã, o Barcelona, e os estados estacionários de ambas as equipes.



## Semi finalistas da Copa do Mundo 2014

O primeiro caso apresentado constitui uma análise composta pelas quatro equipes semi-finalistas da Copa do Mundo 2014. Para a realização desta pesquisa foram coletados os sete jogos de cada equipe durante a competição, sendo assim, foram analisados neste estudo 28 jogos, 17001 transições foram computadas e 4192 ataques iniciados pelas equipes citadas (média de 149,71 inícios por jogo). Dentre os resultados obtidos, pode-se dividi-los em dois conjuntos, um composto pela análise da equipe através dos elementos que compõem a matriz de transição e outro pelo estado estacionário de cada equipe semi finalista da Copa do Mundo 2014.

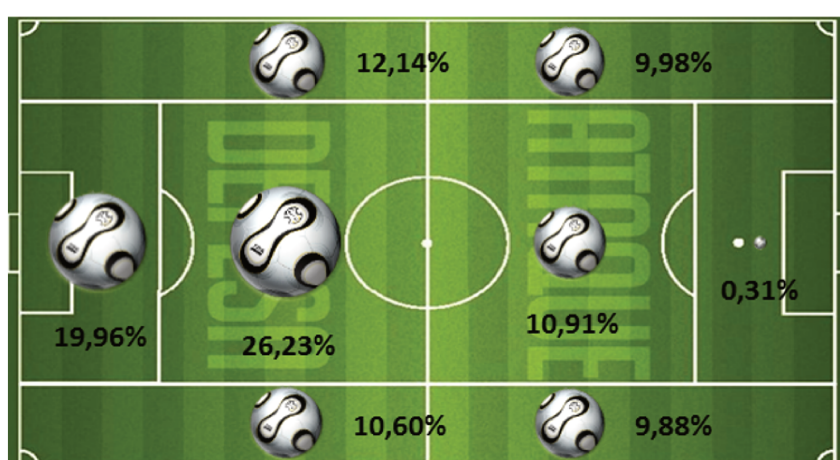


Figura 1.7: Zonas de início do momento ofensivo da Alemanha.  
Bolas de futebol: representação gráfica da porcentagem de ações iniciadas em cada zona.

Na Figura 1.7, estão demonstradas as porcentagem de inícios do momento ofensivo por zona. Pode-se perceber que a equipe da Alemanha na Copa do Mundo de 2014 apresenta como maiores zonas de início de momento ofensivo as zonas 4, 1 e 3. Desta forma, nota-se que grande parte de seus ataques se iniciam em seu campo de defesa (metade do campo mais próxima a área que seu goleiro defende), sugerindo um modelo de construção de momentos ofensivos desde o seu campo de defesa.



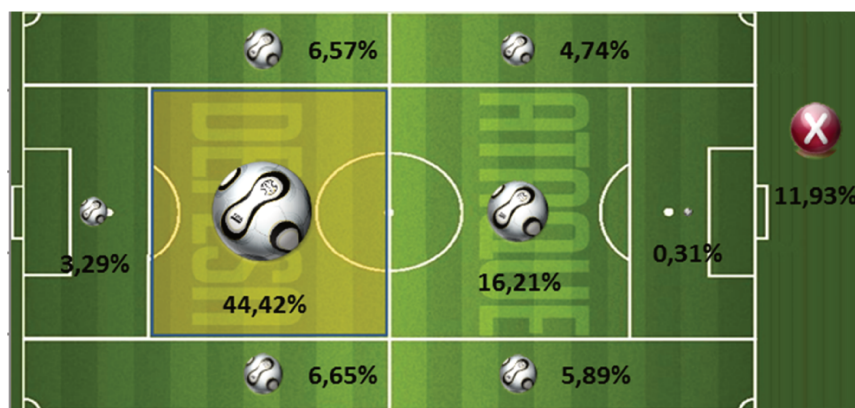


Figura 1.8: Transições da Zona 4 da Alemanha.

Quadrante amarelo: zona de início da transição; bolas de futebol: representação gráfica da porcentagem de destino final da transição; bola vermelha: transições que tiveram como final a perda da posse.

Na Figura 1.8 está ilustrada em amarelo a zona do campo analisada, neste caso a zona 4. Esta zona é a que apresenta maior índice de início de momento ofensivo e, como se pode observar, ela também aponta para uma maior transição da bola na própria zona, além de baixa porcentagem de perdas da posse de bola (11,93%). A segunda zona com maior frequência passa a ser a zona 5, com 16,21%.

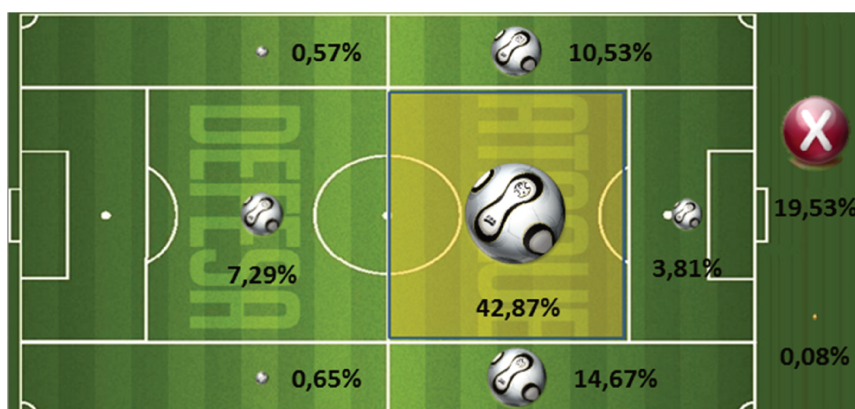


Figura 1.9: Transições da zona 5 da Alemanha.

Quadrante amarelo: zona de início da transição; bolas de futebol: representação gráfica da porcentagem de destino final da transição; bola vermelha: transições que tiveram como final a perda da posse; bola dourada: transições que tiveram como final a realização do gol.

Já na Figura 1.9 estão contidas as porcentagens de transição da bola tendo como ponto de origem a zona 5 (faixa amarela do campo). Nela pode-se notar que apenas 0,08% das bolas que entram nesta zona vão diretamente para o gol. Porém, 3,81% das bolas chegam na zona 8, ou seja, a zona mais próxima ao gol. Nota-se que as zonas que a bola mais transita são as zona 5, 6 e 7, sem contar a perda da posse, o que sugere uma alta posse de bola assim que a mesma chega no campo ofensivo.

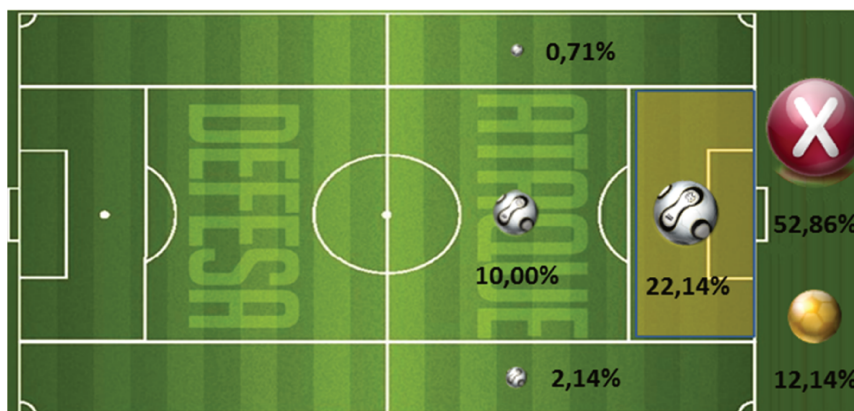


Figura 1.10: Transições da zona 8 da Alemanha.

Quadrante amarelo: zona de início da transição; bolas de futebol: representação gráfica da porcentagem de destino final da transição; bola vermelha: transições que tiveram como final a perda da posse; bola dourada: transições que tiveram como final a realização do gol.

As transições realizadas zona 8 (zona em destaque) é representada na Figura 1.10. Percebe-se que nesta zona do campo a perda da posse de bola é alta em comparação com as demais zonas já analisadas (52,86%) da posse, mas que ao mesmo tempo, 12,14% das bolas que chegam a zona 8 resultam em gols. Sendo assim, a zona 8 constitui a zona de maior impacto direto na realização do gol, o que mostra a importância das demais zonas que tem como característica a transição da bola até a mesma.

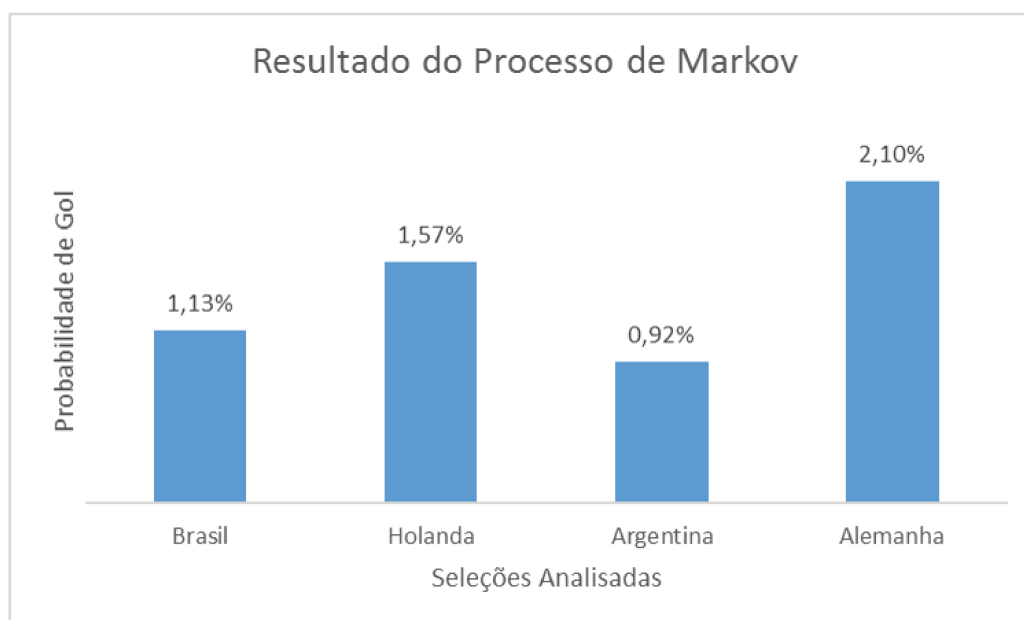


Figura 1.11: Probabilidade de gols das equipes semi finalistas da Copa do Mundo 2014.

Por fim, neste caso, é possível entender o que a forma de transitar a bola da Alemanha resulta no jogo e como ela se diferencia em relação aos seus 3 principais oponentes na Copa do Mundo de 2014. Na Figura 1.11, nota-se que dentre as equipes analisadas, a

que apresenta maior probabilidade de realização de gols é a equipe da Alemanha (2,10%), seguido da equipe da Holanda (1,57%). Se levarmos em consideração apenas as equipes finalistas (Argentina e Alemanha), a diferença passa a ser maior, pois a Argentina apresenta o índice mais baixo (0,92%). Para entender o impacto desta diferença, se ambas as equipes tiverem o mesmo número de ataque, 100 por exemplo, a Argentina teria 39,68% de chances de não realizar um gol sequer, já a Alemanha, apenas 11,97%. Este resultado se obtém elevando a chances de não marcar gols pelo número de ataque.

## Finalistas da *Champions League* 2014-2015

O segundo caso apresentado constitui uma análise composta pelas 2 equipes finalistas da *Champions League* 2014-2015. Para a realização desta pesquisa foram coletados os últimos 6 jogos de cada equipe antes da realização da final do campeonato citado. Foram analisados 12 jogos, 14245 transições computadas sendo 6671 transições de momento defensivo e 7574 transições de momento ofensivo e 3735 inícios realizados nos jogos, sendo 1873 no momento defensivo e 1852 no momento ofensivo. Dentre os resultados obtidos, pode-se dividi-los em três conjuntos. O primeiro é composto pela análise da equipe através dos elementos que compõem a matriz de transição do momento ofensivo, o segundo pela análise da equipe através dos elementos que compõem a matriz de transição do momento defensivo e, por fim, o estado estacionário de ambos os momentos (ofensivo e defensivo) de cada equipe.

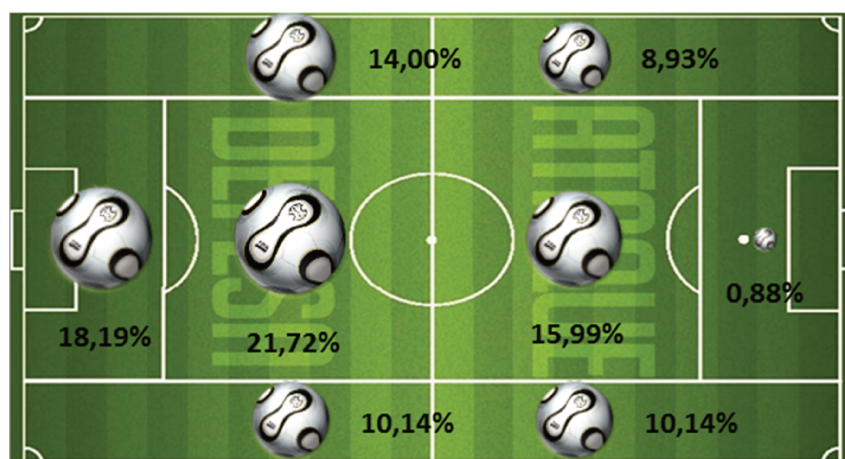


Figura 1.12: Zonas de início do momento ofensivo do Barcelona. Bolas de futebol: representação gráfica da porcentagem de ações iniciadas em cada zona.

Na Figura 1.12, estão contidas as probabilidades de início dos momentos ofensivos do Barcelona por zona. Nela, pode-se notar que suas maiores probabilidades de início ocorrem nas zonas 4, 1 e 5. Além desta informação, pode-se notar que a bola tende a começar no campo defensivo da equipe em 64,05% dos ataques. Este dado sugere que a equipe do Barcelona tende a começar seus ataques no campo defensivo.

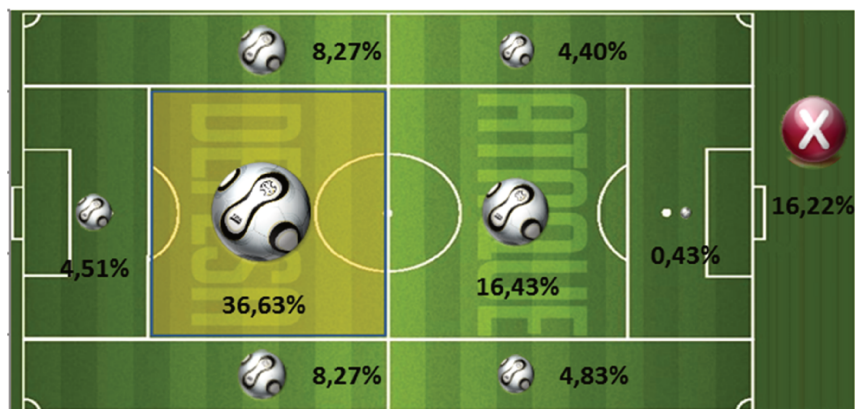


Figura 1.13: Transições da Zona 4 do Barcelona no momento ofensivo.

Quadrante amarelo: zona de início da transição; bolas de futebol: representação gráfica da porcentagem de destino final da transição; bola vermelha: transições que tiveram como final a perda da posse.

A Figura 1.13 representa as transições realizadas a partir da zona 4, destacada na imagem em amarelo. Nela pode-se perceber que a bola tende a ficar na mesma zona (36,63%). Caso saia da zona 4, a bola tende a ir para a zona 5 em 16,43% das vezes e 8,27% da vezes para as zonas 2 e 3, cada uma. Já a transição para a perda da posse de bola é de 16,22%.

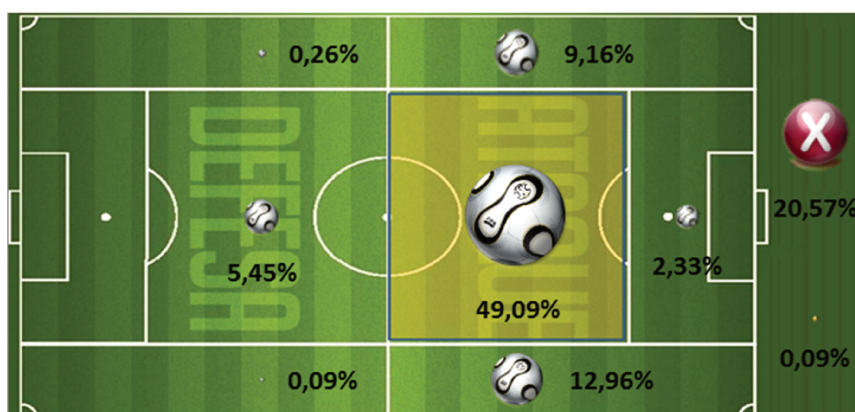


Figura 1.14: Transições da zona 5 do Barcelona no momento ofensivo.

Quadrante amarelo: zona de início da transição; bolas de futebol: representação gráfica da porcentagem de destino final da transição; bola vermelha: transições que tiveram como final a perda da posse; bola dourada: transições que tiveram como final a realização do gol.

Já na Figura 1.14, pode-se notar que na zona 5, a maioria das transições ocorrem para ela mesma: 49,09% das vezes. Outras zonas de destaque são as 6 e a 7, que representam 12,96% e 9,16% das transições, respectivamente. Já a probabilidade da bola ir para o gol é de 0,09% e da perda da posse é de 20,57%.



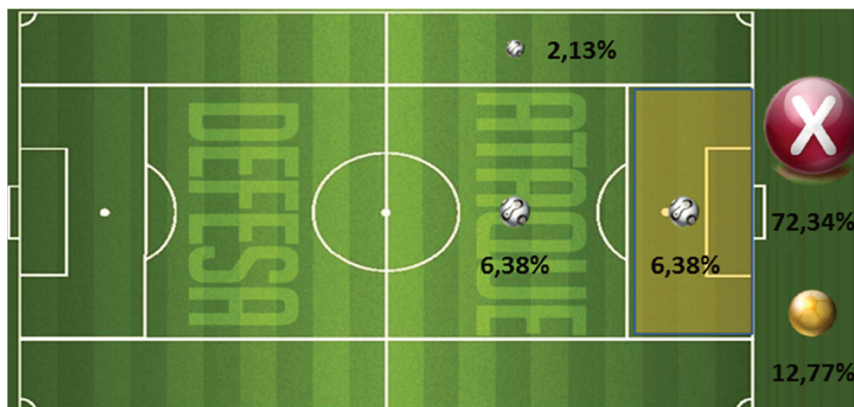


Figura 1.15: Transições da zona 8 do Barcelona no momento ofensivo.

Quadrante amarelo: zona de início da transição; bolas de futebol: representação gráfica da porcentagem de destino final da transição; bola vermelha: transições que tiveram como final a perda da posse; bola dourada: transições que tiveram como final a realização do gol.

Na Figura 1.15, observa-se a probabilidade de transição da bola partindo da zona 8. Desta forma, quando a bola está na zona em destaque, 72,43% das vezes tende a haver a perda da posse ou término do momento ofensivo. Porém, em 12,77% das vezes a bola tende a ir para o gol, resultando na zona de maior transição direta para o gol.

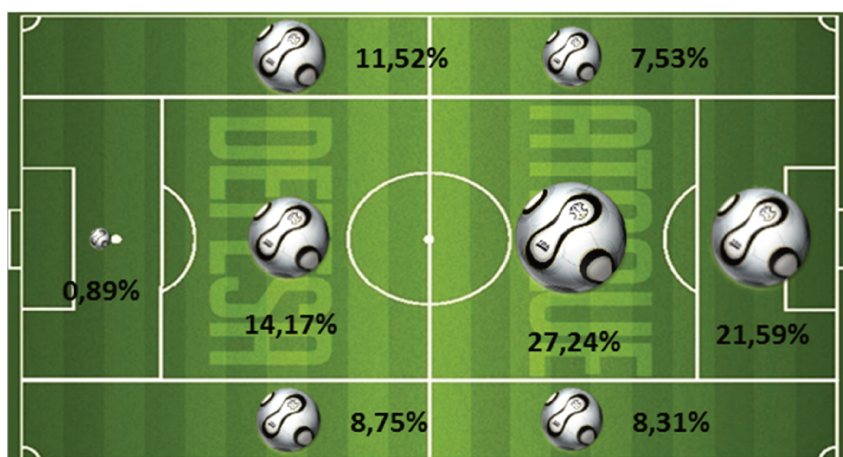


Figura 1.16: Zonas de início do momento defensivo do Barcelona.

Bolas de futebol: representação gráfica da porcentagem de ações iniciadas pelo adversário em cada zona.

A Figura 1.16 é referente as zonas de início no momento defensivo, ou seja, são as zonas que as equipes adversárias iniciam seus ataques e a equipe do Barcelona sua marcação. Na mesma, pode-se notar que as equipes adversárias tendem a começar seus ataques nas zonas 5 e 8. Este dado indica duas possibilidades, a primeira é a de que seus momentos ofensivos tendem a se iniciar de suas zonas de defesa, já a segunda é a de que o final do ataque do Barcelona tende a ser nestas zonas.

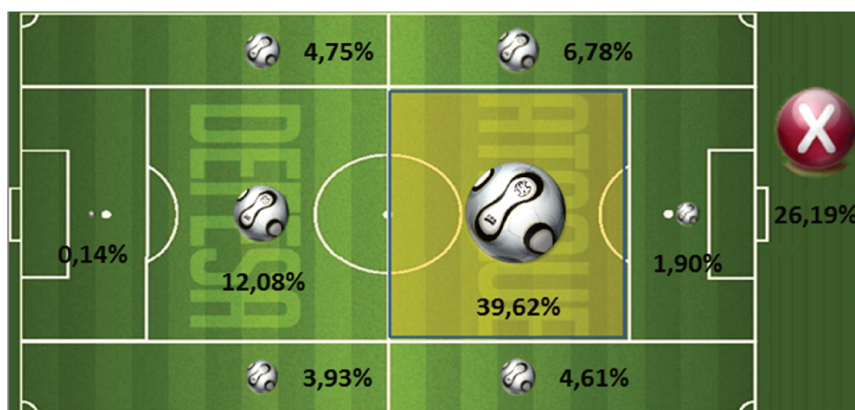


Figura 1.17: Transições da Zona 5 do Barcelona no momento defensivo. Quadrante amarelo: zona de início da transição; bolas de futebol: representação gráfica da porcentagem de destino final da transição; bola vermelha: transições que tiveram como final a ganho da posse.

Sendo a zona 5 a de maior intensidade, a Figura 1.17 apresenta a representação da transição a partir dessa zona em destaque (em amarelo). Como pode se notar, o Barcelona recupera 26,19% das transições que ocorrem, além disso em 39,62% das ocasiões a bola não sai da própria zona. Em apenas 20,90% das vezes, a bola chega ao campo de defesa do Barcelona.

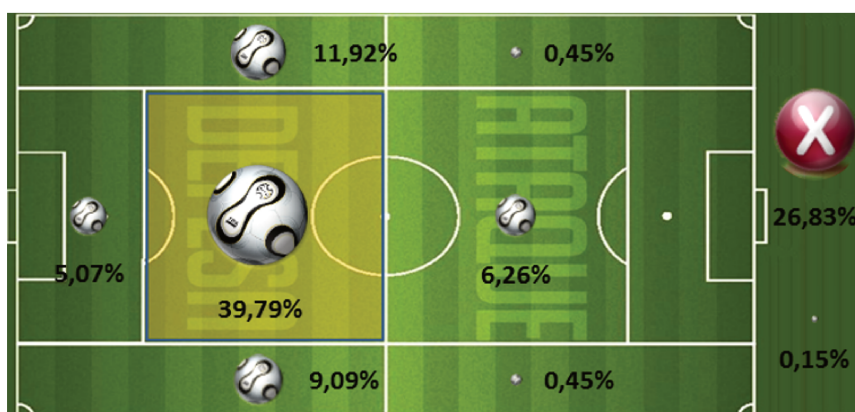


Figura 1.18: Transições da zona 4 do Barcelona no momento defensivo. Quadrante amarelo: zona de início da transição; bolas de futebol: representação gráfica da porcentagem de destino final da transição; bola vermelha: transições que tiveram como final o ganho da posse; bola dourada: transições que tiveram como final sofrer o gol.

Já na zona 4, quando observa-se a Figura 1.18, percebe-se que a bola fica na mesma zona em 39,79% das vezes. Em 26,83% das ocasiões, a equipe do Barcelona recupera a posse de bola ou tem o fim do ataque da equipe adversária. Dois dados demonstram sua capacidade de proteção a meta, o primeiro é que apenas 5,07% das vezes a bola chega na zona 1 (zona mais próxima do gol que o Barcelona defende) e o segundo é que em apenas 0,15% das transição vão para o gol.

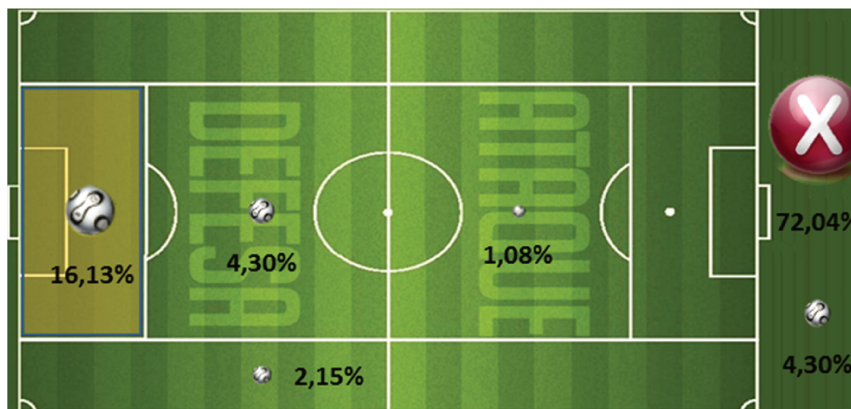


Figura 1.19: Transições da zona 1 do Barcelona no momento defensivo. Quadrante amarelo: zona de início da transição; bolas de futebol: representação gráfica da porcentagem de destino final da transição; bola vermelha: transições que tiveram como final a perda da posse; bola dourada: transições que tiveram como final sofrer o gol.

Na Figura 1.19 são apresentadas as transições da zona 1 no momento defensivo do Barcelona. Quando a bola passa pelo zona 1, em apenas 4,30% das ocasiões a bola vai para o gol. Em 72,04% das vezes a bola é recuperada pelo Barcelona ou o ataque da equipe adversária tem seu fim. Novamente, através desses dados, fica claro sua eficácia defensiva.

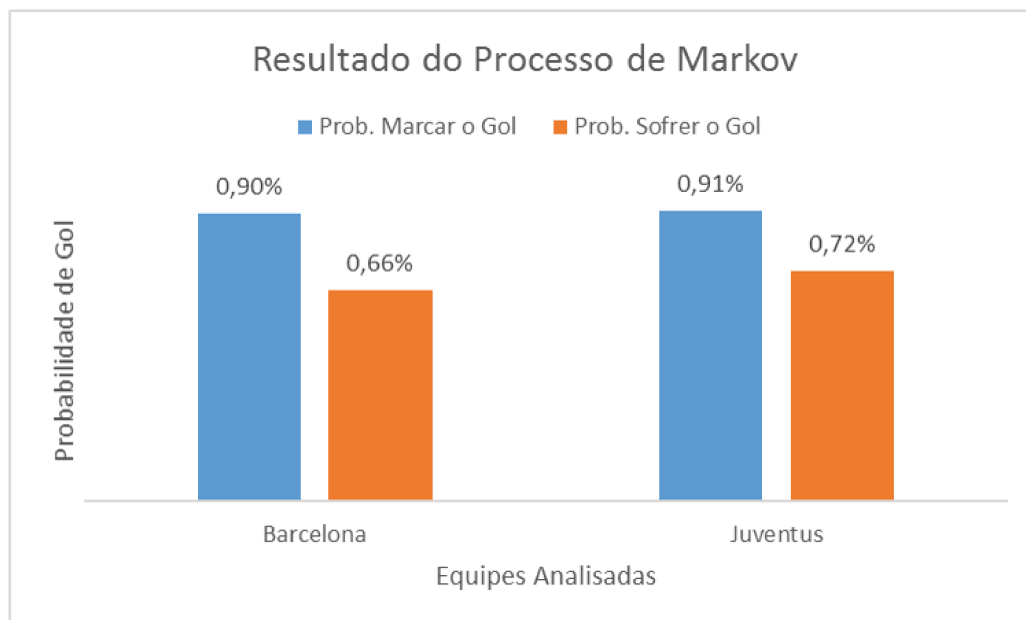


Figura 1.20: Probabilidade de gols marcados e sofridos dos finalistas da *Champions League*.

Por fim, tendo observado os dados dos momentos ofensivo e defensivo do Barcelona, pode se entender melhor seu impacto nas probabilidade de se marcar ou de sofrer o gol. Primeiramente, quando comparadas entre si, a probabilidade de marcar gols da equipe do Barcelona é maior que a de sofrer gols, o que indica que em um dado jogo da temporada

sua chances de vitória são maiores que a de derrota, isso sem levar em consideração a equipe adversária. Quando comparado com o adversário, como demonstrado na Figura 1.20, observa-se que as chances de gols marcados são similares entre as duas equipes, porém as chances de sofrerem gols é o que mais distanciam esses times. O Barcelona apresenta como probabilidade de sofrer gols de 0,66%, já a equipe da Juventus 0,72%. Apesar de pequena a diferença, seu impacto pode ser analisado quando, em 100 momentos defensivos, a chance da Juventus não sofrer gols é de 48,54% e a do Barcelona é de 51,57%.

## 1.7 Conclusão

Este capítulo apresentou uma nova abordagem para o entendimento do jogo de futebol, através das análises apresentadas sobre o comportamento das equipes, que resultam na matriz de transição, seja no resultado da probabilidade de gols marcados e sofridos originados do estados estacionários, ambos presentes no processo de Markov. Sendo assim, passamos a observar possíveis conclusões com relação aos resultados obtidos.

O primeiro ponto a se analisar recai sobre os dados referentes a Copa do Mundo 2014. Quando observados os estados estacionários do momento ofensivo de todas as equipes, nota-se que o ordenamento da maior probabilidade de realização de gols para a menor não é o mesmo que o resultado da competição. Ou seja, a Argentina, que apresenta o menor índice de realização de gols (0,77%) ficou com a segunda colocação da competição. Sendo assim, pode-se entender que o conjunto de dados para sua análise está incompleto para o entendimento pleno do jogo, tendo que incluir seu momento defensivo neste estudo em uma futura realização.

Já no segundo trabalho apresentado neste capítulo, fica claro que apenas uma análise ofensiva não satisfaz as questões apresentadas no estudo. As duas equipes, Barcelona e Juventus, apresentam probabilidades de gols próximas: 0,90% e 0,91%, respectivamente. Sendo assim, equipes com características diferentes passam a ter um resultado similar, o que indica a necessidade de um melhor entendimento do momento defensivo. Neste, a equipe do Barcelona passa a levar vantagem, 0,66% e 0,72%. Também em perspectiva futura, poderia ser realizada a integração dessas matrizes, gerando uma única matriz de jogo.

Em ambos os estudos, pode-se notar que grande parte das ações são concentradas na parte central do campo (zonas 4 e 5), tanto de defesa como de ataque. Uma possibilidade de melhoria passa em novas divisões do campo, tendo em vista quais áreas o analista passe a querer observar. Estudos de diversos modelos de divisões de campo também passam a ser necessários para o desenvolvimento do método proposto.

Em comparação com os trabalhos previamente citados, pode-se ressaltar o problema de previsão de resultados que não foram contemplados neste estudo, como: o local do jogo (Gamerman & Gamerman, 2003; Seligman, 2011; Rue & Salvesen, 2000), os níveis



das equipes observadas (Alves *et al.*, 2008) e os placares obtidos ao longo da temporada (Farias, 2008). Porém, todos estes estudos apresentam como base o resultado das partidas e não o comportamento das equipes como forma de análise. A presente pesquisa é inédita por esta abordagem voltada ao comportamento das equipes ao longo dos jogos.

Por fim, tendo em vista todas estas conclusões, pode se entender que este tipo de estudo tende a auxiliar treinadores, comissões, jornalistas ou apostadores, pois permitem uma redução nos erros no momento da tomada de decisão, facilitando assim o alcance de seus objetivos. Na mesma linha, há um caminho para se percorrer na melhora deste modelo, seja ele pela adição de novos elementos ou remodelação dos elementos já apresentados.

## 1.8 Implicações práticas

Como um dos objetivos desse estudo é a aplicação de modelos matemáticos ao cenário real do futebol, se faz necessário o entendimento desta ferramenta no cotidiano do profissional que vai poder aproveitá-la na prática. Desta forma, esta seção tem como propósito apresentar possíveis aplicações do processo de Markov no futebol, em diferentes cenários.

Dentre tais aplicações destacam-se duas delas, a primeira aplicação é a previsão de resultados em partidas de futebol. Durante este estudo, os resultados de probabilidade de gols, alinhados a outras variáveis do jogo podem fornecer indicadores robustos para este objetivo, que impacta não apenas os administradores presentes em clubes de futebol, mas também treinadores, jogadores e gestores, até jornalistas, apostadores e demais interessados, resultando assim, um indicador de desempenho interessante para o entendimento do jogo e comportamento de equipes.

A segunda aplicação é a análise do jogo aplicada à equipe, ou seja, como os dados de transição da bola em campo sugerem um perfil da equipe, é possível analisá-la e, tendo isso identificado, há a possibilidade de alterações deste perfil para melhorar o jogo e seus resultados, ou mesmo identificar os perfis de outras equipes para confrontos ou *benchmark*. Esta aplicação acaba por ter como interação os treinos e seus objetivos durante a preparação para o jogo.

Sob outro foco, a aplicação de análises táticas baseadas nestas transições podem ser utilizados por diferentes veículos midiáticos como televisão, internet ou rádio, para embasar e, por vezes, demonstrar comentários ou opiniões que os responsáveis teriam. Além da qualidade da informação, outro aspecto favorável é seu fácil entendimento e construção visual.

Essas e outras aplicações que possam derivar da aplicação do processo de Markov auxiliam no entendimento da oportunidade de estreitar a relação entre a ciência e o futebol, principalmente entre os modelos matemáticos e a tomada de decisão no futebol.

## Capítulo 2

# Problema de designação em escala de árbitros de futebol

Neste capítulo apresenta-se um modelo de Programação Linear Inteira para o problema de designação de árbitros em campeonatos de futebol. O modelo proposto tem como objetivo minimizar a distância total percorrida pela soma dos árbitros na competição. As restrições modeladas foram formuladas a fim de equilibrar o número de partidas apitadas pelos árbitros e satisfazer características da competição. O modelo foi testado com dados reais do Campeonato Brasileiro de Futebol da Série A de 2014.

Os resultados apresentados neste capítulo estão presentes no artigo “Um modelo de programação linear inteira para a otimização de recursos na designação de árbitros” Anon *et al.* (2017), publicado na revista Pesquisa Operacional para o Desenvolvimento.

### 2.1 Otimização de recursos na designação de árbitros no futebol

Ligas profissionais de futebol são grandes negócios no mundo e movimentam um volume significativo de recursos financeiros segundo Recalde *et al.* (2013). Diariamente, milhares de torcedores acompanham jornais, rádio, televisão e internet em busca de informações sobre times locais, nacionais e internacionais, além de resultados e estatísticas sobre jogos, campeonatos e jogadores Ribeiro & Urrutia (2012).

Neste esporte, a função dos árbitros envolve garantir que todas as regras sejam respeitadas e que a partida seja conduzida de forma ética, privilegiando o desenvolvimento do jogo como um espetáculo esportivo, como tem em Duarte *et al.* (2007). Neste cenário, a atuação dos árbitros é alvo frequente de grande pressão por parte dos diversos atores envolvidos (clubes, atletas, torcedores, mídia). Diversos fatores podem influenciar na atuação de um árbitro em uma partida, mas os aspectos técnicos e logísticos da profissão têm recebido maior atenção das confederações e entidades de classe. Há uma demanda

crescente pela profissionalização dos árbitros, assim como já ocorreu em outros setores do futebol.

Neste contexto a designação de árbitros para as partidas é uma decisão complexa e de grande responsabilidade, especialmente em campeonatos longos e em territórios extensos, como é o caso do Campeonato Brasileiro de Futebol da Série A. Além disso, em função de sua popularidade e do volume de negócios envolvidos, o futebol é um esporte bastante visado quando se trata de corrupção de resultados. Embora exista bastante preocupação com a lisura esportiva, escândalos como o “*Calciopoli*” no futebol italiano (Distaso *et al.*, 2012) é um exemplo onde árbitros foram envolvidos em casos controversos de resultados Alarcón *et al.* (2014).

Assim, o desenvolvimento de políticas mais apropriadas e eficientes para a escolha de árbitros é fundamental para o desenvolvimento do esporte, tanto do ponto de vista técnico quanto para um gerenciamento mais transparente das escalas de arbitragem.

No entanto, na prática a designação de árbitros na maioria dos campeonatos profissionais não evoluiu na mesma proporção que outras áreas, como por exemplo medicina esportiva, treinamento e *marketing*, que incorporaram uma cultura científica para abordar seus principais problemas. Ao contrário, não é raro de se encontrar campeonatos onde a escala de árbitros é feita sem critérios analíticos claros, dando margem a interferências subjetivas indesejáveis ao esporte.

O problema de escala ou designação dos árbitros tem recebido atenção da comunidade científica, em especial da área de pesquisa operacional, pois sua solução manual além de ineficaz é vulnerável a especulações (Kendall *et al.*, 2010; Ribeiro, 2012). O estudo de Yavuz *et al.* (2008) propôs um modelo que foi utilizado para identificar uma atribuição justa dos árbitros de futebol para a temporada do campeonato turco *Premier League* 2005-2006. Ainda neste contexto, Trick & Yildiz (2007) apresentam um modelo com intuito de minimizar a distância total percorrida pelos árbitros.

Neste trabalho, utilizamos um modelo similar ao proposto em Alarcón *et al.* (2014) para a designação de árbitros para jogos de futebol que utiliza programação linear inteira. O objetivo específico deste artigo é propor e aplicar um modelo para a designação de árbitros utilizando dados reais do Campeonato Brasileiro de 2014, com o intuito de minimizar a distância total percorrida pelos árbitros, além de evitar que o mesmo árbitro apite várias partidas de um mesmo time, ainda é desejável propor uma distribuição uniforme no número total de jogos atribuídos a cada árbitro na competição. Além destes aspectos, o modelo permite a inclusão de outras restrições que podem surgir na prática, como por exemplo um determinado árbitro não ser elegível para apitar um determinado jogo.

## 2.2 Definição do problema

O Campeonato Brasileiro de Futebol da Série A, também conhecido como Brasileirão, possui 20 equipes que se enfrentam na modalidade todos contra todos, em partidas no formato casa-fora. O Campeonato é dividido em 38 rodadas de 10 partidas, totalizando 380 jogos. Na edição de 2014 da competição foram utilizados um total de 59 árbitros, sendo que 39 árbitros apitaram menos de 8 jogos no ano. Como será apresentado na Conclusão, apenas 20 árbitros seriam suficientes para uma escala racional de recursos.

Sendo o Brasil um país de grandes dimensões geográficas, uma das preocupações que devem ser consideradas na designação dos árbitros é justamente a distância percorrida nas viagens para apitar as partidas. O gerenciamento não adequado desta característica pode aumentar os gastos financeiros com deslocamento e levar a um desgaste físico dos árbitros, afetando seu desempenho técnico no jogo.

O modelo proposto neste trabalho pretende minimizar a soma de todas as distâncias percorridas pelos árbitros para apitar as partidas. Para viabilizar a coleta de dados, consideramos que a distância percorrida por um árbitro para apitar uma partida é igual a distância, em linha reta, entre a capital do estado cujo árbitro reside e a capital do estado onde o jogo foi realizado. Esta distância é multiplicada por 2 para computar a viagem de ida e volta. No exemplo apresentado neste artigo, os dados foram coletados pelos autores utilizando a ferramenta computacional disponível no *website* <http://distanciacidades.com/>.

Com base nestas distâncias, foi possível concluir que, no Campeonato Brasileiro de Futebol da Série A de 2014, os árbitros percorreram um total de 833.516 quilômetros em viagens para apitar todas as partidas.

Outro fator que foi considerado no modelo de designação de árbitros é o número de partidas que um mesmo árbitro apita de uma dada equipe, já que a repetição de um mesmo árbitro para diversos jogos de uma equipe pode prejudicar a imparcialidade na condução das partidas, além de levantar dúvidas sobre a lisura de resultados. Desta forma, quanto menor for o número de partidas de uma mesma equipe comandadas por um dado árbitro, melhor será o processo de designação.

A Figura 2.1 apresenta o número máximo de partidas designadas à um mesmo árbitro no Campeonato Brasileiro de 2014 – Série A. Pode-se notar que apenas Chapecoense e Criciúma tiveram no máximo 2 partidas atribuídas ao mesmo árbitro, enquanto 4 equipes tiveram cinco partidas apitadas pelo mesmo árbitro.

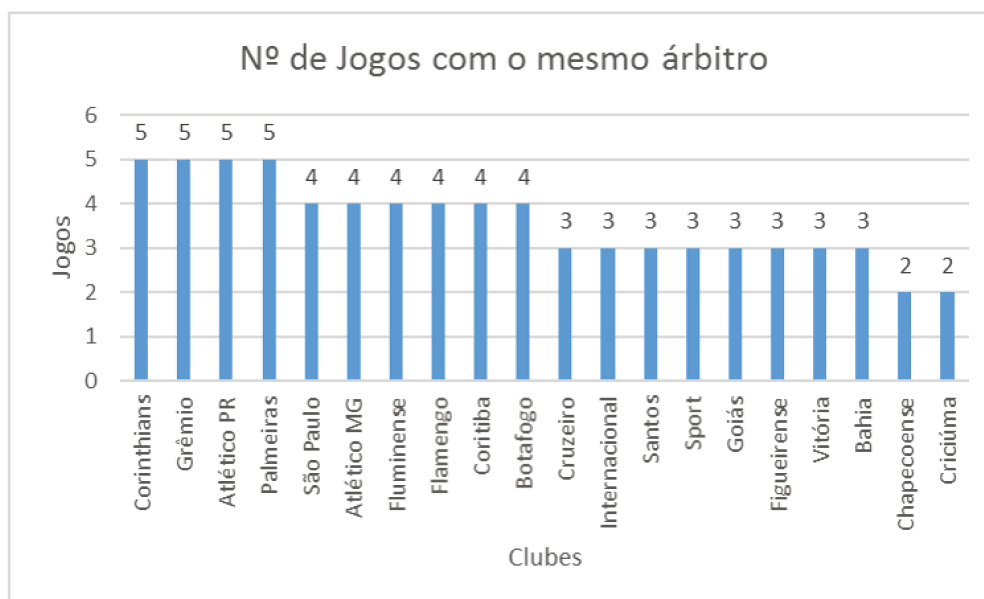


Figura 2.1: Número máximo de partidas designadas a um mesmo árbitro no Campeonato Brasileiro de 2014 – Série A.

Visando obter uma maior homogeneidade na designação considera-se também o número total de partidas atribuídas para cada árbitro no campeonato. A Figura 2.2 apresenta o número de partidas apitadas pelos 20 (dos 59) árbitros que mais atuaram no Campeonato Brasileiro de 2014. Pode-se notar que a diferença entre o árbitro que mais apitou (24: Ricardo Marques Ribeiro) e o que menos apitou (8: Felipe Gomes da Silva) é de 16 partidas. Incluindo os 59 árbitros a diferença passa a ser de 23 partidas.

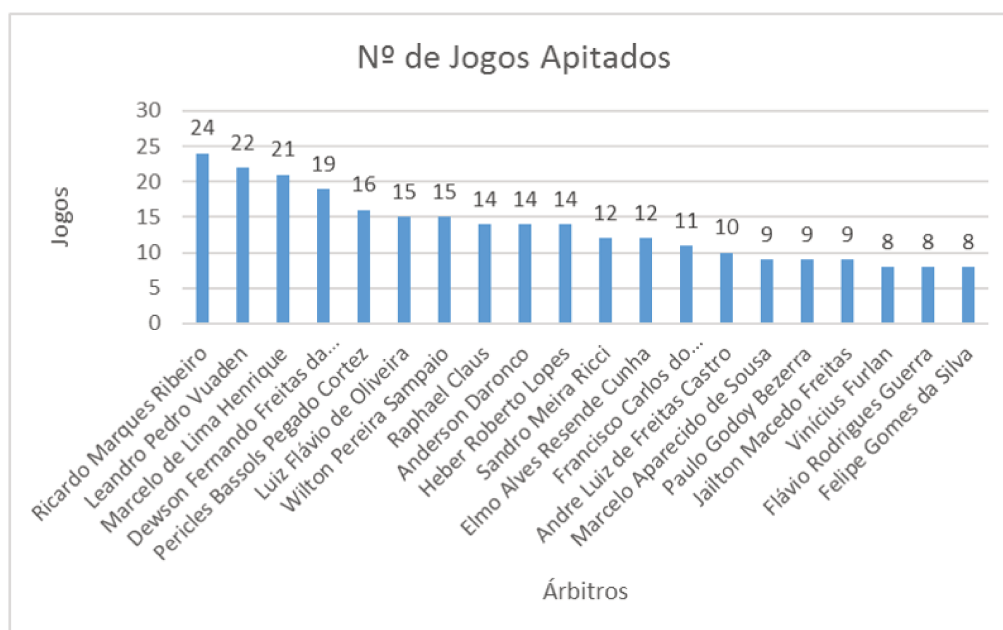


Figura 2.2: Número de partidas apitadas para cada um dos 20 árbitros que mais atuaram no Campeonato Brasileiro de 2014.

O problema de otimização considerado neste trabalho consiste em minimizar a

distância total percorrida pelos árbitros para apitar todas as partidas do Brasileirão, não permitindo que o mesmo árbitro apite mais do que 2 partidas de uma mesma equipe e que a diferença entre o número total de partidas apitadas por cada árbitro no campeonato seja no máximo 2.

Uma questão que surge naturalmente da análise do problema acima é qual é o número mínimo de árbitros necessários para obter-se uma escala factível que atenda os critérios estabelecidos? Sendo 38 jogos de cada equipe e as restrições citadas anteriormente, são necessários no mínimo 19 árbitros para atender esta restrição. Obviamente, quanto mais árbitros estiverem disponíveis, menor será a distância total a ser percorrida.

Neste trabalho, fixou-se o número de 20 árbitros para efeito de comparação com os dados reais. Sendo assim, serão considerados os 20 árbitros que mais atuaram no Campeonato Brasileiro de 2014 e a capital de seus estados de origem será definida como ponto inicial da viagem para apitar cada partida. O nome dos árbitros considerados, bem como o número de partidas que cada um apitou no Brasileirão de 2014, estão apresentados na Figura 2.2.

## 2.3 Modelo proposto

O modelo utilizado neste trabalho é baseado em Alarcón *et al.* (2014) e propõe uma adequação do problema clássico de designação para atender às características peculiares deste caso.

De forma sucinta, o objetivo é definir uma escala de árbitros para todo o Campeonato de forma a minimizar a soma das distâncias percorridas pelos árbitros em viagens para apitar os jogos, sujeito a um conjunto de restrições que garantam um número máximo de partidas de um árbitro com a mesma equipe, bem como limites mínimos e máximos para o total de jogos de um mesmo árbitro no campeonato.

Para especificação do modelo, define-se a seguir os seguintes conjuntos:

$A = \{1, 2, \dots, 20\}$ : Árbitros;

$E = \{1, 2, \dots, 20\}$ : Equipes;

$P = \{1, 2, \dots, 380\}$ : Partidas;

$R = \{1, 2, \dots, 38\}$ : Rodadas;

O modelo terá como entrada os seguintes parâmetros:

$d_{p,a}$ : distância entre o local da partida  $p$  para o local de origem do árbitro  $a$

$$\alpha_{pr} = \begin{cases} 1, & \text{se a Partida } p \text{ pertence a Rodada } r \\ 0, & \text{caso contrário} \end{cases}$$

$$\beta_{pe} = \begin{cases} 1, & \text{se a Partida } p \text{ contém a equipe } e \\ 0, & \text{caso contrário} \end{cases}$$

$Min_a$ : Número mínimo de partidas a serem apitadas por um árbitro = 18;

$Max_a$ : Número máximo de partidas a serem apitadas por um árbitro = 20;

$\gamma_{a,e}$ : Número mínimo de partidas de um árbitro com o mesmo clube = 1;

$\nu_{a,e}$ : Número máximo de partidas de um árbitro com o mesmo clube = 2;

As variáveis de decisão do modelo são as seguintes:  $\forall p \in P, e \in A$

$$x_{pa} = \begin{cases} 1, & \text{se a Partida } p \text{ é apitada pelo Árbitro } a \\ 0, & \text{caso contrário, } \forall p \in P, e \in A \end{cases}$$

Sendo assim, pode-se escrever a função objetivo que deve ser minimizada como:

$$\sum_{p \in P, a \in A} x_{p,a} * d_{p,a} \quad (1)$$

Para garantir que exatamente 1 árbitro seja designado para cada partida devemos incluir a seguinte restrição:

$$\sum_{a \in A} x_{p,a} = 1, \forall p \in P \quad (2)$$

Como cada árbitro pode ser escolhido para apitar no máximo 1 jogo por rodada, tem-se:

$$\sum_{p \in P} \alpha_{p,r} * x_{p,a} \leq 1, \forall r \in R, \forall a \in A \quad (3)$$

A restrição a seguir garante que o número total de partidas apitadas pelo mesmo árbitro esteja no intervalo pretendido.

$$18 \leq \sum_{p \in P} x_{p,a} \leq 20, \forall a \in A \quad (4)$$

Por fim, as duas últimas restrições restringem o número de jogos atribuídos à um mesmo árbitro para uma dada equipe.

$$1 \leq \sum_{p \in P} \beta_{p,e} * x_{p,a} \leq 2 \quad (5)$$

Com o objetivo de validar o modelo apresentado nesta seção e comparar a solução obtida com um cenário real, implementou-se o problema de programação proposto em linguagem MPL, utilizando o solver GLPK para resolvê-lo. Os testes foram executados em um computador Sony Vaio, com processador Intel Core i5-3210M, CPU @2.50GHz, com 6 GB de memória RAM, com sistema operacional Windows 10.

## 2.4 Resultados

Utilizando os dados reais do Campeonato Brasileiro de Futebol de 2014 – Série A e considerando apenas os árbitros listados na Figura 2.2, o modelo de programação linear inteira proposto foi resolvido de forma exata, com tempo médio de 45 minutos. Este tempo de solução é perfeitamente aceitável, visto que a programação da escala de árbitros pode ser feita (e inclusive refeita) com vários dias de antecedência.

Salienta-se que optamos por usar o *software* livre GLPK, ao invés de *softwares* comerciais clássicos (como o CPLEX ou GUROBI) para mostrar que neste caso, o uso de um software gratuito foi suficiente.

O modelo implementado encontrou o ótimo global da função objetivo, atendendo a todas as restrições, com um total de distância percorrida de 514.692Km para cobrir os 380 jogos. Este resultado representa uma redução de 38% em comparação com a distância total percorrida pelos árbitros na competição de 2014. A título de comparação, na competição de 2014 os 20 árbitros considerados neste trabalho apitaram 261 jogos (Figura 2.2) e, apenas para apitar esses jogos, percorreram 605.420Km, enquanto o resultado da otimização prevê um total de 514.692Km para cobrir os 380 jogos do campeonato. Vale ressaltar que, se considerados os 59 árbitros no modelo, a distância total poderia ser reduzida pela relaxação do espaço de busca.

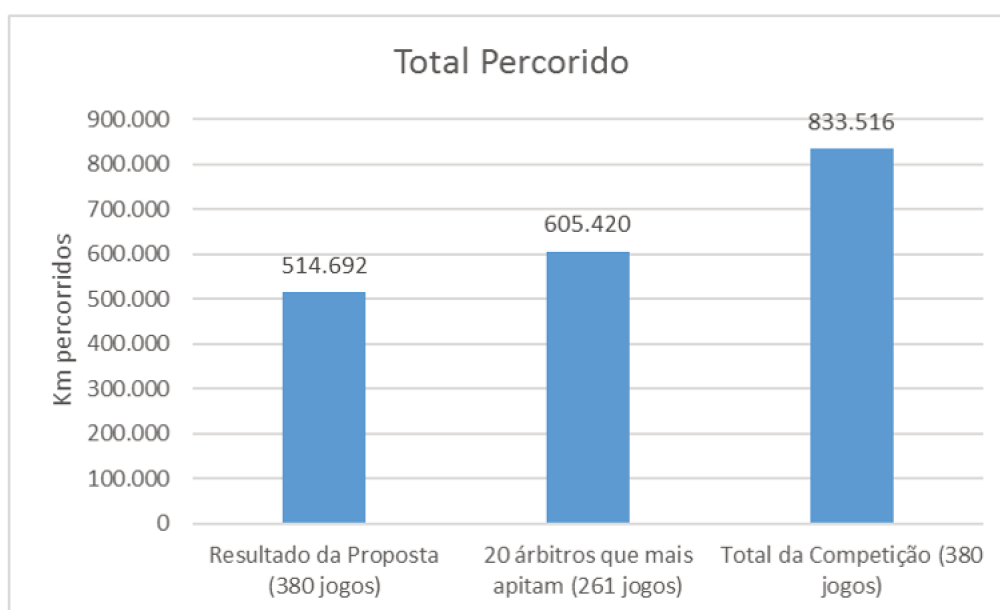


Figura 2.3: Distância total percorrida (em Km) pelos árbitros no Campeonato Brasileiro de 2014 – Série A em comparação com o resultado da otimização.

Em relação ao número total de partidas designadas para cada árbitro na competição, a Figura 2.4 mostra que todos os valores ficaram dentro do limite estabelecido de 18 a 20 partidas. Este resultado é muito mais homogêneo do que o ocorrido no campeonato de 2014, cujo intervalo variou de 8 a 24 partidas (Figura 2.2).



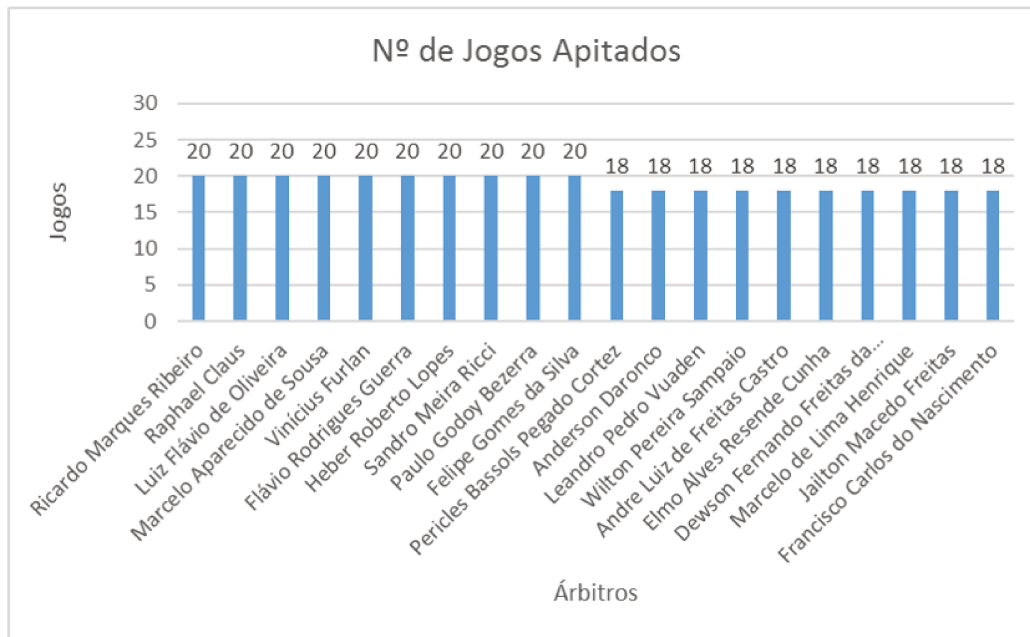


Figura 2.4: Número total de partidas atribuídas a cada árbitro pelo modelo.

Na Figura 2.5 nota-se que as restrições relacionadas ao um número mínimo e máximo de jogos de um mesmo árbitro com a mesma equipe foram atendidas. Todos os árbitros apitam no mínimo 1 partida da equipe e todas as equipes apresentam um máximo de 2 partidas com o mesmo árbitro.

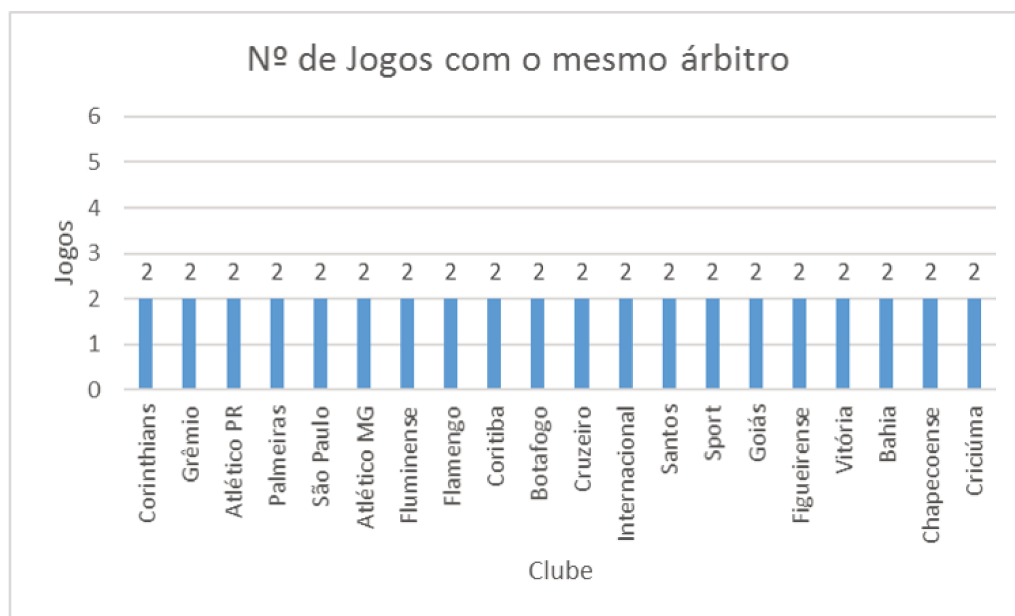


Figura 2.5: Número máximo de partidas designadas a um mesmo árbitro para uma mesma equipe.

Os resultados apresentados nesta seção consideram apenas as restrições que foram inicialmente assumidas. No entanto, o modelo proposto neste trabalho permite que outras restrições sejam incorporadas. Por exemplo se um determinado árbitro (ou conjunto de

árbitros) não puderem apitar uma determinada partida (ou conjunto de partidas) basta zerar as variáveis correspondentes, adicionando restrições do tipo  $x_{pa} = 0$ . De forma análoga, se for desejado impor um determinado árbitro para uma partida, basta atribuir  $x_{pa} = 1$ , para o caso correspondente. Outra consideração que pode ser incorporada, caso seja de interesse da organização, consiste em incluir um número mínimo,  $k$ , de jogos apitados por árbitros de um dado nível (ex: árbitros FIFA) para cada equipe. Para tanto pode se incluir uma restrição do tipo

$$k \leq \sum_{p \in F} \beta_{p,e} * x_{p,a} \quad (6),$$

onde  $F \subset P$  representa o subconjunto de árbitro pertencente a categoria que se deseja destacar.

Além disso, é possível também flexibilizar o número de jogos por árbitros, ou ainda definir uma função objetivo que seja diferente de minimizar a distância, de forma a adequar-se aos interesses da organização.

## 2.5 Conclusão

Este trabalho apresenta um modelo de programação linear inteira para designação de árbitros em Campeonatos de Futebol. A função objetivo consiste na minimização da distância total percorrida, atendendo critérios pré-definidos de homogeneidade na escala.

O modelo foi testado com dados reais do Campeonato Brasileiro de Futebol – Série A de 2014 e mostrou que, utilizando apenas 20 árbitros, é possível obter uma redução de 38% na distância total percorrida. Simultaneamente, homogeneia-se o número de partidas atribuídas a cada árbitro.

O modelo proposto permite ainda a flexibilização de parâmetros, como o número mínimo e máximo de jogos por árbitro e número de partidas no campeonato, bem como a inclusão de novas restrições a critério da organização. Por exemplo, se a comissão de arbitragem decidir fixar um determinado árbitro  $\varphi$  para uma partida  $\rho$ , basta incluir a restrição  $x_{\rho,\varphi} = 1$  e rodar o modelo novamente. A função objetivo pode ser adaptada para atender outros interesses, e não apenas a distância percorrida pelos árbitros em viagens, por exemplo, aspectos técnicos e níveis de dificuldades das partidas.

Conclui-se com isso que o modelo proposto pode ser utilizado como apoio à decisão para o atual cenário de designação dos árbitros e é escalável a novas competições e também diferentes modalidades que não foram consideradas neste estudo.

## 2.6 Implicações práticas

Com este estudo, pode-se perceber que pela utilização do método com dados reais, o modelo do problema de designação aos árbitros de futebol do Campeonato Brasileiro de

2014 torna-se uma possível aplicação prática no cotidiano de um grupo de profissionais ligados ao futebol, no caso: gestores, comissões de arbitragens e árbitros. Porém, quando observado apenas o modelo, ainda sem seu cenário de aplicação, nota-se que o mesmo tende a ter outras utilidades. Assim sendo, esta seção tem como propósito apresentar possíveis aplicações do problema de designação no futebol, em diferentes cenários.

O modelo estudado pode ter em vista com dois focos. O primeiro é sua utilização para problemas de seleção, onde a solução manual tende a ser ineficaz ou vulnerável; e segundo é em problemas de otimização de recursos e processos. Desta forma, a primeira possibilidade de aplicação pode ser encontrada em diversas situações, desde uma seleção de projetos ligados ao esporte em secretarias municipais até a aplicação em cenários de alto rendimento, como seleção de jogadores para uma competição (composição do elenco), até mesmo na escolha de atletas para uma determinada partida. Quando analisado este último caso, seleção de atletas para uma determinada partida, como exemplo nota-se que há um problema em como escolher a equipe que vai para a partida, onde o objetivo pode ser maximizar o rendimento médio da equipe, respeitando restrições como condicionamento físico, posicionamento em campo, dentre outras, para compor uma escolha.

No segundo cenário citado acima, onde se observa a otimização dos recursos e processos e tendo em vista a aplicação realizada neste estudo, nota-se como efeito secundário a minimização de custos de viagens, de estadia e com processos de seleção, sendo este último efeito é notado quando observada a mudança entre se ter uma comissão para a escolha e quando um modelo matemático passa a ser base do processo de tomada de decisão, ou seja, exigindo menos tempo e pessoas para a seleção.

Por fim, há diversas aplicações para o problema de designação no futebol, sendo que o entendimento desta ferramenta e de suas possibilidades de aplicação auxilia no desenvolvimento do esporte e da pesquisa operacional, tanto como ramo de pesquisa científica, como aplicações práticas na modalidade.

## Capítulo 3

# Aplicação de DEA para ranqueamento de eficiência no futebol

Este capítulo, apresenta a aplicação de um modelo de Análise Envoltória de Dados (DEA) com o intuito de obter a eficiência comparada entre as unidades analisadas. Outro resultado obtido foi um ranking baseado na média de pesos obtidos em cada caso da Análise Envoltória de Dados. O trabalho se baseia em dados reais referentes aos 20 clubes mais ricos do mundo em 2015 e os 23 melhores jogadores do mundo segundo o ranking FIFA. Os resultados deste capítulo compõem o artigo “Análise envoltória de dados aplicada para avaliação de jogadores de futebol” que foi submetido ao XLIX Simpósio Brasileiro de Pesquisa Operacional.

### 3.1 Introdução

Analisar a eficiência de um processo de produção é uma atividade complexa, sendo de difícil mensuração para o analista. Esta dificuldade é agravada quando se tem um cenário de múltiplos recursos e/ou múltiplos produtos. Desta forma, entender os conceitos ligados à produção com recurso e produto único (eficácia e produtividade) e recursos e produtos múltiplos (eficiência) é essencial para a análise.

A eficácia é a capacidade de produção de uma unidade tomadora de decisão. Este conceito, segundo Mello *et al.* (2005), está ligado apenas ao quanto a unidade produz e não ao quanto a mesma utiliza de recursos. Observando o cenário do futebol, se uma dada equipe determina como objetivo uma certa conquista de um campeonato e alcança este resultado, esta unidade é eficaz, independentemente de seu gasto financeiro para atingir este objetivo.

Já se analisarmos a razão entre os produtos obtidos e os recursos utilizados para a ob-

tenção do mesmo, passa-se a ter um quociente de produtividade. Este índice é dado por diferentes unidades de medida, singular para cada caso, permitindo obter comparações entre unidades. Desta forma, uma unidade tomadora de decisão é mais produtiva que outra unidade quando a mesma utiliza melhor os recursos para obter o máximo de retorno. No futebol, a relação entre os pontos obtidos em um campeonato e o gasto com o departamento de futebol é um exemplo de indicado de produtividade.

Enquanto nos critérios apresentados há um único recurso e/ou um único produto a ser analisado, a eficiência está relacionada a múltiplos recursos e produtos. Ou seja, compara as produções realizadas, dado os recursos disponíveis para a produção, sendo analisado o que poderia se ter obtido de produção com os mesmos recursos.

Como exemplo, na Figura 3.1, o recurso é representado pelo eixo X, já a produção passa a ser representada pelo eixo Y e a curva em preto indica a fronteira de eficiência. Esta fronteira indica graficamente o máximo que a unidade produz com determinado nível de recursos. O conjunto viável de produção, ou seja, as demais possibilidades de produção, que não a máxima, são representadas na região abaixo da curva de eficiência.

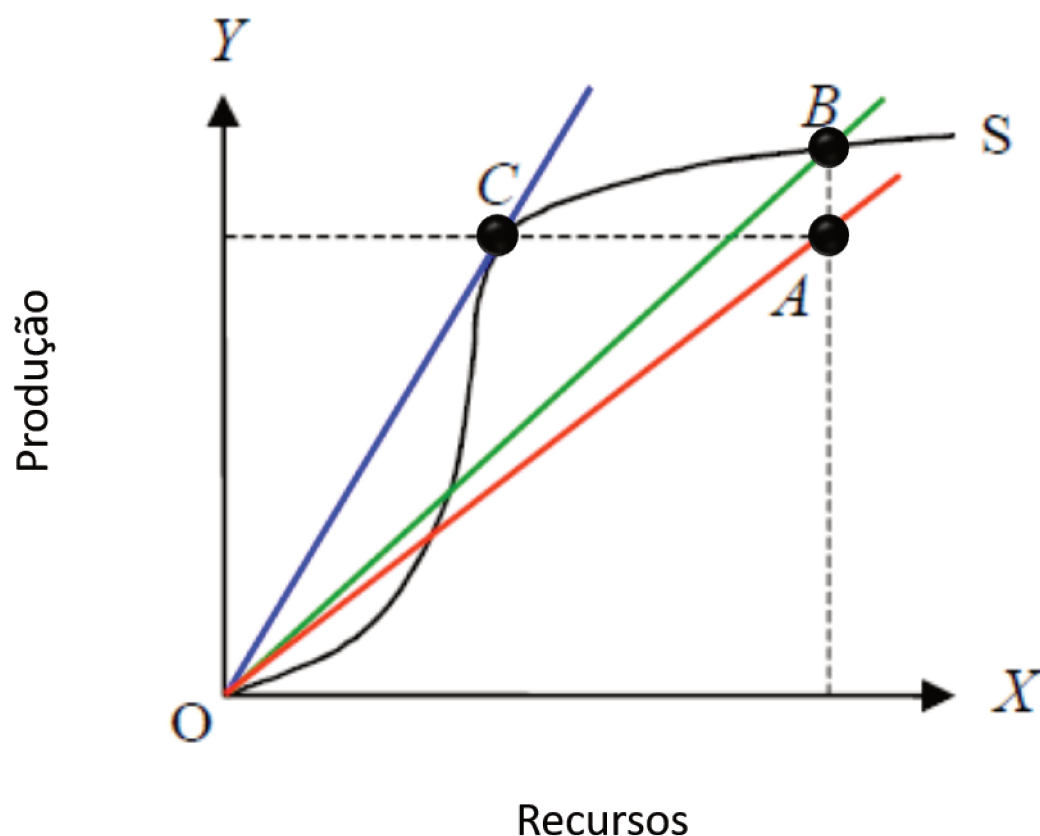


Figura 3.1: Curva de um processo de produção (Mello *et al.*, 2005).

Ainda na Figura 3.1, fica em destaque a diferença entre os conceitos de produtividade e eficiência, principalmente quando observadas as unidades B e C. A curva S representa a fronteira de eficiência. Todos os pontos sob esta curva representam unidades que são

consideradas eficientes. Observa-se que as unidades B e C são eficientes (estão sob a curva S), porém a unidade C é mais produtiva, sendo que esta conclusão se justifica quando comparado os coeficientes angulares das retas OC e OB. Já a unidade A é não eficiente (pois não está sob a curva S) e é menos produtiva que as unidades B e C.

Em situações similares a da unidade A, onde a mesma é não eficiente, Mello *et al.* (2005) destaca que há algumas possibilidades de ações para a unidade ser considerada eficiente. A primeira é a redução de recursos, mantendo os produtos constantes, e a segunda é manter os recursos e aumentar os produtos. A mescla dos dois modos também passa a ser uma opção.

### 3.2 DEA (análise envoltória de dados)

A DEA (Análise Envoltória de Dados) é uma técnica não-paramétrica que utiliza programação matemática para determinar fronteiras de eficiência entre diversas unidades tomadoras de decisão (DMUs). Este cenário é caracterizado por múltiplas DMUs, pela realização de diversos investimentos (*inputs*) e a saída de resultados (*outputs*) de diferentes características. Para Talluri (2000), o DEA é um modelo de análise de produtividade, caracterizado por abranger diversos fatores para medir as eficiências relativas de um conjunto homogêneo de unidades de tomada de decisão. Ao obter a fronteira de eficiência, pode-se estabelecer um ponto de referência para as DMUs, assim como uma avaliação do desempenho da mesma.

Este modelo matemático tem sido utilizado em diversas áreas no objetivo de avaliar o desempenho e o *benchmarking* de escolas, hospitais, agências bancárias, fábricas de produção, etc (Charnes *et al.*, 2013).

Para Gomes *et al.* (2001), o modelo matemático expressado através do DEA, tem como objetivos os seguintes pontos:

- Comparar DMUs que realizam tarefas similares com quantidades de recursos ou resultados diferentes;
- Identificar as DMUs eficientes, medindo e localizando as que não conseguem chegar a este nível. Para as unidades que são ineficientes pode se oferecer uma análise de *benchmark*;
- Levando em consideração algumas condições necessárias, a DEA pode ser útil para problemas de ordenação, como ferramenta de apoio a decisão multicritério;
- Embasar estratégias ou ações de produção que maximizem a eficiência das DMUs avaliadas. Desta forma, corrigindo as DMUs ineficientes através da determinação de alvos;
- Estabelecer taxas de substituição entre as variáveis analisadas permitindo a tomada de decisões gerenciais. O estabelecimento dessas taxas de substituição nem sempre tem solução única;

- Possibilidade de se encontrar *outliers*. Os mesmos podem representar não apenas desvios em comparação ao comportamento médio, mas possíveis *benchmarks* para a análise. Em muitos casos podem significar melhores práticas dentro do cenário analisado;

- Em modelos de fronteiras estocásticas é a necessidade de estimar a fronteira. No caso da aplicação do DEA, essa necessidade não ocorre.

O modelo original da DEA foi proposto em 1978 por Charnes, Cooper e Rhodes (Charnes *et al.*, 1978). O modelo é apresentado como CRS ou CCR (retorno constante de escala), isto é, se há uma variação nas variáveis de entrada (inputs) há também uma variação proporcional nas variáveis de saída (outputs). Segundo Biondi (2002), o modelo CCR tem como ponto central a redução de múltiplos produtos e múltiplos insumos para cada DMU num cenário 'virtual' de um único produto e um único insumo.

Ao determinar a eficiência pela otimização da divisão entre os *outputs* (soma ponderada das saídas) e inputs (soma ponderada das entradas), o modelo CCR consegue generalizar a definição de eficiência apontada por Farrell (1957). Desta forma, o modelo de programação linear permite que cada DMU escolha os pesos de cada variável da forma que a unidade analisada seja beneficiada. Como restrição, os pesos encontrados, aplicados a outras DMUs, não podem gerar uma razão superior a 1.

Porém em alguns casos esta comparação de retorno constante de escala passa a não ser justa, principalmente quando se tem DMUs que vivem cenários diferentes. Se as unidades tomadoras de decisão apresentam cenários de investimentos e retornos próximos, a utilização do CCR-DEA é recomendável, mas se o cenário for discrepante, recomenda-se a utilização do modelo VRS ou BCC-DEA (retorno variável de escala). O modelo BCC-DEA foi proposto por Banker *et al.* (1984) e tem como características ser um modelo de retorno variável de escala, procurando evitar problemas existentes em situações de competição imperfeita.

Ilustrando a diferença entre a aplicação dos modelos citados do DEA, considere um conjunto de DMUs, A, B, C, Q e D, representados na Figura 3.2 (Kumar; & Gulati, 2008). Na mesma, é considerado uma única entrada e uma única saída. Sendo assim, quando observado o modelo CCR, apenas a unidade B é eficiente, ou seja, todas as demais são ineficientes. Já no modelo BCC, as unidades A, B, C e Q são eficientes, sendo apenas a unidade D ineficiente.

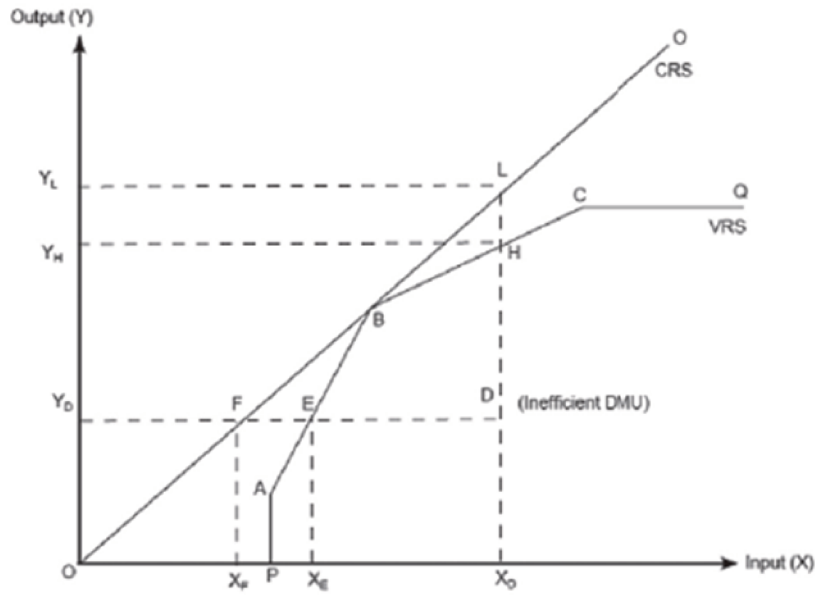


Figura 3.2: Ilustração da eficiência nos modelos BCC e CCR (Kumar; & Gulati, 2008).

Para melhor entendimento dos modelos CCR e BCC do DEA, serão apresentadas, nas seções subsequentes, as equações matemáticas presentes em cada um deles.

## Modelo CCR-DEA

Considere  $n$  DMUs, cuja eficiência deve ser analisada com relação a  $m$  variáveis de entradas, denotadas por  $x_{ik}$  e  $s$  variáveis de saída, denotadas por  $y_{rk}$ . A medida de eficiência para a  $k$ -ésima DMU é definida, no modelo de Charnes *et al.* (1978) por

$$h_k = \max \frac{\sum_{r=1}^s u_r y_{rk}}{\sum_{i=1}^m v_i x_{ik}}, \quad (7)$$

onde

$u_r$  = pesos de *outputs*  $r$ ,  $r = 1, \dots, s$ ;

$v_i$  = pesos de *inputs*  $i$ ,  $i = 1, \dots, m$ ;

$y_{rk}$  = *outputs*  $r$ ,  $k = 1, \dots, n$ ;

$x_{ik}$  = *inputs*  $i$ ,  $i = 1, \dots, m$ .

Os pesos  $v_i \geq 0$  e  $u_r \geq 0$  devem ser escolhidos para cada DMU respeitando a restrição

$$\frac{\sum_{r=1}^s u_r y_{rk}}{\sum_{i=1}^m v_i x_{ik}} \leq 1, \quad \text{para } j=1, \dots, n. \quad (8)$$

A taxa de eficiência  $h_k$  varia entre 0 e 1 e a unidade  $k$  é considerada eficiente se for possível encontrar um conjunto de pesos  $v_1, \dots, v_m$  e  $u_1, \dots, u_s$  tal que  $h_k = 1$  e  $h_j \leq 1$  para todos  $j \neq k$ . Assim, cada unidade pode formular um problema de otimização para encontrar os pesos que mais lhe favorecem na análise comparada.

O problema desse modelo é que ele é um problema de programação não linear, sendo fracionário. Para que o modelo CCR-DEA possa ser formulado como o um problema de programação linear (PL), é proposto por Charnes *et al.* (1978) a seguinte estrutura:



$$\max: h_k = \sum_{r=1}^s u_r y_{rk} \quad (9)$$

Sujeito a:

$$\sum_{r=1}^m u_r y_{rk} - \sum_{i=1}^m v_i x_{ik} \leq 0, K=1, 2, \dots, n \quad (10)$$

$$\sum_{i=1}^M v_i x_{ik} = 1 \quad (11)$$

$$u_r \geq \varepsilon > 0, v_i \geq \varepsilon > 0 \quad (12)$$

O PL acima deve ser resolvido para cada unidade e o resultado final é a determinação de hiperplanos que definem uma fronteira de eficiência. As DMUs que atingem  $h_k = 1$  estão sob a fronteira e as demais, que ficam “envelopadas”, são declaradas relativamente ineficientes.

## Modelo BCC-DEA

O modelo BCC é apresentado por Banker *et al.* (1984) com o objetivo de propor uma alternativa ao modelo CCR. No mesmo, seu diferencial é uma mudança de perspectiva de retorno constante de escala para um retorno variável de escala. Para a realização do mesmo é acrescentado uma variável,  $u_0$  que tem como função ser um fator de escala que auxilia a ajustar a escala de acordo com cada DMU. Desta forma, o problema de programação linear (PL) é apresentado com a seguinte estrutura:

$$\max: h_k = \sum_{r=1}^s u_r y_{rk} - u_0 \quad (13)$$

Sujeito a:

$$\sum_{r=1}^s u_r y_{rk} - \sum_{i=1}^m v_i x_{ik} - u_0 \leq 0, K=1, 2, \dots, n \quad (14)$$

$$\sum_{i=1}^M v_i x_{ik} = 1 \quad (15)$$

$$u_r \geq \varepsilon > 0, v_i \geq \varepsilon > 0 \quad (16)$$

$$u_0 \in \mathfrak{R}$$

onde

$u_r$  = pesos de *outputs*  $r$ ,  $r = 1, \dots, s$ ;

$v_i$  = pesos de *inputs*  $i$ ,  $i = 1, \dots, m$ ;

$y_{rk}$  = *outputs*  $r$ ,  $k = 1, \dots, n$ ;

$x_{ik}$  = *inputs*  $i$ ,  $i = 1, \dots, n$ ;

$u_0$  = fator de escala.

Assim como no modelo CCR, para a realização do modelo BCC, um PL apresentado deve ser resolvido para cada DMU estudada. As DMUs que atingem  $h_k = 1$  estão sob a fronteira e as demais, que ficam “envelopadas”, são declaradas relativamente ineficientes.

## Propostas de ranqueamento através do DEA

Nos modelos tradicionais, pode-se notar que não é possível classificar as DMUs. Esta característica provém da flexibilidade dos pesos no DEA. Desta forma, algumas unidades são eficientes na sua análise, porém quando observadas em contexto global, elas são ineficientes (Talluri, 2000). Assim, muitos autores se dedicaram a achar possíveis complementos ao método original do DEA para que ele pudesse ser usado como ranqueamento. Em um estudo de revisão sobre classificação a partir da Análise Envoltória de Dados, Adler *et al.* (2002) apresentaram 6 grupos para o ranqueamento, sendo eles:

- Métodos de classificação de eficiência cruzada (Sexton *et al.*, 1986);
- Técnicas de classificação de supereficiência (Andersen & Petersen, 1993);
- Método de classificação de referência (Torgersen *et al.*, 1996);
- Ranking com estatísticas multivariadas no contexto da DEA (Friedman & Sinuany-Stern, 1997);
- O ranking das unidades de tomada de decisão ineficientes (Bardhan *et al.*, 1996);
- DEA e métodos de decisão multicritérios (Troutt, 1995).

Neste trabalho será abordado o método de classificação por eficiência cruzada apresentado por Sexton *et al.* (1986). Como apontado por Adler *et al.* (2002), este modelo apresenta como vantagens a capacidade de ranqueamento de unidades eficientes, ranqueamento além do modelo do DEA. Outro fator favorável é sua capacidade de evitar problemas ocasionais com a inviabilidade dos métodos.

## Ranqueamento por eficiência cruzada

Dentre os métodos propostos de realização de um ranking a partir do DEA, o primeiro apresentado foi a matriz de avaliação cruzada, desenvolvida por Sexton *et al.* (1986). O método apresenta como base os resultados obtidos por uma DMU com seus pesos ótimos e com os pesos das demais unidades, ambos encontrados por intermédio da realização da programação linear. Estes resultados das pontuações são expressos na matriz de eficiência cruzada, definida na seguinte equação:

$$h_{kj} = \frac{\sum_{r=1}^s u_r y_{rk}}{\sum_{i=1}^m v_i x_{ij}}, \quad k = 1, \dots, n, \quad j = 1, \dots, n, \quad (17)$$

onde:

$h_{kj}$  = a pontuação atribuída à unidade  $j$  na com os pesos ótimos da unidade  $k$ ;

$u_r$  = pesos de *outputs*  $r$ ,  $r = 1, \dots, s$ ;

$v_i$  = pesos de *inputs*  $i$ ,  $i = 1, \dots, m$ ;

$y_{rk}$  = *outputs*  $r$ ,  $k = 1, \dots, n$ ;

$x_{ik}$  = *inputs*  $i$ ,  $i = 1, \dots, m$ .

Com esta formulação, a diagonal da matriz  $H = [h_{kj}]$ , representa a eficiência obtida para a própria DMU, ou seja,  $h_{11}$  é a eficiência da unidade 1, com o peso ótimo para si mesma. Quando observada a matriz de maneira total, nota-se que  $0 \leq h_{kj} \leq 1$ . Tal

restrição implica na escolha de apenas um modelo do DEA, o CCR. Ao adicionar o fator de escala, o modelo BCC permite que com o pesos das demais DMUs, a unidade apresente índices inferiores a 0.

Para a obtenção final do índice de eficiência cruzada, Green *et al.* (1996) apresentam diferentes maneiras para sua realização. Dentre elas: a aplicação da média, da mediana, do mínimo ou da variância. Neste estudo foi escolhido a aplicação da média, pois, desta forma, o resultado final da classificação passa a representar o comportamento médio das unidades em todas as condições obtidas através da determinação dos pesos, incluindo os que melhor representam a unidade analisada. Esta média, é representada por

$$h_k = \sum_{j=1}^n h_{kj}/n. \quad (18)$$

### 3.3 Propostas do estudo

Estudos aplicando modelos de DEA no futebol vem crescendo, sendo que diversos deles procuram analisar a eficiência de diferentes competições pelo mundo, seja analisando jogadores ou clubes. Como exemplos podemos destacar a aplicação do DEA para a análise de eficiência de clubes na *Bundesliga* (Alemanha) com o estudo de Dieter *et al.* (2004), na *La Liga* (Espanha) com os trabalhos de Gonzalez-Gomez & Picazo-Tadeo (2010), García-Sánchez (2007) e Barros & Garcia-del Barrio (2011), na *Premier League* (Inglaterra) com Guzmán & Morrow (2007) e Haas (2003) e no Campeonato Brasileiro com Pestana Barros *et al.* (2010). Já a aplicação para a análise de eficiência dos jogadores obtem destaque com o estudo na *Bundesliga* (Alemanha) realizado por Tiedemann *et al.* (2011).

Neste trabalho são propostas duas aplicações de DEA em futebol. A primeira é a análise de eficiência dos 20 clubes mais ricos do mundo no ano de 2015. Para este caso, os dados de entrada para a realização do modelo são referentes ao gasto de transferência do clube na temporada e seus gastos em salários com os jogadores. Já como dados de saída, são considerados o aproveitamento dos clubes nas disputas de suas competições nacionais, separados em ligas e copas, competições internacionais e, por fim, o saldo financeiro do clube, composto pela receita gerada menos todos os gastos realizados. Já o segundo caso é composto pela análise de eficiência dos 23 melhores jogadores do mundo em 2015. Para esta análise são considerados como dados de entrada o salário de cada atleta e seu gasto de transferência dividido pelo número de temporadas que o jogador se encontra no clube. Como dados de saída são considerados: o tempo em campo, porcentagem de pontos obtidos e saldo de gols (todos referentes aos jogos realizados em 2015), além da diferença entre o valor do jogador no mercado no momento de sua chegada ao clube e o final do ano de 2015.

Estes estudo têm como objetivo a aplicação dos modelos CCR e BCC para a obtenção das eficiências dos clubes e a comparação dos resultados entre ambos os modelos.

Como objetivo secundários, apresenta-se também o ranqueamento das DMUs de cada caso através de realização do modelo de eficiência cruzada proposto por Sexton *et al.* (1986).

Neste estudo todos os dados foram normalizados através da seguinte equação:

$$Nx_{ij} = (x_{ij} - \min x_j) / (\max x_j - \min x_j)$$

As Tabelas presentes na Figura 3.3 e na Figura 3.4 apresentam os dados utilizados e normalizados dos estudos deste capítulo.

Dados dos 23 melhores jogadores do mundo							
#	JOGADOR	SALÁRIO	TRANSFERÊNCIA	DELTA DE VALOR	TEMPO	PONTOS	SALDO
1	Messi	0,89	0,00	1,00	0,97	0,96	0,92
2	C. Ronaldo	1,00	0,19	0,43	0,98	0,83	0,87
3	Neymar	0,52	0,56	0,50	1,00	1,00	1,00
4	Aguero	0,76	0,16	0,29	0,60	0,64	0,29
5	Benzema	0,52	0,09	0,29	0,61	0,80	0,53
6	Hazard	0,69	0,18	0,36	0,90	0,61	0,20
7	Iniesta	0,52	0,00	0,39	0,70	0,91	0,59
8	Lewandowski	0,56	0,56	0,29	0,90	0,89	0,87
9	Muller	0,52	0,00	0,68	0,82	0,89	0,89
10	Pogba	0,59	0,02	0,49	0,83	0,80	0,44
11	Robben	0,56	0,11	0,14	0,34	0,91	0,54
12	Alexis Sanchez	0,45	0,33	0,26	0,79	0,83	0,42
13	Yaya Toure	0,76	0,07	0,15	0,66	0,78	0,37
14	Bale	0,89	0,72	0,14	0,69	0,76	0,49
15	De Bruyne	0,52	1,00	0,35	0,96	0,79	0,50
16	Ibrahimovic	0,73	0,21	0,00	0,81	0,85	0,77
17	Kroos	0,54	0,47	0,18	0,89	0,81	0,58
18	Mascherano	0,45	0,12	0,07	0,93	0,96	0,90
19	Neuer	0,52	0,12	0,25	0,95	0,86	0,73
20	Rakitic	0,26	0,22	0,21	0,83	0,98	0,93
21	James Rodriguez	0,42	0,67	0,29	0,50	0,83	0,53
22	Luis Suarez	0,69	0,58	0,36	1,00	0,98	0,97
23	Arturo Vidal	0,52	0,93	0,18	0,83	0,92	0,69

Figura 3.3: Dados normalizados de eficiência dos 23 melhores jogadores do mundo em 2015.

Dados dos 20 clubes mais ricos do mundo							
#	CLUBE	SALÁRIO	TRANSFERÊNCIA	LIGA	COPA	CONTINENTAL	RECEITA
1	Real Madrid	0,95	0,19	0,98	0,26	0,79	0,57
2	Barcelona	0,89	0,61	1,00	1,00	1,00	0,93
3	Man. United	0,88	1,00	0,74	0,28	0,00	0,38
4	PSG	1,00	0,36	0,88	1,00	0,45	0,44
5	Bayern Munich	0,84	0,06	0,94	0,61	0,76	0,27
6	Man. City	0,95	0,47	0,84	0,18	0,24	0,72
7	Arsenal	0,77	0,63	0,80	0,50	0,48	0,93
8	Chelsea	0,82	0,00	0,93	0,58	0,48	0,61
9	Liverpool	0,66	0,40	0,66	0,58	0,21	1,00
10	Juventus	0,54	0,15	0,93	0,80	0,73	0,70
11	Borussia Dortmund	0,44	0,45	0,55	0,72	0,39	0,51
12	Tottenham	0,53	0,06	0,68	0,42	0,33	0,68
13	Schalke 04	0,47	0,03	0,57	0,00	0,33	0,41
14	Milan	0,45	0,00	0,55	0,20	0,00	0,00
15	Atlético Madrid	0,33	0,25	0,83	0,30	0,52	0,50
16	Roma	0,42	0,42	0,74	0,20	0,26	0,76
17	Newcastle	0,32	0,19	0,41	0,25	0,00	0,47
18	Everton	0,31	0,30	0,50	0,06	0,53	0,52
19	Internazionale	0,39	0,02	0,59	0,20	0,42	0,39
20	West Ham	0,27	0,25	0,50	0,17	0,00	0,54

Figura 3.4: Dados normalizados de eficiência dos 20 clubes mais ricos em 2015.

### 3.4 Resultados

Após a coleta dos dados citados para cada uma das pesquisas e a normalização dos mesmos, passa-se a apresentar os resultados obtidos na aplicação dos modelo CCR e BCC do DEA. Em ambos os casos, o objetivo foi avaliar cada unidade de forma a entender sua eficiência, compreendendo a influência da aplicação dos diferentes tipos de modelos na análise de eficiência. Posteriormente, passou-se a realizar um ranqueamento das unidades aplicando o modelo de eficiência cruzada a partir dos resultados obtidos no CCR.

#### Eficiência dos 20 clubes mais ricos do mundo em 2015

Na Figura 3.5, se encontram os resultados referentes à análise de eficiência dos 20 clubes mais ricos do mundo em 2015. Nela encontram-se os resultados de ambos os modelos do DEA: CCR e BCC. Desta forma, nota-se que no modelo CCR, 9 das 20 unidades analisadas são eficientes para si mesma. Já no modelo BCC o número de unidades eficientes são de 16. Esta diferença apresentada nos resultados é gerada justamente pela implementação do fator de escala incluso no modelo BCC.

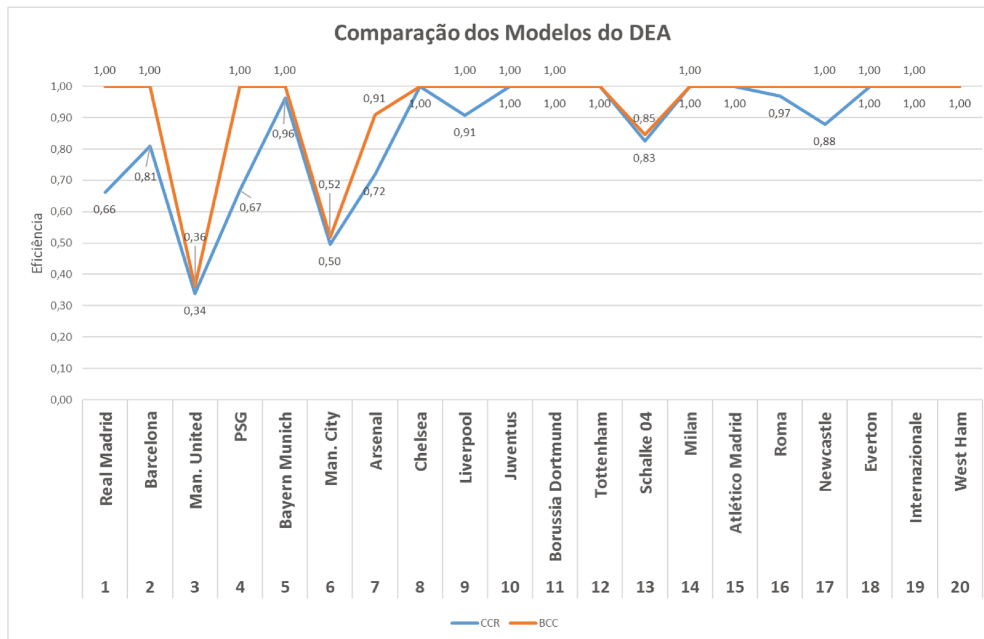


Figura 3.5: Resultados CCR e BCC para eficiência dos 20 clubes mais ricos em 2015.

Após a obtenção dos resultados da eficiência do modelo CCR, o ranking de eficiência foi obtido através do modelo de classificação de eficiências cruzadas (Sexton *et al.*, 1986). Este ranqueamento é apresentado na Figura 3.6, onde pode-se perceber que a melhor DMU classificada é a equipe da Juventus, com média de 0,83 nas análises realizadas. Já a pior equipe avaliada é o Manchester United, com apenas 0,15.

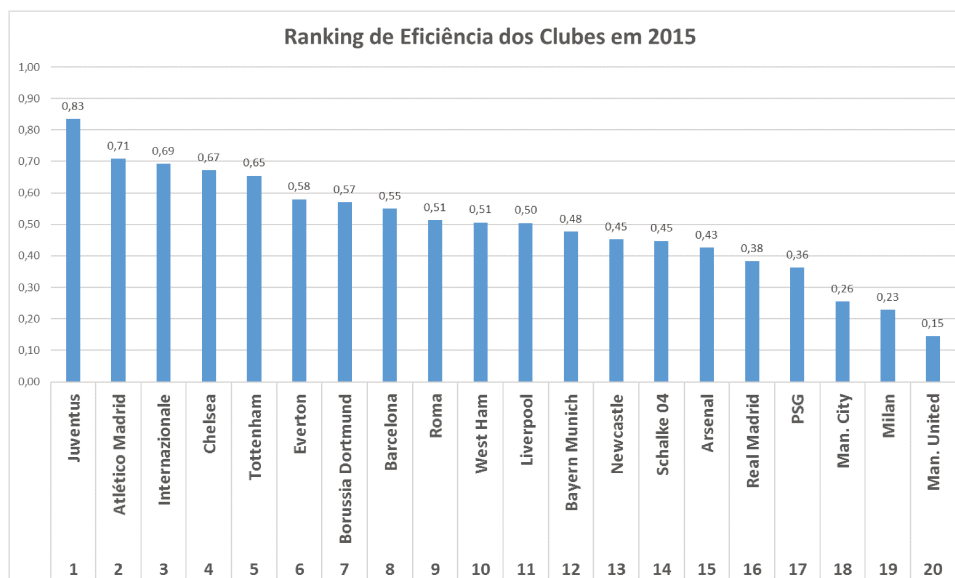


Figura 3.6: Ranking de Eficiência por *Cross Efficiency* dos Clubes.

## Eficiência dos 23 melhores jogadores do mundo em 2015

Após a aplicação dos modelos de DEA citados (CCR e BCC), verifica-se seus resultados na Figura 3.7, onde os resultados de ambos os modelos são referentes a análise de eficiência dos 23 melhores jogadores do mundo em 2015. Nesta análise, pode-se concluir que apenas 3 unidades entre as 23 analisadas são eficientes no modelo CCR e 7 unidades no modelo BCC. Assim como na análise anterior, a mudança do modelo influencia para que mais DMUs cheguem ao nível de eficiência 1.

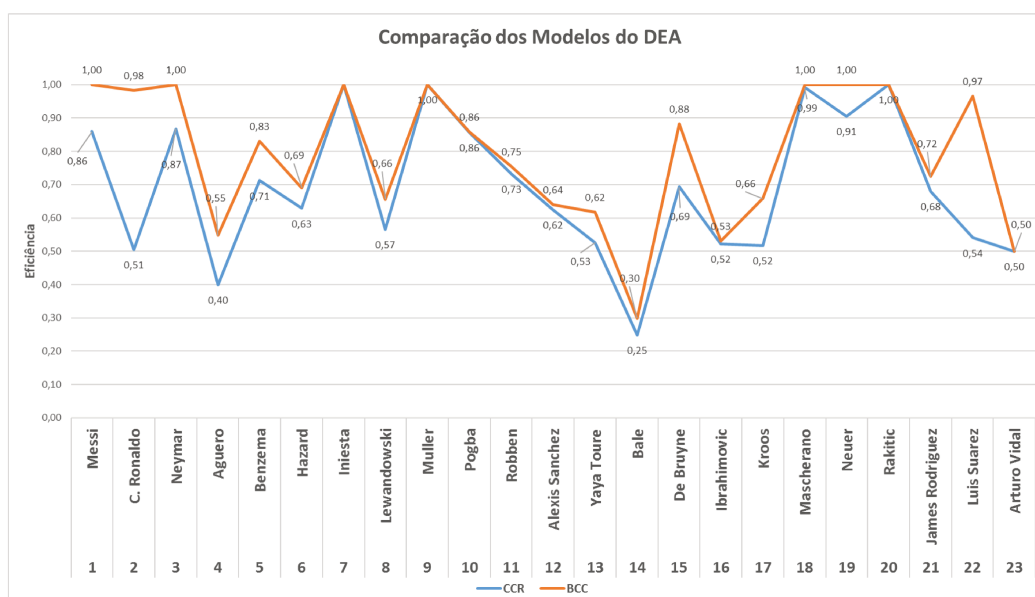


Figura 3.7: Resultados CCR e BCC para eficiência dos 23 melhores jogadores do mundo em 2015.

Com os resultados obtidos, aplicou-se o modelo de ranqueamento proposto por Sexton *et al.* (1986). Este modelo de eficiências cruzadas aponta que a melhor unidade analisada, no caso jogador, é Muller, com média de 0,94 de eficiência em todos os cenários encontrados. A pior DMU tem como média obtida 0,20 e refere-se ao jogador Bale.

Tendo em vista que o estudo tem como base a lista da FIFA de 23 melhores jogadores, uma importante unidade a ser analisada é o vencedor do prêmio. O melhor jogador do mundo eleito, Messi, apresentou como resultados de sua análise no CCR 0,86 e no BCC 1,00. Já na média de desempenho em todos os cenários ele apresenta 0,68.

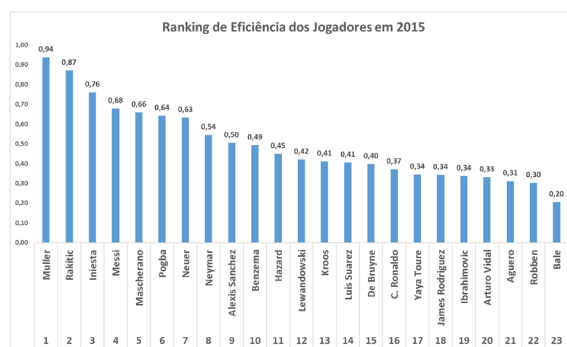


Figura 3.8: Ranking de Eficiência por *Cross Efficiency* dos jogadores.

### 3.5 Conclusão

Com base nos resultados apresentados nos dois cases, pode-se chegar a algumas considerações a respeito da análise de eficiência aplicada ao futebol. A primeira conclusão está relacionada à escolha entre os dois modelos apresentados no estudo, o CCR e BCC. Devido à sua caracterização, o modelo BCC apresenta um retorno variável de escala, podendo assim observar que em ambas as situações, sua aplicação fez com que houvesse um aumento no número de unidade eficientes, chegando a 80% das mesmas na análise de eficiência dos 20 clubes mais ricos do mundo em 2015.

Sendo assim, o modelo BCC passa a ser um bom indicador de não eficiência, pois, sendo ele benevolente com as unidades analisadas, a unidade que não alcance índice igual a 1 passa a ter um forte indicativo de não eficiente, mesmo no melhor cenário possível entre os modelos. Já o modelo CCR, apresenta uma maior rigidez, possibilitando uma análise de maior distinção entre unidades eficientes e não eficientes, além de servir como base para a aplicação de um ranqueamento entre as unidades. Uma segunda conclusão relacionada aos resultados é obtida na análise dos ranqueamentos realizado nestes estudos. Quando analisado o segundo case, observa-se que a seleção dos 23 melhores jogadores do mundo para a elaboração da pesquisa apresenta como dados de retorno uma tendência de que não haja grandes níveis de distinção entre as unidades (jogadores) em sua performance esportiva. Porém, percebe-se que o nível de investimento passa a ser muito discrepante quando analisados. O mesmo ocorre quando observado o caso 1. Mesmo levando em consideração que alguns clubes não obtiveram resultados expressivos na temporada, algumas equipes destacam-se pelo baixo investimento e alto retorno dentro deste cenário.

De maneira abrangente, podemos concluir que os clubes de futebol tendem a superestimar seus investimentos em relação aos resultados, seja no investimento total na formação do elenco, seja no investimento individual no atleta. Ou seja, investem altas quantias de dinheiro e em contrapartida, recebem retornos menores que os necessários para alcançar altas índices de eficiência. Isto reflete, analisando os dados, que os altos investimentos exigem retornos que por vezes são inalcançáveis. Desta forma, se torna necessário uma



reflexão maior dos tomadores de decisão dos clubes de como investir e, principalmente, quanto vale cada retorno, dentro ou fora de campo.

### 3.6 Implicações práticas

Quando observados ambos os estudos de análise envoltória de dados aplicada ao futebol, seja no caso dos jogadores, seja no caso dos clubes, ter o entendimento do uso desta ferramenta no cotidiano das instituições envolvidas com este esporte se torna fundamental para a integração entre o modelo matemático utilizado e o cenário de possível aplicação. Isto posto, esta seção tem como finalidade indicar possíveis aplicações da análise envoltória de dados no futebol, em diferentes contextos.

Entre as possibilidades de aplicações do modelo, passam-se a destacar dois cenários. O primeiro está relacionado à aplicação do modelo clássico da análise envoltória de dados, sem a presença de métodos de ranqueamento. O mesmo passa a ter grande usabilidade quanto ao entendimento da eficiência ou não eficiência de uma unidade analisada (podendo ser considerada como uma unidade as federações, os clubes, ou demais instituições relacionadas ao futebol ou até mesmo projetos, atletas ou treinadores quando se relaciona o investimento individual e seu retorno). Como consequência da aplicação da análise envoltória de dados, podemos destacar a obtenção de uma análise de *benchmark* para as unidades não eficientes, comparação entre diferentes unidades que realizam ações similares e entendimento do cenário que se encontram. Como exemplo, quando se observa o estudo dos 20 clubes mais ricos do mundo pela ótica de um clube, no caso, Manchester United, nota-se que a equipe é não eficiente e que equipes com menos recursos alcançaram resultados mais significativos. Sendo assim, esta equipe obteria através desse estudo uma análise de eficiência, a possibilidade de um *benchmark* com equipes similares em recursos e ainda poderia identificar quais clubes e modelos de gestão estudar para melhorar seus índices.

A segunda possibilidade é a aplicação da análise envoltória de dados para auxílio na tomada de decisão, podendo assim ser utilizada em conjunto com outros métodos de ranqueamento. Como ganho possível com esta possibilidade de aplicação, passa-se a observar o melhor entendimento de mercado, identificação de novas estratégias, além do estabelecimento de novas metas a serem alcançadas. Ainda no exemplo anterior, se analisássemos apenas o Manchester United, pode-se notar que ele obteve o pior índice no ranqueamento. Sendo assim, medidas para a realização de cortes no orçamento e/ou aumento de rendimento seriam necessárias. Em uma segunda análise, pode-se indicar que equipes como a Juventus e o Atlético de Madrid podem ser cases para o clube, readequando assim seu papel no mercado.

Desta maneira, nota-se que a aplicação da análise envoltória de dados no cotidiano dos profissionais do futebol serve como um apoio para as decisões e avaliações de diversos

cenários, resultando assim em uma ferramenta valiosa para o esporte.

## Considerações finais

O presente estudo teve como objetivo principal a apresentação de modelos matemáticos aplicados a cenários do futebol. Esta aplicação inclui a programação linear, a análise envoltória de dados e o processo de Markov, sendo que as mesmas são aplicadas para diversos agentes, como árbitros, clubes e jogadores. A possibilidade de uso de qualquer modelo pode ser ainda mais diversificada e abre perspectivas de diferentes aplicações deste trabalho.

A primeira delas é que há um grande campo de pesquisa a ser estruturado entre pesquisa operacional e esportes, de maneira geral. Seguindo as aplicações neste estudo, a aplicação do processo de Markov pode contemplar comissões técnicas, desde o treinador ao analista de desempenho, também como jornalistas que passem a comentar os jogos, tanto pela obtenção de indicadores de desempenho, como a probabilidade de gols, quanto pela possibilidade de um novo olhar sobre o jogo. Já em relação à programação linear voltada para o problema de designação, pode-se notar sua aplicação para a escala de árbitros, assim como para a determinação de calendários eficazes nas competições e controle de escalação de jogadores. Por fim, quando observado o modelo DEA, sua aplicação vai desde a análise de eficiência dos rendimentos dos clubes, federações e jogadores, como também na avaliação de estratégias realizadas e identificação de cases.

Outra conclusão que se obtém é que estes modelos matemáticos, aplicados de forma correta, podem ser base para tomadas de decisões de seus agentes. Atualmente, muitos são os cenários em que o tomador de decisão tem a sua disposição, grande volume de informações e que o mesmo não consegue compreender sua importância ou selecionar quais são os dados relevantes. Facilitar o papel do tomador de decisão, oferecendo a ele modelos matemáticos é de essencial importância para a otimização de esforços e melhorias do rendimento próprio, de seu clube ou instituição.

Por fim, dentre as considerações, podemos notar que a otimização de processos de tomadas de decisão presentes em confederações, federações, clubes e outros interessados é de fundamental importância para a evolução do próprio esporte. Se passarmos a considerar que a aplicação destes modelos ligados à Pesquisa Operacional à melhora financeira dos clubes, ou à melhora na seleção de atletas (categorias de base ou profissionais), à evolução na prática e à melhora do espetáculo pode ser significativa.

Desta forma, podemos concluir que o cenário de aplicação de pesquisa operacional ao

futebol é inicial, porém promissor. Desta forma, é necessário que haja um aprimoramento nas aplicações seguintes, mantendo assim, a possibilidade de novas linhas de pesquisa que permeiem esporte e pesquisa operacional.

# Referências Bibliográficas

- Adler, Nicole, Friedman, Lea, & Sinuany-Stern, Zilla. 2002. Review of ranking methods in the data envelopment analysis context. *European journal of operational research*, **140**(2), 249–265.
- Alarcón, Fernando, Durán, Guillermo, & Guajardo, Mario. 2014. Referee assignment in the Chilean football league using integer programming and patterns. *International Transactions in Operational Research*, **21**(3), 415–438.
- Alves, Alessandro Martins, Ramos, Thiago Graça, de Mello, João Carlos C B Soares, & Sant’Anna, Annibal Parracho. 2008. Uma proposta de previsão de resultados para o campeonato brasileiro através do modelo logito. *Anais do SPOLM*.
- Andersen, Per, & Petersen, Niels Christian. 1993. A procedure for ranking efficient units in data envelopment analysis. *Management science*, **39**(10), 1261–1264.
- Anon, Iago Cambre, Scaglia, Alcides José, & Torezzan, Cristiano. 2017. Um modelo de programação linear inteira para a otimização de recursos na designação de árbitros. **9**, 1–8.
- Banker, R. D., Charnes, A., & Cooper, W.W. 1984. *Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis*.
- Bardhan, I, Bowlin, W F, Cooper, W W, & Sueyoshi, T. 1996. Models for efficiency dominance in data envelopment analysis. *Part*, **1**, 322–332.
- Barros, Carlos Pestana, & Garcia-del Barrio, Pedro. 2011. Productivity drivers and market dynamics in the Spanish first division football league. *Journal of Productivity Analysis*, **35**(1), 5–13.
- Bergo, F P G, Anido, R, Barros, R M L, & Cunha Sa, Freire J B. 1998. Software para análise topológica de ações no futebol. *Anais do Simpósio Internacional de Ciências do Esporte, São Caetano do Sul, SP: CELAFISCS*, **21**, 90.
- Biondi, L N. 2002. *Neuro-DEA: nova metodologia para determinação da eficiência relativa de Unidades Tomadoras de Decisão. 2001*. Ph.D. thesis, Tese (Doutorado em Engenharia de Produção)-Universidade Federal do Rio de Janeiro, Rio de Janeiro.

- Charnes, A., Cooper, W. W., & Rhodes, E. 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research*, **2**(6), 429–444.
- Charnes, Abraham, Cooper, William W, Lewin, Arie Y, & Seiford, Lawrence M. 2013. *Data envelopment analysis: Theory, methodology, and applications*. Springer Science Business Media.
- Dieter, Haas, Martin, G Kocher, & Matthias, Sutter. 2004. Measuring efficiency of german football teams by data envelopment analysis\*. *Central European Journal of Operations Research*, **12**(3), 251.
- Distaso, Walter, Leonida, Leone, Maimone Ansaldo Patti, Dario, & Navarra, Pietro. 2012. Corruption and Referee Bias in Football: The Case of Calciopoli. *SSRN Electronic Journal*.
- Duarte, Alexandre R, Ribeiro, Celso C, Urrutia, Sebastián, & Haeusler, Edward H. 2007. *Referee assignment in sports leagues*. Springer. Pages 158–173.
- Farias, Fábio Figueiredo. 2008. Análise e previsão de Resultados de Partidas de Futebol. *Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro*.
- Farrell, Michael James. 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, **120**(3), 253–290.
- Franks, Ian M, & Goodman, David. 1986. A systematic approach to analysing sports performance. *Journal of Sports Sciences*, **4**(1), 49–59.
- Friedman, Lea, & Sinuany-Stern, Zilla. 1997. Scaling units via the canonical correlation analysis in the DEA context. *European Journal of Operational Research*, **100**(3), 629–637.
- Gamerman, Dani, & Gamerman, Dani. 2003. Previsão de partidas de futebol usando modelos dinâmicos.
- García-Sánchez, I. M. 2007. Efficiency and effectiveness of Spanish football teams: A three-stage-DEA approach. *Central European Journal of Operations Research*, **15**(1), 21–45.
- Garganta, Júlio. 2001. A análise da performance nos jogos desportivos. Revisão acerca da análise do jogo. *Revista Portuguesa de Ciências do Desporto*, **1**(1), 57–64.
- Gomes, Eliane Gonçalves, de Mello, JCCB, Serapião, B P, Lins, M P E, Biondi, L N, Setti, J R A, & Lima Júnior, O F. 2001. Avaliação de Eficiência de Companhias Aéreas Brasileiras: Uma Abordagem por Análise de Envoltória de Dados. *Panorama Nacional da Pesquisa em Transportes*, **2**, 125–133.

- Gomes, João Carlos Leitão. 2015. *Sistema inteligente de apoio à decisão de apostas em jogos de futebol*. Ph.D. thesis.
- Gonzalez-Gomez, F., & Picazo-Tadeo, a. J. 2010. Can We Be Satisfied With Our Football Team? Evidence From Spanish Professional Football. *Journal of Sports Economics*, **11**(4), 418–442.
- Green, Rodney H., Doyle, John R., & Cook, Wade D. 1996. Preference voting and project ranking using DEA and cross-evaluation. *European Journal of Operational Research*, **90**(3), 461–472.
- Guzmán, Isidoro, & Morrow, Stephen. 2007. Measuring efficiency and productivity in professional football teams: Evidence from the English Premier League. *Central European Journal of Operations Research*, **15**(4), 309–328.
- Haas, Dieter J. 2003. Productive efficiency of english football teams - A data envelopment analysis approach. *Managerial and Decision Economics*, **24**(5), 403–410.
- Hughes, Mike. 1996. Notational analysis. *Science and soccer*, **343**, 361.
- Kendall, Graham, Knust, Sigrid, Ribeiro, Celso C., & Urrutia, Sebastian. 2010. *Scheduling in sports: An annotated bibliography*.
- Kumar, Sunil, & Gulati, Rachita. 2008. An examination of technical, pure technical and scale efficiencies in GCC banking. *American J. of Finance and Accounting*, **1**(2), 152.
- Mello, João Carlos Correia Baptista Soares De, Meza, Lidia Angulo, Gomes, Eliane Gonçalves, & Biondi Neto, Luiz. 2005. Curso de Análise de Envoltória de Dados. *XXXVII Simposio Brasileiro de Pesquisa Operacional*, 2520–2547.
- Pestana Barros, C., Assaf, a., & Sa-Earp, F. 2010. Brazilian Football League Technical Efficiency: A Simar and Wilson Approach. *Journal of Sports Economics*, **11**(6), 641–651.
- Recalde, D., Torres, R., & Vaca, P. 2013. Scheduling the professional Ecuadorian football league by integer programming. *Computers and Operations Research*, **40**(10), 2478–2484.
- Ribeiro, Celso C. 2012. Sports scheduling: Problems and applications. *International Transactions in Operational Research*, **19**(1-2), 201–226.
- Ribeiro, Celso C, & Urrutia, Sebastián. 2004. OR on the ball: Applications in sports scheduling and management. *In: OR/MS Today*. Citeseer.

- Ribeiro, Celso C., & Urrutia, Sebastián. 2012. Scheduling the brazilian soccer tournament: Solution approach and practice. *Interfaces*, **42**(3), 260–272.
- Rue, Havard, & Salvesen, Oyvind. 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.
- Seligman, Eduardo. 2011. Comparação entre os Campeonatos de futebol Brasileiro e Italiano utilizando Cadeias de Markov. 53.
- Severino, Tiago Virg' Teixeira Santos. 2013. *Apostas online: o caso da Primeira liga de Futebol Portuguesa*. Ph.D. thesis, Instituto Superior de Economia e Gestão.
- Sexton, TR, Silkman, RH, & Hogan, A.J. 1986. Data envelopment analysis: Critique and extensions. *Measuring Efficiency: An Assessment of Data Envelopment Analysis*, 73–105.
- Shamblin, James E, & Stevens, G T (Gladstone Taylor). 1979. *Pesquisa operacional: uma abordagem básica*. Atlas.
- Talluri, Srinivas. 2000. Data Envelopment Analysis: Models and Extensions. *Journal of Portfolio Management*, **31**, 8–11.
- Tiedemann, Torben, Francksen, Tammo, & Latacz-Lohmann, Uwe. 2011. Assessing the performance of German Bundesliga football players: A non-parametric metafrontier approach. *Central European Journal of Operations Research*, **19**(4), 571–587.
- Torgersen, Arne Martin, Forsund, Finn R., & Kittelsen, Sverre a. C. 1996. Slack-adjusted efficiency measures and ranking of efficient units. *Journal of Productivity Analysis*, **7**(4), 379–398.
- Trick, Michael A, & Yildiz, Hakan. 2007. *Bender's cuts guided large neighborhood search for the traveling umpire problem*. Springer. Pages 332–345.
- Troutt, Marvin D. 1995. A maximum decisional efficiency estimation principle. *Management Science*, **41**(1), 76–82.
- Winston, Wayne L. 2003. *Operations Research: Applications and Algorithms*.
- Yavuz, Mesut, Inan, Umut H., & Figlali, Alpaslan. 2008. Fair referee assignments for professional football leagues. *Computers and Operations Research*, **35**(9), 2937–2951.