



PREVISÃO DE PÚBLICO COM MODELO DE REGRESSÃO LINEAR PARA
JOGOS DO BOTAFOGO DE FUTEBOL E REGATAS

Caio Fernando dos Santos Araujo

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Produção, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Produção.

Orientador: Edilson Fernandes de Arruda

Rio de Janeiro

Abril de 2018

PREVISÃO DE PÚBLICO COM MODELO DE REGRESSÃO LINEAR PARA
JOGOS DO BOTAFOGO DE FUTEBOL E REGATAS

Caio Fernando dos Santos Araujo

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM
CIÊNCIAS EM ENGENHARIA DE PRODUÇÃO.

Examinada por:

Prof. Edilson Fernandes de Arruda, D.Sc.

Prof. Glaydston Mattos Ribeiro, D.Sc.

Prof. Lino Guimarães Marujo, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
ABRIL DE 2018

Araujo, Caio Fernando dos Santos

Previsão de público com modelo de Regressão Linear para jogos do Botafogo de Futebol e Regatas / Caio Fernando dos Santos Araujo – Rio de Janeiro: UFRJ/COPPE, 2018.

IX, 61p.: il.; 29,7 cm.

Orientador: Edilson Fernandes de Arruda

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia de Produção, 2018.

Referências Bibliográficas: p. 59-61.

1. Regressão linear. 2. Futebol. 3. Estádios. I. Arruda, Edilson Fernandes de. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Produção. III. Título.

AGRADECIMENTOS

Ao finalizar este trabalho e tornar-me mestre, olho para trás e vejo a sorte que tive na vida ao ter tantos professores e professoras na vida: os das salas de aula e os que tive na minha família.

Ao orientador desse trabalho, Edilson, preciso genuinamente agradecer por topar orientar esta pesquisa enquanto ainda era um breve trabalho sobre métodos estocásticos e, mais que isso, por nunca ter desistido de mim. O trabalho no Botafogo tomou meu tempo e dedicação a ponto de eu quase abandonar de vez essa pesquisa. Graças ao meu orientador, finalizo este trabalho.

Aos companheiros e chefes do Botafogo, pessoas que trabalham com enorme disposição e amor ao clube, em ritmo e carga de trabalho que só quem conhece perto sabe.

Aos professores da UFRJ, importantes na minha formação da Graduação e do Mestrado. É com enorme orgulho que me torno mestre na mesma casa e salas em que me formei Engenheiro de Produção.

Agradeço aos professores do CAp/UERJ, que me formaram aluno e cidadão: lá entrei criança, aos 6, e saí adulto, aos 18.

E como sempre gosto de deixar o melhor para o final, preciso agradecer à minha família, aqui incluindo minha fantástica namorada e minha legião de amigos, por aturar meu jeito de ser, extremamente devoto à coisa mais fantástica já produzida neste planeta: o Futebol. Quis o destino que eu pudesse trabalhar com a maior paixão da minha vida, e a forma como minha família valorizou e incentivou minha formação educacional permitiram isso.

Uma das minhas maiores alegrias nestes 29, quase 30, anos de vida foi o dia em que proporcionei e vi o escandaloso sorriso orgulhoso da minha mãe no dia da minha colação de grau como Engenheiro de Produção. Como disse, nessa mesma casa na Ilha do Fundão. Eu infelizmente não vou poder ver aquele seu hipnotizante sorriso ao tornar-me mestre agora, Mãe, mas sei que de alguma forma você está irradiante de orgulho da pessoa que o filho que você gerou está se tornando. Obrigado.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

PREVISÃO DE PÚBLICO COM MODELO DE REGRESSÃO LINEAR PARA
JOGOS DO BOTAFOGO DE FUTEBOL E REGATAS

Caio Fernando dos Santos Araujo

Abril/2018

Orientador: Edilson Fernandes de Arruda

Programa: Engenharia de Produção

A previsão de público para jogos de futebol é feita quase sempre utilizando informações de compra antecipada de ingressos e baseada na experiência dos organizadores. Este trabalho propõe um método de cálculo do público previsto para jogos do Botafogo de Futebol e Regatas utilizando Regressão Linear. Para isso, foram propostas variáveis que tentam explicar o que leva o torcedor ao estádio, que englobam fatores como aspectos econômicos e resultados esportivos. Foram usados borderôs (boletim financeiro) de jogos do Botafogo no Rio de Janeiro durante os Campeonatos Brasileiros de 2010, 2011, 2012, 2013 e 2014. Uma estimativa confiável do público facilitaria, entre outras coisas, o dimensionamento do corpo operacional para os jogos (como a quantidade de catracas que serão abertas). O modelo permite conclusões sobre o que faz o torcedor assistir a um jogo de futebol no estádio.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

ATTENDANCE FORECAST FOR BOTAFOGO DE FUTEBOL E REGATAS'
MATCHES BY MEANS OF LINEAR REGRESSION

Caio Fernando dos Santos Araujo

April/2018

Advisor: Edilson Fernandes de Arruda

Department: Industrial Engineering

In Brazil, the attendance of football games is usually estimated using advance ticket purchase information and the previous experience of the organizers. This paper presents a method for estimating the expected attendance to matches of Botafogo de Futebol e Regatas, which employs Linear Regression. Variables were suggested in an attempt to explain what makes fans attend a match, which encompass factors such as economic aspects and sports results. We used financial newsletters from games of Botafogo at Rio de Janeiro during the 2010, 2011, 2012, 2013 and 2014 Brazilian Championship seasons. A reliable estimate of the public would facilitate, among other things, the sizing of the operating facilities to serve games (such as the number of gates to be opened). The proposed method may assist the decision maker to understand what bring fans to the stadium.

SUMÁRIO

| | |
|--|----|
| 1. Introdução | 1 |
| 2. Revisão Bibliográfica | 8 |
| 3. Modelagem: previsão do publico | 20 |
| 3.1 Como é feito na prática | 20 |
| 3.2 Método | 21 |
| 3.3 As variáveis | 24 |
| 3.4 O modelo e os resultados | 39 |
| 4. Discussão | 48 |
| 5. Conclusão | 57 |
| Referências | 59 |

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1 - Relação entre previsão de público e custos/receitas. | 6 |
| Figura 2 - Os motivos que afastam os torcedores ingleses dos estádios em jogos em casa e fora. | 17 |
| Figura 3 - Os 3 motivos mais importantes, para os torcedores ingleses, para ir a uma partida. | 18 |
| Figura 4 - Um exemplo de borderô de uma partida. | 22 |
| Figura 5 - Comportamento da curva logarítmica de 1 a 31 | 30 |
| Figura 6 - Política de preços da temporada 2017/8 do Tottenham Hotspur (Inglaterra). | 32 |
| Figura 7 - Histograma com preços mínimos praticados nos jogos do Botafogo FR... .. | 33 |
| Figura 8 - Histograma com preços mínimos praticados nos jogos do Fluminense FC. | 34 |
| Figura 9 - Comparação entre o público estimado e o público real para 18 jogos do Botafogo. | 46 |
| Figura 10 - Comparação entre o público estimado e o público real para os 7 jogos do Fluminense. | 47 |
| Figura 11 - Entrada de torcedores por faixa de horário – Botafogo x Luverdense (01/08/2015). | 49 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 - Informações de público, renda e custos de 2 clássicos cariocas no Campeonato Brasileiro de 2017. | 3 |
| Tabela 2 - Média de público por estádio e por Campeonato Brasileiro do Fluminense. | 23 |
| Tabela 3 - Correlação do público do Botafogo com a pontuação obtida nos 5 jogos anteriores. | 25 |
| Tabela 4 - Correlação do público do Fluminense com a pontuação obtida nos 5 jogos anteriores. | 26 |
| Tabela 5 - Classificação de “cedo”, “médio ou “tarde” para os jogos das bases de acordo com o horário de início da partida e dia da semana. | 28 |
| Tabela 6 – Comportamento das variáveis HB1 e HB2 de acordo com horário de início 28 | 28 |
| Tabela 7 - Comportamento da variável que mede a distância da data da partida para o dia de pagamento..... | 31 |
| Tabela 8 - Classificação dos rivais do Botafogo, na base observada, quanto à rivalidade. | 37 |
| Tabela 9 - Classificação dos rivais do Fluminense, na base observada, quanto à rivalidade. | 38 |
| Tabela 10 - Comportamento das variáveis RB1 e RB2 de acordo com o grau de rivalidade com o adversário..... | 38 |
| Tabela 11 - Resumo das variáveis dependentes construídas para os modelos de previsão. | 39 |
| Tabela 12 - Resultados da Regressão Linear para o modelo do Botafogo utilizando o MQO..... | 40 |
| Tabela 13 - Resultados da Regressão Linear para o modelo do Fluminense utilizando o MQO..... | 42 |
| Tabela 14 - Resultados da Regressão Linear para o Botafogo com apenas 80% da base utilizando o MQO..... | 44 |
| Tabela 15 - Resultados da Regressão Linear para o Fluminense com apenas 80% da base utilizando o MQO..... | 45 |
| Tabela 16 - Diferença nos resultados do modelo de estimativa do Botafogo para a realidade 46 | 46 |
| Tabela 17 - Diferença nos resultados do modelo de estimativa do Fluminense para a realidade. 47 | 47 |
| Tabela 18 - Itens de custo, por faixa esperada máxima de público, no Estádio Nilton Santos em março/2017 (valores originais foram normalizados e multiplicados por um fator único). 51 | 51 |
| Tabela 19 - Evolução dos custos totais e por pessoa, por faixa esperada máxima de público, no Estádio Nilton Santos em março/2017 (valores originais foram normalizados e multiplicados por um fator único). 51 | 51 |
| Tabela 20 - Incremento de custos ou perda de receitas pelo uso do estimador de público do Botafogo (valores originais foram normalizados e multiplicados por um fator único). 53 | 53 |

Tabela 21 - Comparativo do prejuízo acumulado abrindo partida na faixa de público com limite superior imediatamente acima à do público estimado com prejuízo acumulado abrindo partida na faixa de público superior à do público observado. 55

Tabela 22 - Comparativo do prejuízo acumulado (aumento de custos ou perda de receita) abrindo partida na 1ª faixa de público acima do público estimado com prejuízo acumulado abrindo partida na faixa de público superior à do público observado. 56

1. Introdução

O Botafogo de Futebol e Regatas é um clube com mais de 111 anos de história ligada ao futebol no Rio de Janeiro. Apesar de ser um clube de torcida espalhada pelo Brasil, é na capital do estado do Rio de Janeiro que o Botafogo joga a maioria absoluta de suas partidas oficiais de futebol: campeonatos regionais, nacionais e internacionais. E, nos últimos anos, quase todos os jogos em que o clube é mandante são sediados no Estádio Nilton Santos, o “Engenhão”.

Construído para os Jogos Pan-americanos de 2007, o então Estádio Olímpico João Havelange foi rebatizado em 2017. Passou a se denominar Estádio Nilton Santos, em homenagem ao histórico jogador do Botafogo, clube que administra o estádio desde 2007, após vencer licitação. Por se situar no bairro do Engenho de Dentro, esse estádio carioca também é rotineiramente chamado de “Engenhão”.

Devido a uma suposta falha no projeto¹, o estádio ficou em obras de reparo desde março de 2013 até o final de 2014. Nesse período, eventos abertos ao público foram suspensos, incluindo partidas de futebol. Em 2016, houve novo período sem uso pelo Botafogo: o estádio sofreu mudanças e adaptações para ser a mais importante instalação dos Jogos Olímpicos Rio 2016, sediando todos os eventos de atletismo, além de algumas partidas de futebol masculino e feminino.

Após 11 anos frequentando o Nilton Santos, exceto pelos dois períodos explicados acima, já podemos dizer que a torcida do Botafogo está completamente acostumada a ir ao estádio, o que se permite tentar construir modelos que expliquem o comportamento desse torcedor. O clube mandou partidas oficiais em estádios espalhados pelo estado do Rio de Janeiro e pelo restante Brasil nos anos de 2013, 2014 e 2016 (períodos em que o estádio esteve fechado). Pela proximidade (menos de 7km de distância) e pelo tamanho (únicos estádios do Rio com capacidades para mais de 40 mil espectadores cada), talvez o único estádio que se assemelhe ao “Engenhão” seja o Maracanã: até por isso, esse estádio sofrerá tratamento semelhante na modelagem estatística.

O Estádio Nilton Santos é dividido em 4 macro setores: Norte, Sul, Leste e Oeste, sendo os dois últimos setores tendo dois pavimentos cada – inferiores e superiores. Norte

¹ https://odia.ig.com.br/_conteudo/esporte/botafogo/2016-10-04/estudo-afirma-que-fechamento-do-engenhao-por-quase-dois-anos-foi-desnecessario.html

e Sul abrigam cerca de 4.500 torcedores cada. Os setores inferiores (Leste e Oeste) têm capacidade de 7.334 torcedores, enquanto os superiores (Leste e Oeste) suportam 10.775 cada. Normalmente a torcida do Botafogo só não fica na ala Sul, setor reservado aos visitantes. Em clássicos contra Vasco, Flamengo e Fluminense, a torcida do Botafogo fica, na maioria das vezes (situações em que a torcida visitante tem 50% da carga de ingressos, e não apenas 10% como nos demais jogos), apenas nos setores Norte, Oeste Inferior e Oeste Superior. Cabe ressaltar, no entanto, que essa distribuição de torcedores valeu apenas até 2015 e em partidas do Campeonato Carioca (como determinado pela Federação de Futebol do Estado do Rio de Janeiro). No Campeonato Brasileiro de 2017, por exemplo, o Botafogo ficou com 90% da capacidade para sua torcida também nos clássicos (como sempre fez quando era mandante contra os demais clubes do país), destinando os 10% restantes para a torcida do time visitante (no setor Sul), mesmo que se tratasse de um dos demais grandes clubes do Rio de Janeiro.

Deve-se comentar que os clássicos (com mando do Botafogo ou não) são cada vez mais frequentes no Estádio Nilton Santos e cada vez menos ocorrem no Maracanã. Esse último estádio ficou fechado entre 2010 e 2013, em preparação para a Copa do Mundo de 2014. Após a sua abertura para os clubes em 2013, o Maracanã apresenta valores de locação e custos operacionais bastante altos², sendo muito pouco vantajoso mandar uma partida oficial lá com menos de 30 mil pagantes³. Esse fator é alarmante, uma vez que a média de público pagante nos estádios brasileiros é bastante inferior a esse número. Em 2015, apenas o Corinthians ultrapassou a barreira dos 30 mil torcedores (em média), com média de 33.480⁴ pagantes. No ano seguinte, 2016, mais uma vez apenas o Corinthians teve média superior aos 30 mil pagantes: 30.336⁵. Em 2017 o cenário melhorou, com o Corinthians novamente tendo a melhor média de público pagante no Brasil e ultrapassando a barreira dos 30 mil, com 34.896⁶, mas ganhando a companhia de São Paulo (33.635) e Palmeiras (30.496). No entanto, considerando que o Maracanã é um estádio carioca e que apenas 2 clubes do Rio de Janeiro figuraram no “top 10” do ranking de média de público pagante em 2017 (o Flamengo, em 5º, com

² <http://www.lance.com.br/futebol-nacional/quanto-custa-para-jogar-maracana-tem-preco-mais-salgado-brasil.html>

³ https://odia.ig.com.br/_conteudo/esporte/2017-07-09/maracana-esta-cada-vez-mais-carro-e-inviavel-para-publicos-abaixo-de-30-mil.html

⁴ <http://globoesporte.globo.com/futebol/noticia/2015/12/publico-2015-corinthians-tem-maior-media-palmeiras-melhores-rendas.html>

⁵ <http://app.globoesporte.globo.com/futebol/publico-no-brasil/2016/>

⁶ <https://globoesporte.globo.com/numerologos/noticia/corinthians-tem-maior-media-de-pagantes-em-2017-sao-paulo-o-recorde-de-publico.ghtml>

20.802 pagantes, e o Botafogo, em 10º, com 15.172), é extremamente preocupante o padrão elevado do valor de locação e também do custo operacional do Maracanã. Com capacidade máxima de 45 mil pessoas e com custos significativamente mais baixos, o Estádio Nilton Santos cada vez mais se apresenta como a alternativa mais viável para o futebol do Rio de Janeiro, ao menos nos jogos de médio e grande apelo.

Para que se tenha ideia da diferença de custos operacionais entre o Nilton Santos e o Maracanã, e já demonstrando a importância de um controle eficaz de custos num estádio de futebol, a tabela 1 mostra informações retiradas dos borderôs de 2 clássicos cariocas no Campeonato Brasileiro de 2017, com apenas de 21 dias de distância temporal entre as partidas.

Tabela 1 - Informações de público, renda e custos de 2 clássicos cariocas no Campeonato Brasileiro de 2017.

Fonte: www.cbf.com.br

| Data | Estádio | Partida | Pagantes/ Presentes | Receita Bruta | Custo Operacional | Lucro Líquido |
|-------------|----------------|-----------------------|--------------------------------|--------------------------|------------------------------|--------------------------|
| 21/06/2017 | Nilton Santos | Botafogo x Vasco | 13.287/ 15.048 | R\$ 363.610 | R\$ 107.536 | R\$ 62.376 |
| 12/07/2017 | Maracanã | Fluminense x Botafogo | 13.670/ 14.851 | R\$ 414.045 | R\$ 309.737 | -R\$ 235.514 |

Como se pode notar, são partidas de mesma magnitude em termos de público: pouquíssima diferença de públicos presente e pagante. A renda também é similar. Apesar disso, é gritante a diferença em termos de custo operacional. De fato, entre as partidas comparadas, jogar no Maracanã custou cerca de 188,03% a mais que jogar no Nilton Santos. É importante ressaltar que não estão contabilizados neste parâmetro valores de locação dos estádios. Sendo ambas as partidas consideradas clássicos cariocas e com aparatos de segurança e trânsito semelhantes, e lembrando também que os 2 jogos foram noturnos (isto é: ambos tiveram gastos com refletores), nota-se que é aceitável a comparação entre os jogos. Apesar de a partida do Maracanã ter gerado uma receita bruta 13,87% superior, os altíssimos custos operacionais e demais fatores fizeram com que a partida gerasse um prejuízo acumulado de mais de R\$ 235 mil, isto é: 477,57% inferior ao lucro líquido (Receita Bruta, descontando Impostos, Custos Operacionais e Despesas Operacionais) do jogo no Nilton Santos.

Ao final deste trabalho, espera-se que fique claro como é complexa a gestão operacional de uma partida oficial de futebol, e como uma previsão de público assertiva

pode evitar custos desnecessários.

A primeira parte desta dissertação construirá um método de cálculo do público previsto para jogos do Botafogo de Futebol e Regatas utilizando Regressão Linear. Para tanto, foram propostas variáveis com o intuito de explicar o que leva o torcedor ao estádio. Essa motivação engloba fatores como aspectos econômicos, resultados esportivos e até questões como data e horário de realização das partidas. Foram usados borderôs (boletins financeiros) de jogos do Botafogo no “Engenhão” durante os Campeonatos Brasileiros de 2010, 2011 e 2012, assim como os jogos do Botafogo no Maracanã nos Campeonatos Brasileiros de 2013 e 2014. Analisando os jogos do Botafogo no Rio, com estádios que distam cerca de 6,6 km entre si, buscaremos entender os motivos que levam o torcedor do clube às partidas deste no Campeonato Brasileiro da Série A.

Uma estimativa confiável do público facilitaria, entre outras coisas, o dimensionamento do corpo operacional para os jogos (o que inclui, por exemplo, a quantidade de catracas que serão abertas). Apesar de ainda estar em fase de desenvolvimento, o modelo permite conclusões sobre o que motiva o torcedor a assistir um jogo de futebol no estádio.

Para efeito de comparação, também serão analisados jogos do Fluminense Football Club, outro tradicional e centenário clube de futebol profissional do Rio de Janeiro, válidos pelos Campeonatos Brasileiros de 2013 e 2014 e realizados no Rio de Janeiro. O Fluminense foi escolhido pelo fato de ter torcida de tamanho semelhante à do Botafogo (1,6% da população nacional torce para o Fluminense, enquanto 1,7% torce para o Botafogo)⁷, além de ter mandado seus jogos no Estádio Nilton Santos e no Maracanã, de maneira semelhante ao Botafogo, como veremos mais à frente.

Podemos fazer uma analogia entre a previsão de público de um jogo e a previsão de demanda de uma empresa, da mesma forma como a simulação da entrada dos torcedores no estádio pode ser comparada ao planejamento operacional da empresa que passa a conhecer a sua demanda. Um dos desafios desse trabalho é pensar o torcedor como um cliente (por isso a analogia do futebol com outras indústrias), raciocínio nem sempre utilizado no futebol brasileiro.

Apesar de poder parecer algo completamente novo no futebol, a previsão de público é algo necessário para a realização de uma partida oficial no Brasil. No estado

⁷<http://globoesporte.globo.com/futebol/noticia/2016/12/pesquisa-mostra-que-torcida-sem-time-fica-na-frente-de-fla-e-corinthians.html>

do Rio de Janeiro, onde o Botafogo FR manda suas partidas, o clube mandante deve, por exemplo, emitir um certificado de liberação do estádio junto ao Corpo de Bombeiros para cada partida. E o certificado tem valores diferentes dependendo da quantidade de público informada. Existem várias faixas (intervalos) de público e o preço de liberação aumenta em função dessas. Naturalmente, é mais caro liberar o estádio para receber mais torcedores. A carga de ingressos posta à venda é limitada ao certificado emitido, o que força o clube a fornecer sua melhor estimativa. Além disso, a previsão de público fornecida pelo time mandante irá municiar o plano de jogo, elaborado dias antes da realização da partida. Esse plano contém a definição de efetivo policial, horário de abertura de portões, entre outras informações e definições operacionais.

Pode-se imaginar, assim, que a previsão de público tem papel importante na realização de uma partida de futebol. Se a previsão subestimar o público disposto a ir ao jogo, o clube mandante perderá receitas, dentre as quais podemos destacar obviamente a venda de ingressos. Entretanto, existem também receitas indiretas como venda de alimentos e bebidas, produtos oficiais, vagas de estacionamento, entre outras possibilidades. E não é só a perda de receita que é um fator negativo de uma previsão que subestime o público disposto a ir à partida: pode ocorrer confusão na venda de ingressos (no Brasil ainda se compram muitos ingressos no dia da partida, no local do jogo), o que pode gerar filas e uma experiência ruim para o torcedor, afastando-o do estádio em partidas futuras e, aí sim, comprometendo ainda mais receitas em um círculo vicioso.

Por outro lado, uma previsão de público também inadequada, mas que superestime o público presente ou o pagante não faz perder receitas. Todavia, acarretará em um aumento de custos operacionais, o que é tecnicamente equivalente. Como já mencionado, esperando (muito) mais torcedores do que o que realmente virão para a partida, paga-se mais caro por certificados e licenças desnecessárias, além de se ter custo de pessoal e operacional maior. A título de exemplificação, em 2017, no Estádio Nilton Santos, a equipe de administração trabalhou com 7 faixas de público para calcular os custos operacionais de uma partida: de 0 a 5000 pessoas, de 5.001 a 9.999, de 10.000 a 14.999, de 15.000 a 19.999, de 20.000 a 29.999, de 30.000 a 39.999 e, finalmente, de 40.000 a 45.000 (lotação máxima). Foram 17 itens de custo, que variam desde custos de pessoal operacional a aluguel de grades, manutenção e limpeza, entre outros. O aumento do custo operacional do estádio é de 13,4% da faixa de 10.000 a 14.999 quando comparado à faixa de 5.001 a 9.999. Se compararmos a faixa de 15.000 a 19.999 com a

de 5.001 a 9.999, o aumento de custo chega a 29,0%. Ou seja: superestimar o público pode custar muito caro aos realizadores da partida (os clubes).

A figura 1, abaixo, resume o dilema dos gestores de clubes e estádios no momento de prever o público de uma partida de futebol.

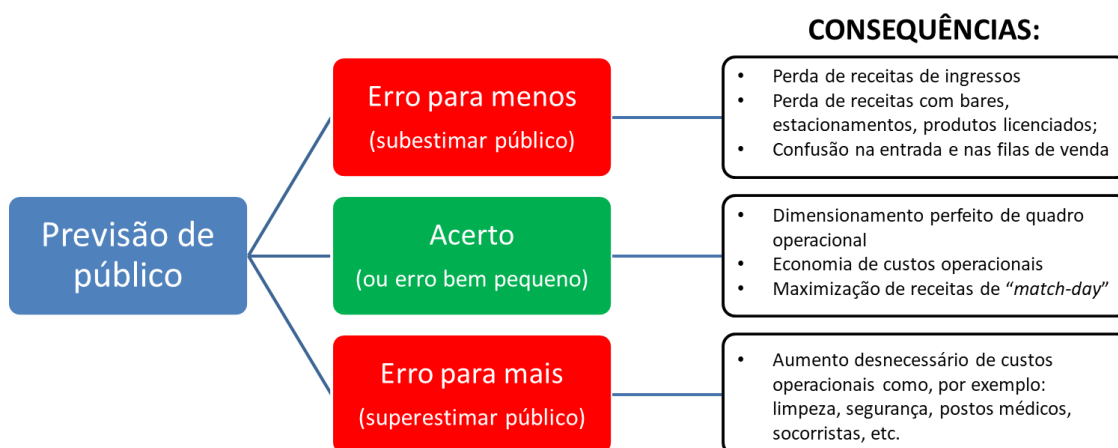


Figura 1 - Relação entre previsão de público e custos/receitas.

Fonte: O autor.

Atualmente a previsão de público é feita jogo a jogo, mas muitas vezes de forma empírica (baseada na experiência dos organizadores). Em geral, também se faz uso da estatística, porém apenas e tão somente por meio da observação (e não simulando ou se utilizando de modelos de regressão) de partidas anteriores.

O principal objetivo deste trabalho é construir um modelo matemático, através da métodos estatísticos, de previsão de público para jogos do Botafogo de Futebol e Regatas. O modelo de previsão do Fluminense será construído para efeitos comparativos, de modo auxiliar.

O modelo de previsão também poderá dar pistas e informações sobre os principais motivos que podem fazer o torcedor ir ao jogo, sendo descobrir essas razões um objetivo paralelo deste trabalho.

Por fim, tornar o processo de previsão de público mais confiável, porém sem desprezar o valioso conhecimento da prática, conseguindo evitar custos desnecessários e não perder receitas, é mais um objetivo deste trabalho. Como mencionado anteriormente, uma boa previsão de público pode proporcionar aumento de receitas em

jogos “bons” e “grandes”, diminuir custos em jogos “pequenos” e “ruins” e, por fim, proporcionar sempre uma boa a experiência ao torcedor, que é um consumidor muito particular. Apaixonado, mas que merece respeito e conforto. Afinal, frequentar um estádio de futebol, ao contrário do que possa parecer, sempre tem um fator racional.

2. Revisão Bibliográfica

Apesar de ainda ser incipiente no Brasil, o estudo da demanda por eventos esportivos é encontrado em uma série de artigos e trabalhos estrangeiros, principalmente nos Estados Unidos e no Reino Unido. Existem várias tentativas de se determinar o que leva um torcedor a um estádio de futebol (ou o afasta dele). No entanto, a ideia de se aplicar um modelo apenas para um time ainda é inovadora. Uma vantagem dessa abordagem é que a torcida de um time específico apresenta comportamento e tamanho bem definidos. Além disso, os times comumente mandam seus jogos em um único mercado – no caso do Botafogo, a cidade do Rio de Janeiro. Em contraste, a enorme maioria dos trabalhos constrói modelos que fazem uso de base de dados variada, englobando campeonatos inteiros, misturando vários times. Tais estudos, ao passo que se prestam a explicar o comportamento geral do torcedor, não nos permitem captar comportamentos específicos de torcedores de certo time ou certa região.

Fugindo um pouco do padrão de tentar inferir o comportamento geral do torcedor, Lima (2015) analisa o comportamento do torcedor do Esporte Clube Bahia residente na região metropolitana de Salvador, sob um enfoque econômico. Nota-se imediatamente a relação com o presente trabalho. Entretanto, embora também se proponha a fazer um modelo focado em um único time, o presente trabalho vai além da ótica puramente econômica. O foco aqui é mais geral: pretende-se aproveitar resultados e conclusões de trabalhos anteriores com foco em previsão de público e na busca de fatores que influenciam a ida do torcedor ao estádio. Além disso, propõe-se um método de avaliar motivações diversas, não apenas econômicas, para a presença no estádio de uma torcida específica, associada a um mercado local conhecido.

Há tempos tradicionais no futebol europeu, os carnês de ingresso para a temporada (os famosos “*season tickets*”) infelizmente ainda são pouco comuns no Brasil. Outra iniciativa, entretanto, floresceu. Até o início dos anos 2000, começaram a surgir nos principais clubes brasileiros os programas de Sócio-Torcedor, que experimentaram grande crescimento nos últimos 10 anos. Ainda assim, em que pese o advento de uma modalidade de associação voltada para o futebol e com benefícios na aquisição de ingressos, são poucos os clubes que vendem carnês de ingresso, com a maioria dos clubes apenas dando desconto em ingresso para os associados ou garantindo ingressos apenas para as modalidades de associação mais cara.

Ao não vender antecipadamente os ingressos da temporada, o clube perde

previsibilidade de receita. Além disso, acaba ficando refém da decisão do torcedor, a cada jogo, de comparecer ou não ao espetáculo. Nesse cenário, ainda é muito comum que o torcedor deixe para comprar seu ingresso faltando pouco tempo para a partida, deixando a decisão de ir ou não ao jogo bastante sensível a aspectos econômicos, esportivos, de horário, climáticos, etc. Naturalmente, tudo isso traz muita volatilidade ao processo de gerenciamento das receitas e despesas do clube. Logo, estudar a presença dos torcedores de um único clube (excluindo, então, efeitos geográficos e distorções amostrais) pode ser bastante interessante, e é precisamente isso que se propõe neste trabalho.

Alguns dos precursores da ideia de se estudar a presença de torcedores em estádios com uso da econometria são Hart, Hutton e Sharot (1975). Pelo fato dos precursores serem ingleses e também pelo fato de o Reino Unido ser um país com grandes tradições no esporte e na academia, a maioria dos estudos do tipo com o futebol vêm desse país. No entanto, a escola norte-americana também publicou uma série de trabalhos pioneiros sobre o assunto, mas como o futebol está longe de ser o esporte mais popular nos Estados Unidos, em geral os estudos versam sobre outros esportes. Pode-se citar por exemplo a investigação de Demmert (1973) em relação ao público presente em jogos de Beisebol e o estudo de Noll (1974), que analisa e compara os fatores que determinam a presença de público nas ligas dos EUA de beisebol, basquete, futebol americano e hóquei. Note que o estudo foca nos quatro esportes mais populares daquele país.

Kuypers (1996) tenta inferir as razões pelas quais os ingleses assistem a jogos de futebol, tanto nos estádios quanto pela televisão. Há conclusões esperadas, como o fato de a qualidade dos times impactar positivamente na presença, e a tendência do público ser maior quanto mais importante for o jogo (o que depende da fase do campeonato e da distância para os times extremos da tabela). Um resultado não tão imediato é a conclusão de que a transmissão televisiva ao vivo de um jogo (de uma rede de TV britânica escolhida – o que não dá conclusões sobre quaisquer outras transmissões) não tem impacto no público das partidas. Conclusão semelhante à de Souza (2004), que, por sua vez, analisou a influência da transmissão de um canal de assinatura brasileiro na disposição dos torcedores em ir ao estádio.

De certa forma, essas conclusões vão contra o senso comum, de que a transmissão televisiva ao vivo, principalmente de TV aberta, oferece “concorrência” ao estádio. É comum o argumento de que uma transmissão gratuita do espetáculo, no

conforto do lar do torcedor, ofereceria uma concorrência poderosa à experiência in loco. Observa-se, no entanto, de maneira empírica e complementar aos estudos, que a transmissão ao vivo em TV aberta apresenta-se como um motivo que desestimula a ida ao estádio apenas se conjugada a (e só depois de cumprida) uma série de fatores. Para jogos em horários ruins (mais suscetíveis a problemas de transporte público x segurança pública – sobretudo numa metrópole sul-americana, como o Rio de Janeiro), caros e de baixo apelo, a transmissão pode fazer a diferença. Para jogos normais ou com apelo, a transmissão televisiva ao vivo em TV aberta não influencia na partida: é entendimento comum entre os dirigentes esportivos que talvez até ajude a promover. Deixando de lado o jogo no estádio, TVs por assinatura (canais fechados) experimentam audiências maiores quando seus eventos são mostrados simultaneamente por televisões abertas: a crença é que a TV aberta, ao mostrar a partida, promove e fala do jogo bastante em seus programas nos dias anteriores, ajudando a criar uma expectativa maior sobre o evento. Se isso funciona para TV fechada (que é a “mesma” coisa da TV aberta e está a canais de TV de distância) aumentar seu público, pode ajudar também no estádio, que oferece experiência mais completa e diferente da TV.

No futebol norueguês, uma liga que não figura entre as mais fortes e ricas do mundo, o trabalho de Kringstad, Solberg e Jakobsen (2018) encontra resultados muito interessantes: enquanto a transmissão ao vivo de jogos locais em TV aberta é positivamente correlacionada com a presença de público nos estádios, a transmissão de jogos das grandes ligas nacionais europeias (o *big-five*: Inglaterra, Alemanha, França, Espanha e Itália), de forma contrária, causa um efeito substitutivo na presença nos estádios. Além disso, o trabalho é mais um a atestar que o dia da semana em que a partida é jogada tem influência: no caso norueguês, é esperado maior público quando joga-se nos finais de semana.

O resultado reforça a ideia de que há particularidades em cada cultura: no caso, de um país (Noruega). O ideal é se fazer uma análise restrita a uma liga ou, possivelmente, a um determinado clube ou torcida, como visto em Lima (2015) e defendido também por Strnad, Nerat e Kohek (2017), que comprovam que o modelo individual, focado em apenas um time (Barnsley, Ipswich ou Preston) gera previsões de público melhores do que modelos gerais, esses últimos sendo a regra na literatura. A distância (física) para a sede do time visitante é fator importante para a presença de público em jogos do Ipswich e do Barnsley, por exemplo, mas não para o Preston. Este tipo de análise, particular, será apresentado no presente trabalho, com um modelo focado

no Botafogo de Futebol e Regatas.

A incerteza do resultado é apontada como um fator que contribui para a ida do torcedor inglês ao estádio por Kuypers (1996) e também por Forrest e Simons (2002), com os dois trabalhos utilizando cotações das tradicionais casas de apostas inglesas para medir a probabilidade (a priori) de vitória dos times. Souza (2004) chega à mesma conclusão para o caso do futebol brasileiro, mas variando a medida da incerteza do resultado, que pode ser estimada também pela diferença prévia de pontos entre os dois times envolvidos na partida. A incerteza do resultado é um resultado amplamente verificado e comprovado em trabalhos da área, não só no futebol: Falls e Natke (2016) também apresentam evidências, para o caso do futebol americano universitário, de validade da teoria. Contudo, este resultado pode nem sempre ser fator importante no esporte. De fato, Lemke, Leonard e Tlhokwane (2010) verificam, em seu estudo sobre a Major League Baseball (Liga de Beisebol dos Estados Unidos), que quanto maior a expectativa prévia de vitória do time da casa, maior a influência direta e positiva sobre a demanda de ingressos. Esse resultado, talvez, pode ser consequência do fato de a temporada de beisebol ter 162 jogos, seguidos de playoffs, o que certamente altera a forma como o espectador encara uma partida da temporada regular. É natural que o espectador de futebol encare de maneira diferente temporadas disputadas em *pontos corridos*.

Madalozzo e Villar (2009) e Strnad, Nerat e Kohek (2017) também não acham, em seus respectivos trabalhos sobre futebol, resultados significativos para atestar a incerteza do resultado como fator importante na presença de público ao estimarem esse efeito por meio da diferença de pontos entre as equipes. É interessante identificar que, apesar de os dois trabalhos utilizarem técnicas de previsão e modelagem distintas entre si (respectivamente Regressão Linear e Redes Neurais), eles compartilham resultados como, por exemplo: o fato de o interesse do público aumentar de acordo com o andamento do campeonato e também quando o time da casa enfrenta um rival, além de, é claro, a posição do time da casa na classificação do campeonato ser fator relevante para a presença de público. No entanto, talvez pelo fato de Madalozzo e Villar (2009) estudarem o futebol brasileiro e Strnad, Nerat e Kohek (2017) o futebol inglês, somente o primeiro trabalho observe evidência de que o dia da semana em que a partida ocorre e a classificação do time visitante sejam fatores importantes na presença dos fãs.

Outra conclusão comum a quase todos os trabalhos revisados é que a qualidade da partida tem efeito positivo no público, e não só em futebol, já que Welki e Zlatoper

(1999), que também concluem que a qualidade do jogo é fundamental para a presença de público no futebol americano. No entanto, a forma de medir a qualidade varia bruscamente entre os trabalhos. García e Rodríguez (2002) estimam a qualidade do jogo pela quantidade de jogadores com passagens por seleções nacionais envolvidos no jogo, medida essa também utilizada por Kuypers (1996) e Souza (2004). Já o orçamento dos clubes serve como parâmetro para medir a qualidade e foi proposto por Falter e Pérignon (2000) e também utilizado por García e Rodríguez (2002). Adaptar os orçamentos dos clubes como um estimador da qualidade dos times não parece, ao menos para um campeonato brasileiro, uma boa medida. Observa-se que em campeonatos ingleses é altíssima a correlação entre a pontuação dos times (desempenho esportivo) e os salários pagos aos atletas (força econômica): isto é mostrado por Szymanski e Hall (2003) e Szymanski (2013 e 2014) com coeficientes de explicação sempre próximos a 90%. Além de gasto salarial não necessariamente equivaler a orçamento, sabemos que campeonatos sul-americanos, principalmente os brasileiros, não experimentam uma relação do tipo: o desempenho esportivo não costuma depender tanto do orçamento dos clubes, que não têm eficiência administrativa comparada a dos ingleses. Muitas vezes o orçamento dos clubes brasileiros não é cumprido à risca, seja a instituição grande ou pequena: a pressão das arquibancadas ou a falta de profissionalismo dos dirigentes pode fazer com que se gaste mais do que se pode. E as consequências indesejadas desse tipo de comportamento podem se estender ao terreno esportivo. Sendo assim, é interessante buscar outro tipo de estimador para a qualidade das equipes brasileiras: até porque é conhecida a máxima que “antes de o Campeonato Brasileiro começar, são 12 favoritos”, diferente dos grandes campeonatos europeus (quando há no máximo 4 ou 5 candidatos ao título).

Szymanski (2001) concluiu que, em seu estudo, a soma das posições dos times envolvidos no jogo foi estatisticamente significativa e teve impacto na presença do público, isto é, quanto mais bem posicionados os times, maior a probabilidade de um público expressivo.

A rivalidade entre os times envolvidos no jogo também pode ser importante para a presença do público. Bortoluzzo, Iaropoli e Machado (2011) mostram que as rivalidades locais normalmente têm um impacto positivo na presença. Os mesmos autores também comprovam um resultado apontado por Madalozzo e Villar (2009), que mostram que a renda per capita das cidades que sediam os jogos tem impacto negativo na presença do público, um resultado aparentemente contraintuitivo, mas que pode apontar para a

existência de mais opções de lazer concorrendo com o futebol. No entanto, os mesmos trabalhos concluem que, como esperado, o preço do ingresso tem um impacto negativo no público.

Outro efeito que os trabalhos tentam medir e estimar é o momento do time da casa e o seu impacto no público: se o retrospecto recente for bom, espera-se um impacto positivo no público. E, mais uma vez, há uma série de formas de estimar a fase do time e a importância do jogo. Nesse aspecto, destaca-se a medida proposta por Kuypers (1996) e também usada por García e Rodríguez (2002), que estimam a importância do jogo para o campeonato com uma função que multiplica o número de jogos restantes pela diferença de pontos do time da casa para o líder, considerando também a partida a ser jogada. Quanto menor a medida, mais importante o jogo, e maior o público esperado (desde que se mantenham constantes as outras variáveis).

Como se pode observar, há uma série de efeitos e de maneiras para estimá-los. Falter e Pérignon (2000) dividem as variáveis em três grupos a fim de facilitar o tratamento: variáveis relacionadas ao ambiente econômico, variáveis relacionadas à qualidade da partida e variáveis ligadas aos incentivos que um torcedor tem para ir ao estádio. Esse tipo de classificação é utilizada em diversos trabalhos, e o trabalho de Bortoluzzo, Iaropoli e Machado (2011) explica e discute os tipos de variáveis.

Com base nas variáveis apresentadas pelos trabalhos dos autores citados e nas linhas de raciocínio apresentadas, a presente dissertação procura mostrar o que motiva o torcedor do Botafogo a ir ou não ao estádio. Obviamente, muitos dos trabalhos apontam fatores que impactam torcedores de outros times, brasileiros ou não, o que não significa que os mesmos fatores servirão ao estudo de caso, já que há diferenças culturais e socioeconômicas, dentre outras discrepâncias.

● Técnicas de previsão

Várias são as técnicas de previsão utilizadas na literatura em geral. Com foco nos trabalhos sobre esporte, a seção abordará as principais técnicas e trabalhos de referência que as utilizam.

A Regressão Linear com o método dos mínimos quadrados ordinários é amplamente usada na literatura, sendo o modelo mais visto nos trabalhos de previsão de público no esporte. O uso deste modelo pressupõe que o erro seja distribuído

aleatoriamente numa curva normal. Esta forma de abordagem é utilizada, por exemplo, nos trabalhos de Hart, Hutton e Sharot (1975), Forrest e Simons (2002), García e Rodríguez (2002), Souza (2004) e Madalozzo e Villar (2009), Lemke, Leonard e Tlhokwane (2010) e Bortoluzzo, Bortoluzzo, Machado *et al* (2017). Todos esses trabalhos abordam o futebol, mas é interessante observar que há registros também desta modelagem no trabalho de Drever e McDonald (1981), que trataram da questão de público em jogos de futebol australiano (*Aussie Rules*), um esporte originado na Austrália no século XIX e popular até hoje, com regras distintas às do futebol como praticado no Brasil, e com características derivadas do rúgbi e do futebol gaélico.

São vários os trabalhos que fazem uso de um modelo Tobit para efetuar a regressão. Pode-se citar, por exemplo, trabalhos como o de Bortoluzzo, Iaropoli e Machado (2011) e Welki e Zlatoper (1999), Kuypers (1996). O modelo Tobit também faz uma regressão linear, e é escolhido quando há “censura” nos dados: os estádios têm restrições de capacidade, o que faz com que nem sempre os ingressos vendidos representem a demanda, pois jogos lotados normalmente terão demanda maior do que o número de ingressos vendidos (capacidade do estádio). Infelizmente, esse é poucas vezes o caso em estádios brasileiros.

O modelo Tobit é usado também por Mirabile (2015), em novo trabalho sobre futebol americano, desta vez na liga universitária. O trabalho traz uma rara e interessante discussão sobre o local em que as partidas analisadas são jogadas: campo neutro, quando nenhuma das duas equipes é local. É observado, desta forma, que a qualidade das equipes (resultados recentes) e o tamanho das universidades (se são times “grandes”) têm efeitos positivos sobre o público, que também tende a ser maior em estádios e arenas mais modernos e cobertos. No entanto, é esperado um efeito negativo no público se a partida é jogada numa cidade em que há algum time profissional local (seja em futebol americano ou em basquete, hóquei ou beisebol) durante sua temporada. Outro fator de impacto negativo à presença de público é a distância do local da partida às cidades-sede dos times.

Falls e Natke (2016) também usam um modelo Tobit para abordar a presença de público em futebol americano em ligas de estudantes. Os autores observam resultados clássicos como, por exemplo, o efeito negativo que o preço dos ingressos causa no público. O interesse dos fãs aumenta de acordo com a evolução do campeonato se o time tem bom desempenho (ganha jogos).

Há também uso de Tobit em trabalhos que utilizam uma série de métodos, abordados mais a frente, como a produção de Szymanski (2001), Bortoluzzo, Bortoluzzo, Machado *et al* (2017) e de Lemke, Leonard e Tlhokwane (2010). Este último trabalho, aliás, compara diversas técnicas aplicando-as na presença de público em jogos de Beisebol nos Estados Unidos: além de Tobit, apresenta resultados com Tobit com Efeitos Fixos, Regressão Linear com Mínimo Quadrados Ordinários e Regressão Linear com Mínimo Quadrados Ordinários com Efeitos Fixos. O trabalho de Welki e Zlatoper (1999) é mais um que utiliza Tobit com Efeitos Fixos, além do Tobit, como já dito.

Um modelo estatístico de efeitos fixos pretende controlar os efeitos das variáveis omitidas que variam entre indivíduos, mas permanecem constantes ao longo do tempo. Assim, supõe-se que o intercepto varia de um indivíduo para o outro, mas é constante ao longo do tempo. Em termos esportivos, usar um modelo de efeitos fixos significa atribuir a cada clube ou time mandante suas próprias variáveis binárias de controle (*dummy*). Em formato de painel, cada time teria sua variável binária e, sempre que o dado observado fosse deste time em específico, a variável teria valor “1”, com as dos outros clubes tomando valor “0”, de modo a individualizar a observação (por considerar que cada torcida ou local tem características próprias).

É comum encontrar trabalhos que utilizam técnicas denominadas de “Efeitos Fixos”, simplesmente: nada mais do que uma regressão linear, sem censura nos dados, mas com efeitos fixos (individualização das observações, por clube, time ou franquia). Podemos elencar o trabalho de Meier, Konjer e Leinwather (2016), em raro trabalho sobre futebol feminino, e também a recente produção de Kringstad, Solberg e Jakobsen (2018), além de Lemke, Leonard e Tlhokwane (2010), como já mencionado.

É frequente, também, ver abordagens do tipo “log-linear”, explicando com parâmetros lineares uma transformação logarítmica da observação de público: os trabalhos de Forrest e Simmons (2000), Falter e Pérignon (2000) e Villa, Molina e Fried (2011) são bons exemplos.

Há, também, técnicas pouco vistas e que aparecem em poucos trabalhos. Lima (2015) utiliza cointegração de séries temporais e Kuypers (1996), além das modelagens com Tobit, utiliza um modelo de Heterocedasticidade e Autocorrelação em Grupo. Villa, Molina e Fried (2011) utilizam vários modelos lineares mistos, além de uma abordagem com efeitos autocorrelacionados e outra com erros autocorrelacionados. Bortoluzzo, Bortoluzzo, Machado *et al* (2017), em seu trabalho também utilizam uma Regressão

Linear Generalizada com distribuição Gama, além de usar Tobit e Mínimos Quadrados sem censura nos dados, como já exposto.

Strnad, Nerat e Kohek (2017), por sua vez, diferenciam-se de boa parte da literatura ao usarem Redes Neurais para preverem público em jogos de futebol de três clubes ingleses (Barnsley, Ipswich e Preston). Apesar da forma diferente de abordar o assunto, o trabalho encontra resultados encontrados que são amplamente difundidos na literatura, como o fato de a rivalidade local ser um fator que aumenta a presença de público no estádio e que o interesse pelos jogos aumenta conforme a temporada e o campeonato avançam.

Este trabalho abordará a questão da presença de público nos estádios Nilton Santos e Maracanã com Regressão Linear pelo método dos Mínimos Quadrados, como boa parte da literatura: trabalhos de futebol, futebol americano e mesmo beisebol. A escolha se deve também pelo fato como as observações de dados foram feitas e as variáveis construídas, utilizando os trabalhos de referência. Ao focar sempre em um único clube sendo o mandante da partida (um modelo com foco no Botafogo ou no Fluminense), não será necessária a abordagem por efeitos fixos, uma vez que a análise já será individualizada.

● **Produção de conteúdo na “indústria do futebol”**

A extensa bibliografia britânica sobre o assunto talvez influencie discussão sobre o tema em nível gerencial e estratégico entre os “dirigentes” ingleses. Anualmente, e há quase 20 temporadas, a *Premier League* (Campeonato Inglês de Futebol) conduz uma extensa pesquisa com os torcedores de cada um dos 20 clubes que jogam o principal campeonato nacional do mundo. A pesquisa (que já foi chamada de *Fan Survey* e hoje é conhecida como *Fan Panel*) gera relatórios de mais de 70 páginas com diversas conclusões sobre a forma como os torcedores consomem o produto da Premier League, dissecando as características dos torcedores, as opiniões sobre a qualidade do jogo, do serviço e dos estádios, a experiência e a forma de assistir às partidas pela TV e pela internet e, claro, dá uma série de informações sobre o trajeto dos torcedores aos estádios.

O relatório sobre a temporada 2007/08, por exemplo, mostra os fatores que contribuem para a ida (ou não) os torcedores a jogos dentro ou fora de casa (figura 2):

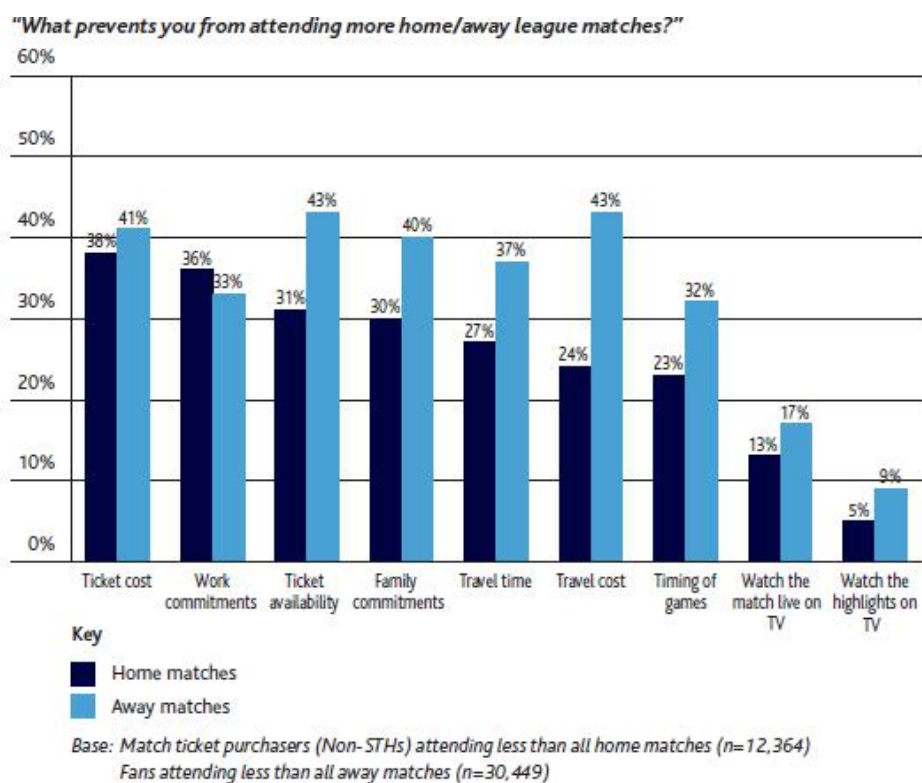


Figura 2 - Os motivos que afastam os torcedores ingleses dos estádios em jogos em casa e fora.

Fonte: Premier League Fan Survey 2007/08

Percebe-se que, dentre os motivos apresentados, o fato de a partida ser transmitida pela TV (ao vivo ou os melhores momentos) pouco influencia na decisão de não ir, como apontado em diversos trabalhos acadêmicos.

Da mesma maneira, quando perguntados sobre os elementos mais importantes que contribuem para a decisão de comparecer a uma partida, os torcedores apontaram seu veredito, detalhado na figura 3.

TABLE 5.1 BEST MATCH ELEMENTS*"Which of the following are the 3 most important elements about going to a game?"*

| <i>Base: Match attenders only</i> | 2005/06 <i>(n=15,750)</i> | 2006/07 <i>(n=23,100)</i> | 2007/08 <i>(n=30,862)</i> |
|-----------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| The quality of the football | 69% | 67% | 83% |
| Seeing the team play live | 91% | 91% | 78% |
| The passion of the crowd | N/A | 69% | 64% |
| Meeting family and friends | 15% | 23% | 21% |
| Singing/chanting | 22% | 23% | 21% |
| Stadium facilities | 8% | 11% | 20% |
| Other | N/A | 7% | 3% |

N/A = Not asked in previous survey.

Figura 3 - Os 3 motivos mais importantes, para os torcedores ingleses, para ir a uma partida.

Fonte: Premier League.

Novamente, como apontado e sugerido na bibliografia, a qualidade da partida é fundamental para a presença do torcedor no estádio.

Mais do que confirmar dados e conclusões conseguidos pelo meio acadêmico, a condução de pesquisa do tipo (com dados quantitativos e qualitativos) contribui para o amadurecimento da indústria do futebol, que passa a conhecer com mais profundidade o comportamento do seu elemento mais importante: o torcedor.

Acredita-se que o futebol brasileiro, já carente de profissionalização e produção acadêmica sobre gestão (sobretudo com viés matemático e estatístico), deveria investir em pesquisas como as feitas pela Premier League. Ao invés das relevantes, mas superestimadas pesquisas de tamanho de torcida (cujo grande benefício parece ser alimentar discussões de bares), nem sempre conduzidas com o rigor matemático devido, pesquisar sobre o que motiva o torcedor a ir ao estádio, ou entender o que o afasta das partidas (no campo ou na TV), talvez ajude a reduzir a fuga de torcedores para concorrentes como NBA (Liga profissional de Basquete Norte-Americana), NFL (Liga profissional de Futebol Americano dos Estados Unidos), UFC (maior campeonato mundial de artes marciais), campeonatos de futebol estrangeiros e até, para os mais jovens, os chamados *e-sports* (jogos de videogame).

O que deve motivar os gestores e pensadores do futebol brasileiro é que se deve estimular a pesquisa para que se ganhe vantagens competitivas, e para que se melhore

a qualidade do serviço. O entendimento do futebol como sendo parte da indústria do entretenimento (tendo como concorrentes não somente outros esportes, mas qualquer outra forma de entretenimento) é fundamental para o crescimento e resgate do setor. O fanático e o apaixonado podem nunca mudar de gosto ou de paixão, mas manter e atrair os demais fãs (esmagadora maioria) é necessário.

3. Modelagem: previsão do público

O presente capítulo visa introduzir um modelo de previsão de público para jogos mandados pelo Botafogo de Futebol e Regatas no Estádio Nilton Santos (Engenhão). Começa-se, todavia, com uma discussão da prática corrente entre clubes de futebol brasileiros. Em seguida, apresenta-se e discute-se o modelo proposto e suas variáveis.

3.1 Como é feito na prática

Seja para precificar a partida (tendo na maioria das vezes o objetivo de maximizar os lucros) ou para recolher taxas para legalizar a realização da partida, haverá pelo menos um funcionário do clube de futebol que terá uma expectativa prévia sobre o público da próxima partida. Sendo assim, de maneira completamente empírica ou com apoio de estatística apenas para consolidar os relatórios de público de jogos recentes, clubes de futebol estão acostumados a prever público dos seus jogos oficiais.

Sob a ótica do gestor que deve prever o público, é de praxe considerar 4 (quatro) parâmetros básicos, em torno de 3 a 6 dias da realização da partida. Estima-se que esses parâmetros tenham influência na decisão do torcedor de ir ao estádio. São eles: a fase do time (mandante), o dia da partida (se dia útil ou não), o horário da partida e o adversário. Curiosamente, o clube não comanda diretamente nenhum desses parâmetros. Naturalmente a fase do clube depende do bom futebol que o time tenha (ou não) jogado, mas o esporte admite vitórias, empates e derrotas que nem sempre estão associadas apenas ao bom ou mau futebol do time mandante. Além disso, na maioria dos casos, a equipe que trabalha com administração de estádios e arenas (ou se ocupa de questões para a realização da partida) não trabalha diretamente com o Departamento de Futebol.

Horário, dia e adversário são determinados pela tabela, sendo que dia e horário sofrem (principalmente no Brasil) influência muito grande das redes de Televisão (que buscam posicionar em seus horários de transmissão jogos com maior apelo de audiência). É sabido, contudo, que o horário bom para a TV não necessariamente é um horário bom para o torcedor que vai comparecer ao estádio, vide os jogos semanais em torno das 22h no meio da semana.

Após observar esses 4 parâmetros básicos, e tendo uma ideia aproximada do apelo da partida para seu torcedor, o gestor do clube precifica a partida, devendo ter a

sensibilidade de que um preço baixo demais pode causar prejuízos (ou desperdiçar receitas), e que um preço alto demais pode afastar torcedores (e assim causar prejuízos). Não é uma decisão fácil, e é comum que se busquem (mentalmente ou através de planilhas simples) jogos anteriores com perfil “prévio” semelhante ao do próximo jogo.

Dessa forma, desenvolver um fluxo processual preciso de previsão e/ou entender a fundo as variáveis que influenciam na decisão do torcedor de ir à partida pode ajudar o tomador de decisão, seja na pura previsão de público ou na precificação da partida.

3.2 Método

Para construir o modelo de Regressão Linear, são necessárias informações que constituam uma base de dados. Várias das informações são retiradas do borderô, que é o documento financeiro oficial de uma partida. Nele, é registrada a quantidade de pessoas em cada setor do estádio e quanto cada torcedor pagou pelo ingresso, além de gratuidades e descontos. A figura 4 a seguir apresenta um exemplo desse documento.



FEDERAÇÃO DE FUTEBOL DO ESTADO DO RIO DE JANEIRO
BOLETIM FINANCEIRO DO JOGO REALIZADO EM 23 de setembro de 2012

| JOGO: | | BOTAFOGO FR | | X | | SC CORINTHIANS | | ESTÁDIO: Olímpico João Havelange | | BF | | | |
|-----------------------|-------------------------------|------------------------------------|-----------------------------------|---------------|-----------|----------------|--|----------------------------------|--|---------------------|--|----|--|
| Competição (Espécie): | | OFICIAL | | X | | AMISTOSO | | Engenheiro | | | | | |
| Categoria: | | Internacional | | | | Interestadual | | X | | Local | | | |
| Competição: | | CAMPEONATO BRASILEIRO SÉRIE A 2012 | | | | | | LOCAL: | | RIO DE JANEIRO - RJ | | 52 | |
| INGRESSOS | | | | | | | | | | | | | |
| LOCALIZAÇÃO | | À VENDA | | DEVOLVIDOS | | VENDIDOS | | PREÇO | | ARRECADAÇÃO | | | |
| RECEITAS | NORTE INTEIRA | 2.204 | 89 | 2.115 | R\$ 20,00 | R\$ 42.300,00 | | | | | | | |
| | NORTE MEIA | 859 | 270 | 589 | R\$ 10,00 | R\$ 5.890,00 | | | | | | | |
| | OESTE SUPERIOR INTEIRA | 1.828 | 696 | 1.132 | R\$ 30,00 | R\$ 33.960,00 | | | | | | | |
| | OESTE SUPERIOR MEIA | 2.318 | 1.208 | 1.110 | R\$ 15,00 | R\$ 16.650,00 | | | | | | | |
| | OESTE INFERIOR INTEIRA | 1.208 | 300 | 908 | R\$ 30,00 | R\$ 27.240,00 | | | | | | | |
| | OESTE INFERIOR MEIA | 1.446 | 776 | 670 | R\$ 15,00 | R\$ 10.050,00 | | | | | | | |
| | SUL INTEIRA | 929 | 403 | 526 | R\$ 20,00 | R\$ 10.520,00 | | | | | | | |
| | SUL MEIA | 838 | 614 | 224 | R\$ 10,00 | R\$ 2.240,00 | | | | | | | |
| | LESTE SUPERIOR INTEIRA | 1.304 | 777 | 527 | R\$ 30,00 | R\$ 15.810,00 | | | | | | | |
| | LESTE SUPERIOR MEIA | 1.600 | 1.223 | 377 | R\$ 15,00 | R\$ 5.655,00 | | | | | | | |
| | LESTE INFERIOR INTEIRA | 2.298 | 568 | 1.730 | R\$ 30,00 | R\$ 51.900,00 | | | | | | | |
| | LESTE INFERIOR MEIA | 1.976 | 602 | 1.374 | R\$ 15,00 | R\$ 20.610,00 | | | | | | | |
| | GEO PREMIUM 1 | 14 | 0 | 14 | R\$ 30,00 | R\$ 420,00 | | | | | | | |
| | GEO PREMIUM 2 | 31 | 0 | 31 | R\$ 15,00 | R\$ 465,00 | | | | | | | |
| | SOCIO TORCEDOR OESTE INFERIOR | 177 | 0 | 177 | R\$ 15,00 | R\$ 2.655,00 | | | | | | | |
| | SOCIO TORCEDOR OESTE SUPERIOR | 1.904 | 0 | 1.904 | R\$ 15,00 | R\$ 28.560,00 | | | | | | | |
| | SOCIO TORCEDOR LESTE INFERIOR | 347 | 0 | 347 | R\$ 15,00 | R\$ 5.205,00 | | | | | | | |
| | GRATUIDADES POR FORÇA DE LEI | 5.088 | 1.853 | 3.235 | | | | | | | | | |
| TOTALS | | 26.369 | 9.379 | 16.990 | RECEITA | R\$ 280.130,00 | | | | | | | |
| DESPESAS | 1 TAXA FERJ | R\$ 14.006,50 | 21 TAXA DE BOMBEIROS | R\$ 1.200,01 | | | | | | | | | |
| | 2 INSS 5% DECRETO 832 | R\$ 14.006,50 | 22 CREDENCIAMENTO | R\$ 3.660,00 | | | | | | | | | |
| | 3 TAXA DE ARBITRAGEM | R\$ 8.500,00 | 23 TAXA DE ILUMINAÇÃO | R\$ 7.500,00 | | | | | | | | | |
| | 4 INSS 20% SI/ARB | R\$ 1.700,00 | 24 ALUGUEL DO CAMPO | R\$ 28.500,00 | | | | | | | | | |
| | 5 NF COOPAFERJ 4* ARB E OBS | R\$ 1.159,95 | 25 CARRO FORTE | R\$ 950,00 | | | | | | | | | |
| | 6 INSS 15% SI/4* ARB+OBSERV | R\$ 173,99 | 26 ALUGUEL DE GRADES | R\$ 7.000,00 | | | | | | | | | |
| | 7 SEGURO DOS ARBITROS | R\$ 158,27 | 27 POSTO MÉDICO | R\$ 12.200,00 | | | | | | | | | |
| | 8 DESPESA OPERACIONAL | R\$ 10.000,00 | 28 DESPESA OPERACIONAL DO ESTÁDIO | R\$ 35.000,00 | | | | | | | | | |
| | 9 INSS 20% S/DESP OPERAC | R\$ 2.000,00 | 29 REEMBOLSO PASSAGEM ARBITRAGEM | R\$ 343,00 | | | | | | | | | |
| | 10 DELEGADO | R\$ 1.000,00 | 30 INGRESSO PROMOCIONAL | R\$ 36.420,00 | | | | | | | | | |
| | 11 INSS 20% S/ DELEGADO | R\$ 200,00 | | | | | | | | | | | |
| | 12 TRANSPORTE (VAN) | R\$ 185,00 | | | | | | | | | | | |

Figura 4 - Um exemplo de borderô de uma partida.

Fonte: www.cbf.com.br

O borderô é a melhor e mais precisa fonte de informação sobre uma partida oficial no que diz respeito aos seus custos e receitas, além da distribuição dos torcedores presentes.

Neste trabalho foram utilizados dados dos borderôs dos jogos do Botafogo no Estádio Nilton Santos e no Maracanã nos Campeonatos Brasileiros de 2010, 2011, 2012, 2013 e 2014. A inclusão de jogos dos 2 estádios é justificada por pesquisa feita pelo primeiro autor com torcedores botafoguenses em 2013, em um grande portal esportivo brasileiro (globoesporte.com⁸), quando o Estádio Nilton Santos encontrava-se interditado e o Botafogo mandava seus jogos no Maracanã: de 3.299 entrevistados, 2.219 (67,26%) moravam na Região Metropolitana do Rio de Janeiro. Destes 2.219, 665 (29,97%) responderam ir mais aos jogos pelo fato de serem no Maracanã; 650 (29,29%) responderam que iam mais aos jogos antes porque eram no Estádio Nilton Santos; e, finalmente, 904 (40,74%) responderam não ter preferência por um dos estádios. Isto é:

⁸ <http://globoesporte.globo.com/rj/torcedor-botafogo/platb/2013/11/01/pesquisa-o-botafoguense-no-estadio/>

pode-se considerar, a partir desta pesquisa, que não há prejuízo estatístico em estimar uma equação única para explicar a presença de público (em Campeonatos Brasileiros) do Botafogo em jogos no Maracanã ou no Estádio Nilton Santos, uma vez que a preferência por cada um dos estádios apresentou perfil semelhante.

Além disso, foram utilizados também dados dos borderôs dos jogos do Fluminense no Maracanã nos Campeonatos Brasileiros de 2013 e 2014. Em 2010, 2011 e 2012, o Fluminense também mandou a maioria dos seus jogos no Estádio Nilton Santos. No entanto, (i) pelo fato de o Fluminense não ser considerado dono do Estádio Nilton Santos, (ii) pela média de público do Fluminense em 2012 (sagrou-se campeão, e jogou no Estádio Nilton Santos) ser menor do que em temporadas nas quais equipe a jogou no Maracanã (com desempenho esportivo inferior, como mostra a Tabela 2) e também (iii) pela falta de pesquisa que compare a preferência do torcedor tricolor em relação aos estádios, apenas consideraremos jogos do clube no Maracanã para estimar a equação. Como explicitado mais adiante, a inclusão de jogos do Fluminense se deve também à política de preços mais variada praticada pelo clube em alguns campeonatos. Ora, uma variação maior dos preços permite uma avaliação mais precisa do impacto do valor do ingresso no ânimo do torcedor para ir ao estádio. A base do Botafogo, infelizmente, apresenta um perfil de preços bem menos variados, o que torna difícil quantificar a influência desse fator no comportamento do torcedor alvinegro.

Tabela 2 - Média de público por estádio e por Campeonato Brasileiro do Fluminense.

Fonte: O autor.

| Parâmetros Fluminense | Brasileiro 2012 | Brasileiro 2013 | Brasileiro 2014 |
|----------------------------|------------------------|------------------------|------------------------|
| Média de Público | 13.992,03 | 22.113,05 | 20.390,59 |
| Estádio | Engenhão | Maracanã | Maracanã |
| Classificação Final | 1º lugar | 17º/15º lugar | 6º lugar |
| Preço Mínimo Médio | R\$ 17,06 | R\$ 31,25 | R\$ 26,47 |

A base de dados do Botafogo tem 90 observações, isto é, 90 jogos: 68 ocorridos em finais de semana (ou feriados) e 22 ocorridos em dias úteis. A base do Fluminense tem 33 observações, sendo 26 dos jogos em finais de semana (ou feriados) e 7 em dias úteis. Os jogos em dias úteis são diferenciados dos outros por meio de uma variável binária auxiliar.

Construída a base de dados, variáveis devem ser escolhidas para tentar explicar a variável dependente: no nosso caso, o público total de Botafoguenses (ou Tricolores) na partida. O software escolhido para fazer a Regressão Linear pelo Método dos Mínimos Quadrados Ordinários (MQO) é o EViews®, pela intimidade do autor com o programa.

3.3 As variáveis

São vários os fatores que influenciam um torcedor a sair de casa para ir (ou não ir) a um estádio de futebol. De forma rápida, podemos imaginar e levantar alguns aspectos que pesam na decisão do torcedor:

- O adversário a ser enfrentado;
- O dia da semana do jogo;
- O horário do jogo;
- O preço do ingresso;
- Se está perto ou não do final do mês (perto do recebimento do pagamento);
- O momento da equipe;
- A fase (início, meio ou fim) do campeonato;
- Se o jogo está sendo transmitido pela TV (ou se o último foi transmitido);
- A previsão e as condições climáticas;
- Uma série de fatores difíceis de serem medidos (como a pura vontade do torcedor de sair de casa ou as condições de “diversão” e acomodação que o estádio oferece).

A variável dependente será, obviamente, o público na partida (no caso, o do Botafogo ou do Fluminense). O borderô quase sempre deixa claro o exato número de pessoas que pagaram ingresso para assistirem ao jogo na torcida do time A ou do time B (público pagante). No entanto, as gratuidades (como idosos) não são divididas por setor e, assim, não podem ser atribuídas a nenhuma das torcidas. Neste trabalho, decidimos aproximar o público presente (soma do público pagante com as gratuidades) do Botafogo ou do Fluminense supondo que a proporção de torcedores do time da casa e do time visitante é a mesma para o público pagante e para as gratuidades. Isto é: se em determinado jogo foram registrados, por exemplo, 1.000 pagantes do Palmeiras e 9.000 pagantes do Botafogo além de 5.000 gratuidades, o público presente estimado do Botafogo será de 13.500 pessoas, já que o Botafogo teve 90% do público pagante da partida $(9.000 + (9.000/10.000)\% \text{ de } 5000 \text{ gratuidades} = 9.000 + 0,9 \cdot 5.000 = 9.000 +$

4.500 = 13.500). Essa variável será chamada (a única variável dependente) de *PubBot* (na estimação para o Botafogo) ou *PubFlu* (na estimação para o Fluminense).

A seguir, serão relatadas as variáveis propostas nesse trabalho.

● Retrospecto recente do time

É esperado que a presença de público seja maior quando o time está bem. Mas como quantificar se um time está bem ou não? Como no Campeonato Brasileiro as vitórias somam 3 pontos à classificação do time, o empate soma 1 e a derrota soma 0, a decisão foi quantificar o momento da equipe pela pontuação dos jogos recentes.

Desta forma, para cada jogo da base de dados, foram computadas as pontuações dos jogos anteriores do Botafogo e do Fluminense, mesmo que os jogos anteriores não fossem todos do Campeonato Brasileiro (Campeonato Carioca, Copa do Brasil, Copa Libertadores e Copa Sul-Americana).

Para decidir a quantidade de jogos que entrariam no modelo, foram calculadas as correlações dos públicos com as pontuações dos 5 jogos anteriores, a fim de estimar até onde iria a “memória” do torcedor. As correlações são apresentadas nas Tabelas 3 e 4.

Tabela 3 - Correlação do público do Botafogo com a pontuação obtida nos 5 jogos anteriores.

Fonte: O autor.

| | 1º jogo Anterior | 2º jogo Anterior | 3º jogo Anterior | 4º jogo Anterior | 5º jogo Anterior |
|--------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Correlação com o público | 0,406 | 0,258 | 0,210 | 0,164 | -0,088 |

A Tabela 3 mostra que, como esperado, a memória do torcedor “diminui” conforme nos afastamos temporalmente, e a correlação cai até o 4º jogo anterior. A correlação torna-se negativa (algo que foge à realidade) no 5º jogo, o que faz com que o modelo seja constituído de modo a utilizar a pontuação dos 4 jogos anteriores. Assim, serão utilizadas as variáveis *R1*, *R2*, *R3* e *R4*, separadamente, tendo, cada uma, valor 3 quando o resultado for vitória, 1 para empate e 0 para derrota. Quando um dos jogos anteriores foi de uma competição eliminatória, consideramos 3 pontos para classificação e 0 para desclassificação, não importando se foi uma vitória, empate ou derrota.

Tabela 4 - Correlação do público do Fluminense com a pontuação obtida nos 5 jogos anteriores.

Fonte: O autor.

| | 1º jogo Anterior | 2º jogo Anterior | 3º jogo Anterior | 4º jogo Anterior | 5º jogo Anterior |
|--------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Correlação com o público | 0,257 | 0,046 | 0,210 | -0,098 | -0,334 |

A Tabela 4 mostra que, como esperado, a memória do torcedor do fluminense também “diminui” conforme nos afastamos temporalmente, e a correlação cai até o 5º jogo anterior. No entanto, a correlação torna-se negativa (algo que foge à realidade) com relação ao 4º e ao 5º jogo, o que faz com que o modelo seja constituído de modo a utilizar a pontuação dos 3 jogos anteriores. Assim, serão utilizadas, para a equação do Fluminense, as variáveis $R1$, $R2$ e $R3$, separadamente, tendo, cada uma, valor 3 quando o resultado for vitória, 1 para empate e 0 para derrota. Assim como foi para o caso do Botafogo, quando um dos jogos anteriores foi de uma competição eliminatória, consideramos 3 pontos para classificação e 0 para desclassificação, não importando se foi uma vitória, empate ou derrota.

● Momentos especiais

As variáveis $R1$, $R2$, $R3$ (e $R4$) não conseguem medir, no entanto, situações ditas especiais como um título conquistado ou a estreia de um grande atleta. Assim, foi criada a variável $MomP$ (momento positivo) que será binária: terá valor 1 quando a situação for especial e tiver impacto positivo na presença (apresentação ou estreia de um grande jogador, ou até um jogo subsequente a um título) e será 0 quando o jogo for normal.

Da mesma forma, foi criada a variável $MomN$ (momento negativo) que será binária: terá valor 1 quando a situação for especial e tiver impacto negativo na presença (jogo subsequente à perda de um título ou uma goleada sofrida) e será 0 quando o jogo for normal.

Na base do Botafogo, em 90 observações, apenas 7 jogos tiveram a variável $MomP$ no valor 1 e 2 jogos com a variável $MomN$ com valor 1. Já na base do Fluminense, que é menor (33 jogos), 2 jogos tiveram a variável $MomP$ valendo 1, e nenhum jogo teve a variável $MomN$ valendo 1.

● Dia da semana

Acredita-se que o dia da partida influencia diretamente na presença de público. Jogos em finais de semana e em feriados atraem mais famílias e idosos do que jogos durante a semana, basicamente pela questão do horário e por ser um dia de folga. Assim, foi criada mais uma variável binária chamada *FDSFer*, que será 1 para jogos em finais de semana ou feriados (68 jogos de 90 observações para o Botafogo; 26 para 33 do Fluminense) e 0 para jogos em dias úteis (22 observações para o Botafogo e 7 para o Fluminense).

● Horário da partida

O horário da partida está intimamente ligado com o dia da semana. Um jogo às 16h de domingo está em um horário “nobre”, mas se fosse às 16h de uma quarta-feira seria certamente um desastre de público: são dois tipos diferentes de jogos, com públicos também distintos. Para tentar medir esse efeito, será proposta uma variável relacionada aos dias da semana e horários de jogos. É notável que os horários de jogo e dias da semana têm uma influência conjunta.

Antes de tudo, no entanto, para todos os jogos da base (90 do Botafogo e 34 do Fluminense), foi verificado que houve jogos em 10 horários distintos. Cruzando esses horários com o dia da semana na partida (isto é: se foi um dia útil ou um dia de final de semana ou feriado), foi convencionado que há 3 classificações: “cedo”, “médio” ou “tarde”. A tabela 5, a seguir, mostra os horários em que foram observadas partidas na base e sua respectiva classificação, caso ela tenha sido num dia útil ou não útil. Células com “-“ indicam que não foram observados jogos nesta combinação de dia/hora na base de dados.

Tabela 5 - Classificação de “cedo”, “médio” ou “tarde” para os jogos das bases de acordo com o horário de início da partida e dia da semana.

Fonte: O autor.

| Horário | Final de semana/Feriado | Dia útil |
|----------------|--------------------------------|-----------------|
| 16:00 | Cedo | - |
| 16:20 | Cedo | - |
| 17:00 | Cedo | - |
| 18:00 | Médio | - |
| 18:30 | Médio | - |
| 19:00 | Médio | - |
| 19:30 | Tarde | Cedo |
| 20:30 | - | Médio |
| 21:00 | Tarde | Médio |
| 21:50 | - | Tarde |
| 22:00 | - | Tarde |

Obviamente, não se pode incluir “cedo” ou “tarde” na modelagem. Assim, será usada uma modelagem binária, composta pelas variáveis *HB1* e *HB2*. Como há três possibilidades qualitativas para o início de uma partida (“cedo”, “médio” ou “tarde”), a combinação das variáveis *HB1* e *HB2* vai traduzir estas classificações da seguinte maneira:

Tabela 6 – Comportamento das variáveis *HB1* e *HB2* de acordo com horário de início

Fonte: O autor.

| Classificação do horário de início da partida | HB1 | HB2 |
|--|------------|------------|
| Cedo | 0 | 1 |
| Médio | 1 | 0 |
| Tarde | 1 | 1 |

Uma partida começando às 19:30 de um sábado (ou seja: segundo a convenção exposta na tabela 5, uma partida considerada “tarde”), *HB1* terá valor 1 e *HB2* valor 1. No entanto, caso uma partida tenha início às 19:30 de um dia útil (“cedo”), *HB1* terá valor 0 e *HB2* valor 1.

● **Distância do dia de pagamento**

O torcedor de futebol muitas vezes é de uma classe menos abastada, o que faz com que muitos torcedores deixem de ir ao jogo caso o mês “já esteja no fim”, isto é, a data do jogo seja longe do seu último recebimento. Há diversas maneiras de (tentar) estimar a “perda de intenção” de ir ao jogo conforme os dias passam ao longo do mês. Foram propostas várias formas de descrever essa variável. Em primeiro lugar, deve-se decidir qual é o início do mês do calendário de “recebimento”: o 1º dia do mês do calendário usual, o 5º dia do mês do calendário usual ou o 5º dia útil. Isto é: quando começará o mês do assalariado?

Depois de definir o alvo, deve-se decidir a taxa de decrescimento da intenção de ir ao jogo em função do dia do mês. A ideia é estimar como o salário vai sendo consumido à medida que o tempo passa. Como opções de modelagem é possível utilizar escala linear ou exponencial, por exemplo. Quando se opta por uma escala linear, com os dias passando como passam no calendário (dia 1, 2, 3...), estaremos assumindo que o nível de renda disponível para o torcedor despender em ingresso (ou o interesse do torcedor em ir ao jogo) cairá linearmente.

Foram testadas 15 variáveis diferentes para modelagem desse efeito, já que foram escolhidas 5 escalas (linear, logarítmica, semanal e 2 exponenciais – uma exponencial do dia e outra exponencial do dia dividido por 30) e 3 formas de começar a contagem do calendário (1º dia do mês, 5º dia do mês e 5º dia útil do mês).

Como são muitas as pessoas não recebem já no 1º dia do mês, é natural utilizar as variáveis do 5º dia ou do 5º dia útil.

Por apresentar a maior correlação prévia e também por apresentar o coeficiente na regressão de maior significância estatística, a variável utilizada será uma de escala logarítmica com o mês começando no dia 5.

A escala logarítmica para um mês de 31 dias se comporta da seguinte maneira (sendo 1 o dia 5, 2 o dia 6, 3 o dia 7, etc.):

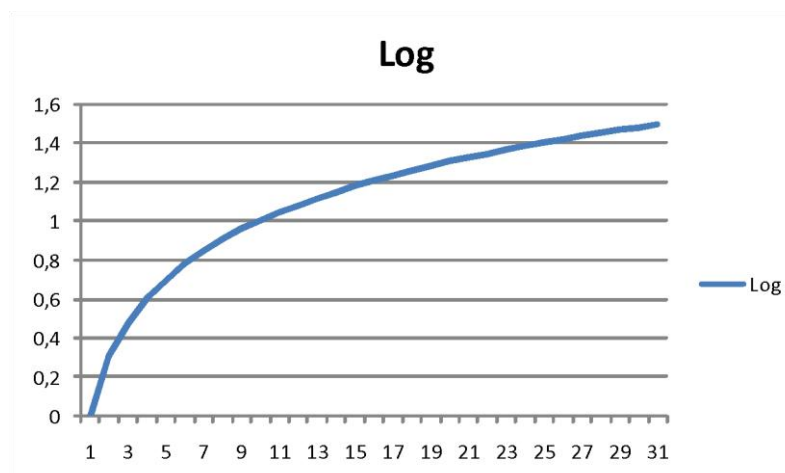


Figura 5 - Comportamento da curva logarítmica de 1 a 31

Fonte: O autor.

Percebe-se, ao se observar a figura 5, que a curva cresce em uma alta taxa nos primeiros dias (basta imaginar a derivada em cada dia), tendendo a uma estabilidade no final do mês. Considerando que a variável tenha coeficiente negativo, então o interesse em ir ao jogo será perdido de maneira rápida, conforme se afasta do dia 5 nos dias próximos (mas posteriores) ao dia 5, e cairá de maneira mais estável a partir daí e até o dia 4 do mês seguinte. Isso faz sentido, imaginando-se que há muitas contas a serem pagas nos dias próximos ao recebimento, sobrando uma porção cada vez menor do salário para o lazer (futebol).

Como exemplo do comportamento da variável criada ($LogD5c_1$), a tabela a seguir mostra possíveis valores para determinadas datas.

Tabela 7 - Comportamento da variável que mede a distância da data da partida para o dia de pagamento

Fonte: O autor.

| Data da partida | Variável “Distância dia 5” | LogD5c_1 |
|------------------------|-----------------------------------|-----------------|
| 05/09/2011 | 1 | 0,000 |
| 06/09/2011 | 2 | 0,301 |
| 07/09/2011 | 3 | 0,477 |
| 08/09/2011 | 4 | 0,602 |
| 09/09/2011 | 5 | 0,699 |
| 10/09/2011 | 6 | 0,778 |
| 11/09/2011 | 7 | 0,845 |
| 12/09/2011 | 8 | 0,903 |
| 13/09/2011 | 9 | 0,954 |
| 14/09/2011 | 10 | 1,000 |
| 15/09/2011 | 11 | 1,041 |

O conhecimento da influência relatada acima é importante, pois permite que o clube faça uma política de precificação dos ingressos que pode variar de acordo com o dia do mês.

● **O preço do ingresso**

É automático imaginar que o preço tem influência na partida. No Brasil, a questão tende a ser ainda mais importante. Como falado, não é comum no Brasil a venda de *season ticket*, como é nos esportes norte-americanos ou no futebol europeu. Consequentemente, não está na cultura do torcedor brasileiro portar o ingresso com antecedência para todas as partidas da temporada (ou do campeonato nacional). Além disso, também é extremamente incomum que clubes brasileiros determinem e/ou divulguem com grande antecedência o preço de ingressos. Como parte de uma iniciativa inovadora, o Fluminense categorizou os adversários do Campeonato Brasileiro de 2014 em 3 tipos (rivais locais, rivais nacionais e sem rivalidade) e anunciou, antes do início do torneio, o valor do ingresso mais barato para cada uma dessas partidas: R\$ 10 contra clubes sem rivalidade, R\$ 20 contra clubes rivais nacionais e R\$ 30 contra clubes rivais

locais⁹. O clube, no entanto, abandonou a medida poucos meses depois, ainda no meio do campeonato.

É possível lembrar também de “pacotes” lançados para os jogos finais de um campeonato¹⁰ ou situações também pontuais, mas talvez o exemplo mais consistente tenha sido o do Corinthians no Brasileiro de 2017¹¹. No entanto, é notável que são raros os exemplos de políticas de preço antecipadas no futebol brasileiro.

Clubes ingleses divulgam antecipadamente, para os sócios ou os não sócios, os preços de ingressos para os jogos em que terão mando de campo. Como exemplo, o Arsenal FC da Inglaterra divulga os preços da temporada¹² de acordo com a categoria/tipo da partida¹³. Seu enorme rival londrino, o Tottenham, também faz o mesmo¹⁴, como mostramos a figura abaixo para a temporada 2017/18 da Premier League (Campeonato Inglês).

| BLOCK SECTIONS | | CATEGORY A | CATEGORY B | CATEGORY C | PACKAGE 1 PRICING | | | | |
|-------------------------|--|---|---|---|------------------------------------|--------|--------|---------|--------|
| | | Arsenal Chelsea Liverpool Manchester City Manchester United West Ham | Crystal Palace Everton Leicester City Newcastle United Southampton Watford | AFC Bournemouth Brighton & Hove Burnley Huddersfield Stoke City Swansea City West Brom Albion | Chelsea Burnley Swansea City | | | | |
| CATEGORY 1 (LEVEL 1) | 102-104, 120-126, 141-143 | £75.00 | £30.00 | £55.00 | £22.00 | £45.00 | £18.00 | N/A | N/A |
| | 104-105, 118-120, 126-127, 140-141 | £70.00 | £28.00 | £50.00 | £20.00 | £40.00 | £16.00 | N/A | N/A |
| CATEGORY 2 (LEVEL 1) | 105-107, 117-118, 127-128, 139-140, 108-112, 129-138 | £54.00 | £22.00 | £45.00 | £18.00 | £35.00 | £14.00 | N/A | N/A |
| | 108-112, 130-137 | £52.00 | £21.00 | £42.00 | £17.00 | £32.00 | £13.00 | N/A | N/A |
| CATEGORY 3 (LEVEL 5) | 501-505, 522-531, 548-552 | £48.00 | £20.00 | £45.00 | £18.00 | £35.00 | £14.00 | £18.00 | £48.00 |
| | 501-505, 522-531, 548-552 505-522, 531-548 | £45.00 | £18.00 | £40.00 | £16.00 | £30.00 | £12.00 | £105.00 | £42.00 |
| CATEGORY 4 (LEVEL 5) | 505-506, 547-548 | £35.00 | £14.00 | £30.00 | £12.00 | £20.00 | £8.00 | N/A | N/A |
| CLUB WEMBLEY GOLD | 202-205, 222-231, 248-251 | £95.00 | £38.00 | £80.00 | £32.00 | £70.00 | £28.00 | N/A | N/A |
| CLUB WEMBLEY SILVER | 206-209, 218-221, 232-235, 244-247 | £75.00 | £30.00 | £65.00 | £26.00 | £55.00 | £22.00 | £185.00 | £74.00 |
| | 209-218, 235-244 | £70.00 | £28.00 | £60.00 | £24.00 | £50.00 | £20.00 | £170.00 | £68.00 |

* Concession prices apply to supporters over 65 and under 18 years of age

Figura 6 - Política de preços da temporada 2017/8 do Tottenham Hotspur (Inglaterra).
Fonte: <http://www.tottenhamhotspur.com/tickets/ticket-prices/>

⁹ <http://globoesporte.globo.com/futebol/times/fluminense/noticia/2014/05/flu-divulga-precos-dos-ingressos-para-partidas-em-casa-pelo-brasileirao.html>

¹⁰ <http://globoesporte.globo.com/futebol/times/flamengo/noticia/2016/10/flamengo-lanca-pacote-para-jogos-com-botafogo-coritiba-e-santos.html>

¹¹ <https://globoesporte.globo.com/futebol/times/corinthians/noticia/corinthians-lanca-pacote-de-ingressos-para-brasileirao-veja-os-precos.ghtml>

¹² <https://www.arsenal.com/tickets/non-member-ticket-prices>

¹³ <https://www.arsenal.com/tickets/matchcategories>

¹⁴ <http://www.tottenhamhotspur.com/tickets/ticket-prices/>

No Brasil, cada jogo tem um preço, e normalmente este é determinado a poucos dias da partida, de acordo com os objetivos do clube ou a importância da partida.

Há diversas possibilidades de se tentar construir uma variável que capte o efeito do aumento, da manutenção ou mesmo da diminuição do preço do ingresso. Com diversos preços para diferentes setores em um mesmo estádio, qual deles utilizar?

Apesar de parecer óbvio e imediato pegar o popular “ticket médio” da partida (renda bruta dividida pelo público pagante), esta variável teria um viés, já que se deseja, justamente, ter o público como variável explicada.

Por conta da modelagem envolver diferentes estádios (inviabilizando, assim, a aplicação de um preço ponderado por setor), será escolhido o preço mínimo da partida. Isto é: para cada um dos 90 jogos do Botafogo e 33 jogos do Fluminense, foi registrado o ingresso de inteira mais barato vendido em cada partida.

A ideia é que o ingresso mais barato seja, talvez, aquele que mais influa na decisão de ir ou não ao jogo: caso o menor preço seja visto como restritivo para um grupo de torcedores, certamente os ingressos mais caros serão, no mínimo, tão restritivos quanto, podendo restringir ainda mais torcedores.

A Figura 7 mostra, na base de dados do Botafogo, um histograma de preços mínimos.

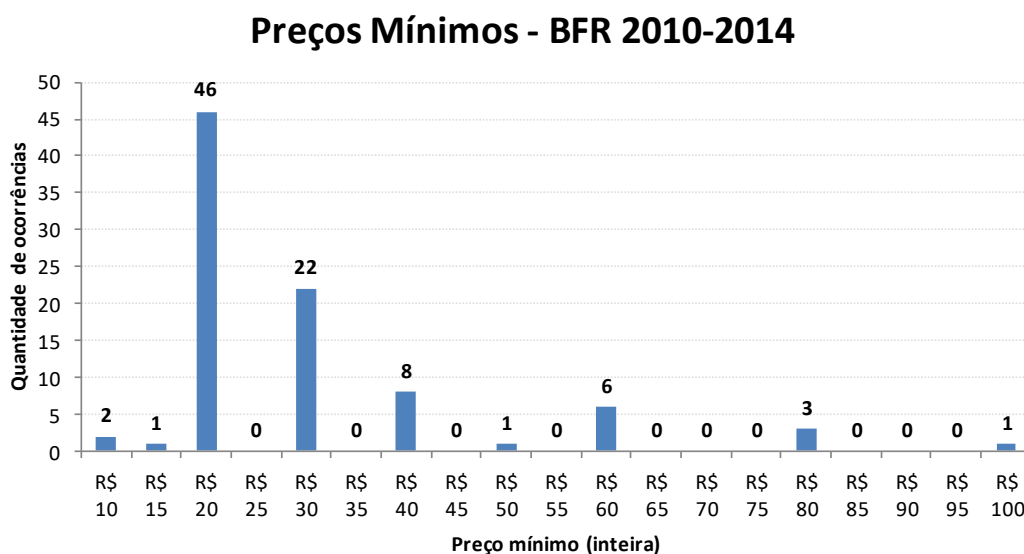


Figura 7 - Histograma com preços mínimos praticados nos jogos do Botafogo FR

Fonte: O Autor

Observa-se na Figura 7 como a grande maioria (76% da base) dos jogos tiveram preço mínimo entre R\$ 20 e R\$ 30. Com tão baixa variação nessa base grande, pode-se esperar que o preço não seja uma variável significativa na equação, já que não houve variação suficiente para se medir sua influência.

Já a Figura 8, a seguir, mostra o histograma de preços mínimos nos jogos da base do Fluminense.

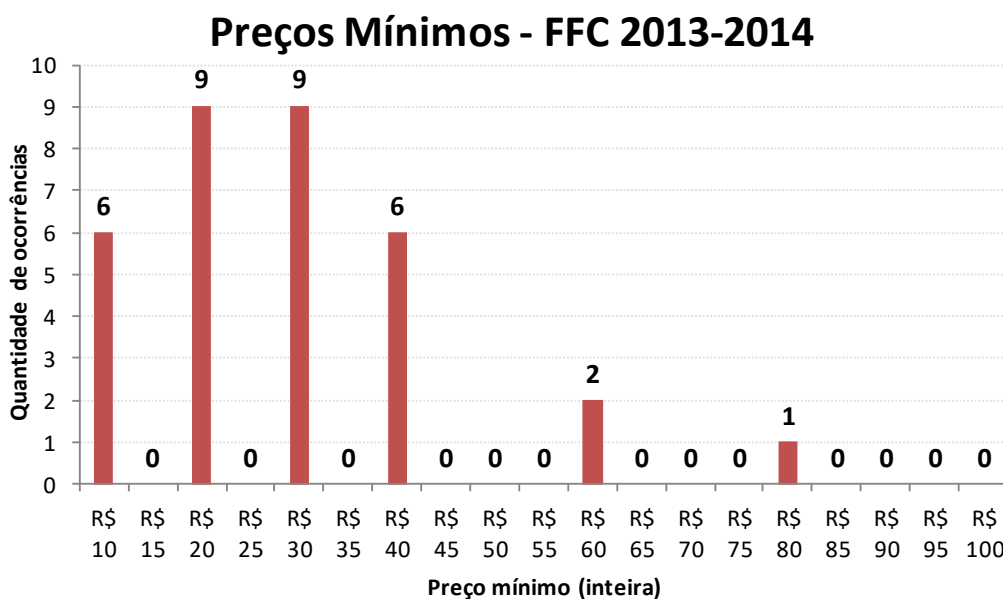


Figura 8 - Histograma com preços mínimos praticados nos jogos do Fluminense FC.

Fonte: O Autor

O histograma de preços do Fluminense é mais bem distribuído. 55% dos jogos tiveram preço de ingresso mais baixo entre R\$ 20 e R\$ 30, contra 76% da base do Botafogo. Se analisarmos de R\$ 10 a R\$ 30, o número aumenta para 73%: número idêntico se olharmos de R\$ 20 até R\$ 40. Desse modo, é esperada maior significância para este parâmetro na equação do Fluminense.

Como a observação de preços dá-se num intervalo amplo (uma diferença de aproximadamente 4 anos e meio entre a data do 1º e do último jogo da base do Botafogo: 08/05/2010 – 15/11/2014), é preciso corrigir o valor do dinheiro ao longo do tempo. É de se esperar que, por efeitos como a inflação, R\$ 1 em maio de 2010 tenha poder de compra diferente que a mesma quantia em novembro de 2014. Ou seja: faz-se necessário registrar todos os preços mínimos tomados nas 2 bases e colocá-los sob um mesmo padrão.

Com base no Índice de Preços ao Consumidor Amplo (IPCA), indicador econômico do Instituto Brasileiro de Geografia e Estatística (IBGE), todos os preços (90 da base do Botafogo e 33 da base do Fluminense) foram inflacionados para a base Janeiro/2015.

Assim, finalmente, a variável preço utilizada na regressão será o preço mínimo (de ingresso de inteira) da partida inflacionado a Janeiro/2015: *PMinJan15*.

• Importância do jogo

Um jogo é dito importante quando o time tem objetivos claros e bem definidos em um campeonato. Isto é: quando este time busca (e tem condições de) ganhar o campeonato ou também alcançar classificação para competições internacionais (como a Taça Libertadores). E quanto mais importante for um jogo para o time da casa (aqui o Botafogo e/ou o Fluminense), maior será, espera-se, o público no estádio.

Acredita-se também que os jogos vão se tornando mais importantes conforme o campeonato torna-se mais “agudo”, isto é, mais próximo do término (aqui, cabe lembrar também que o Campeonato Brasileiro tem 38 rodadas). Assim, adapta-se a utilização de uma das fórmulas propostas por Kuypers (1996), que relaciona o momento do campeonato com a situação do time frente aos outros.

A 2ª fórmula proposta por Kuypers (1996) é:

- $(PB \times GL)$, sendo PB = pontos atrás do líder (*points behind*) e GL = número de jogos que faltam (*games left*).

Quando o time for o líder da competição, o indicador é forçado ao resultado 1 (aqui considera-se que o time será líder sempre que $PB = 0$, ainda que a liderança não esteja sendo alcançada por critérios de desempate). Quando não há mais chance matemática de título, o resultado é arbitrado em 0 (zero). Kuypers (1996) argumenta que a fórmula pode ser adaptada também para a fuga do rebaixamento, considerando então a diferença de pontos para a zona do rebaixamento.

Propõe-se, assim, a variável *ImpJogo1*, uma aplicação da fórmula descrita acima, mas apenas com foco na liderança (e não para o rebaixamento). Vale destacar que quanto menor for a variável, mais importante será a partida, e espera-se um impacto positivo no público.

● O adversário

Se na variável anterior o foco era o time mandante, agora cabe também observar o time visitante, o adversário.

Espera-se que adversários melhores atraiam mais público, o que valoriza confrontos entre bons jogadores (e até a chamada incerteza do resultado, ou *Uncertainty of Outcome*, muito citada nos trabalhos pesquisados). Alguns trabalhos, como os de Kuypers (1996) e Souza (2004), contabilizam o número de jogadores com passagens por seleções nacionais envolvidos no jogo, inclusive do time adversário. Pode-se, contudo, argumentar que essa é uma realidade agora distante do futebol brasileiro, devido ao baixo número de jogadores de times brasileiros em seleções nacionais de futebol, e por isso deve-se estimar a qualidade do time adversário de outra forma.

Tentando captar o aspecto histórico dos clubes, e entendendo que o que pode fazer um torcedor comparecer (ou deixar de ir) a uma partida, no que diz respeito ao adversário, é a rivalidade entre seu clube e os demais, este trabalho propõe a divisão os adversários em 3 categorias distintas:

- Rivais Locais (mesma cidade/região);
- Rivais Nacionais;
- Não Rivais.

Como mencionado anteriormente, clubes ingleses (como Tottenham, vide figura 6, e Arsenal) categorizam seus adversários na temporada de modo semelhante na hora de precificar seus ingressos. Da mesma maneira, o Fluminense já categorizou suas partidas também em 3 maneiras (rivais locais, rivais nacionais e demais clubes), como também já mencionado.

Por conta da rivalidade local, espera-se que jogos ditos “clássicos” (contra Flamengo, Fluminense e Vasco) atraiam mais torcedores do Botafogo, assim como é esperado que jogos contra Botafogo, Flamengo e Vasco atraiam mais torcedores do Fluminense. Além disso, claro, jogos desse tipo atraem mais atenção da mídia e da imprensa, que provê em geral maior cobertura prévia da partida, e assim acaba por promover o jogo, gerando um impacto (esperado) positivo no público.

Considerando o conceito de “doze grandes clubes brasileiros”, ideia amplamente difundida na imprensa e entre os torcedores, que diz que Botafogo/RJ, Flamengo/RJ, Fluminense/RJ, Vasco da Gama/RJ, Corinthians/SP, Palmeiras/SP, São Paulo/SP,

Santos/SP, Atlético Mineiro/MG, Cruzeiro/MG, Grêmio/RS e Internacional/RS contemplam o maior peso histórico do futebol nacional (títulos, ídolos e torcedores), pode-se considerar que rivais nacionais (para Botafogo e Fluminense) serão os 4 grandes de São Paulo (Corinthians, Palmeiras, São Paulo e Santos), os 2 grandes de Minas Gerais (Atlético-MG e Cruzeiro) e os 2 grandes do Rio Grande do Sul (Grêmio e Internacional). Há que se destacar que esses 8 clubes sempre figuram entre as maiores torcidas nacionais¹⁵, o que também provê efeito na atenção e tipo de cobertura que a imprensa oferece para as partidas.

A tabela abaixo mostra que as 90 partidas observadas do Botafogo apresentaram 29 adversários diferentes, que foram categorizados entre rivais locais, rivais nacionais e não rivais:

Tabela 8 - Classificação dos rivais do Botafogo, na base observada, quanto à rivalidade.

Fonte: O autor

| Rivais - Base Botafogo | | |
|-------------------------------|-------------------------|----------------------|
| Não Rivais | Rivais Nacionais | Rivais Locais |
| América-MG | Atlético-MG | Flamengo |
| Atlético-GO | Corinthians | Fluminense |
| Atlético-PR | Cruzeiro | Vasco |
| Avaí | Grêmio | |
| Bahia | Internacional | |
| Ceará | Palmeiras | |
| Chapecoense | Santos | |
| Coritiba | São Paulo | |
| Criciúma | | |
| Figueirense | | |
| Goiás | | |
| Guarani | | |
| Náutico | | |
| Ponte Preta | | |
| Portuguesa | | |
| Prudente | | |
| Sport | | |
| Vitória | | |

Para o Fluminense, foram 33 observações e 22 clubes diferentes, divididos da seguinte maneira:

¹⁵ <http://blogs.diariodepernambuco.com.br/esportes/2016/12/25/levantamento-do-parana-pesquisas-com-10-mil-entrevistados-aponta-6-clubes-do-nordeste-com-torcidas-acima-de-1-milhao/>

Tabela 9 - Classificação dos rivais do Fluminense, na base observada, quanto à rivalidade.

Fonte: O autor

| Rivais - Base Fluminense | | |
|---------------------------------|-------------------------|----------------------|
| Não Rivais | Rivais Nacionais | Rivais Locais |
| Atlético-PR | Atlético-MG | Botafogo |
| Bahia | Corinthians | Flamengo |
| Chapecoense | Cruzeiro | Vasco |
| Coritiba | Grêmio | |
| Criciúma | Palmeiras | |
| Figueirense | Santos | |
| Goiás | São Paulo | |
| Náutico | | |
| Ponte Preta | | |
| Portuguesa | | |
| Sport | | |
| Vitória | | |

Mais uma vez, não se pode incluir “Rival Local” ou “Não Rival” na modelagem. Desta forma, mais uma será usada uma variável binária, composta pelas variáveis RB1 e RB2. Como há três possibilidades qualitativas para o grau de rivalidade (“Rival Local”, “Rival Nacional” ou “Não Rival”), a combinação das variáveis *RB1* e *RB2* vai traduzir estas classificações da seguinte maneira:

Tabela 10 - Comportamento das variáveis RB1 e RB2 de acordo com o grau de rivalidade com o adversário.

Fonte: O autor

| Classificação do adversário quanto à rivalidade | RB1 | RB2 |
|--|------------|------------|
| Não Rival | 0 | 1 |
| Rival Nacional | 1 | 0 |
| Rival Local | 1 | 1 |

Exemplificando, para uma partida da base do Botafogo contra o Santos (ou seja: segundo a convenção exposta na tabela 8, uma partida considerada contra um “Rival Nacional”), *RB1* terá valor 1 e *RB2* valor 0. No entanto, caso uma partida do Fluminense seja contra o Náutico (“Não Rival”), *RB1* terá valor 0 e *RB2* valor 1.

Por fim, para facilitar o entendimento, a tabela 11 resumirá todas as 14 variáveis dependentes construídas, com seus nomes e os efeitos que buscam captar.

Tabela 11 - Resumo das variáveis dependentes construídas para os modelos de previsão.

Fonte: O autor

| Nome da Variável | Tipo da Variável |
|-------------------------|---|
| FDSFER | Dia da semana - dia útil ou não (binária) |
| HB1 | Classificação do horário de início da partida (binária) |
| HB2 | Classificação do horário de início da partida (binária) |
| MOMP | Momento positivo do clube mandante analisado (binária) |
| MOMN | Momento negativo do clube mandante analisado (binária) |
| RB1 | Classificação do adversário quanto à rivalidade (binária) |
| RB2 | Classificação do adversário quanto à rivalidade (binária) |
| R1 | Resultado do jogo anterior |
| R2 | Resultado do 2º jogo anterior |
| R3 | Resultado do 3º jogo anterior |
| R4 | Resultado do 4º jogo anterior |
| PMINJAN15 | Preço mínimo da partida (inflacionados a janeiro/2015) |
| IMPJOGO1 | Importância da partida |
| LOGD5C_1 | Log da distância para dia 5 do mês |

3.4 O modelo e os resultados

Para o modelo do Botafogo, as 90 observações com as respectivas 14 variáveis dependentes (descritas no item anterior) e a variável independente (PubBot) foram inseridas no programa EViews®, bastante utilizado para implementar modelos de regressão linear. Para o modelo do Fluminense, o mesmo foi feito com as 33 observações, as 11 variáveis dependentes (em relação ao modelo do Botafogo, não entraram as variáveis R4 e MomN).

Pelo Método dos Mínimos Quadrados Ordinários (MQO), obtém-se o seguinte resultado para o modelo do Botafogo, retirado do programa EViews®:

Tabela 12 - Resultados da Regressão Linear para o modelo do Botafogo utilizando o MQO.

Fonte: O autor

Dependent Variable: PUBBOT

Method: Least Squares

Date: 04/02/18 Time: 01:06

Sample: 1 90

Included observations: 90

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|--------|
| C | 15907.80 | 3855.395 | 4.126115 | 0.0001 |
| FDSFER | 3218.927 | 1416.466 | 2.272506 | 0.0259 |
| HB1 | -5371.095 | 1833.000 | -2.930221 | 0.0045 |
| HB2 | -4291.171 | 1862.383 | -2.304129 | 0.0240 |
| MOMP | 11465.65 | 2410.079 | 4.757377 | 0.0000 |
| MOMN | -3646.923 | 4181.465 | -0.872164 | 0.3859 |
| RB1 | -4753.770 | 1698.420 | -2.798936 | 0.0065 |
| RB2 | -3513.916 | 1678.902 | -2.092984 | 0.0397 |
| R1 | 1772.009 | 453.3685 | 3.908539 | 0.0002 |
| R2 | 393.0032 | 474.4655 | 0.828307 | 0.4101 |
| R3 | 1173.352 | 460.4030 | 2.548533 | 0.0129 |
| R4 | 1037.379 | 476.7761 | 2.175820 | 0.0327 |
| PMINJAN15 | 25.90783 | 36.24637 | 0.714770 | 0.4770 |
| IMPJOGO1 | -6.310396 | 4.993382 | -1.263752 | 0.2102 |
| LOGD5C_1 | -1554.177 | 1687.068 | -0.921230 | 0.3599 |
| R-squared | 0.607684 | Mean dependent var | 12809.79 | |
| Adjusted R-squared | 0.534452 | S.D. dependent var | 7955.778 | |
| S.E. of regression | 5428.312 | Akaike info criterion | 20.18766 | |
| Sum squared resid | 2.21E+09 | Schwarz criterion | 20.60429 | |
| Log likelihood | -893.4445 | Hannan-Quinn criter. | 20.35567 | |
| F-statistic | 8.298046 | Durbin-Watson stat | 1.364116 | |
| Prob(F-statistic) | 0.000000 | | | |

Observa-se na tabela 12 que o coeficiente R^2 (0,608) e o R^2 ajustado (0,534) são razoáveis, se levarmos em conta a complexidade do fenômeno estudado e a quantidade de fatores que o explicam. O futebol tem aspectos econômicos, esportivos, culturais e sociais, o que deixa clara a dificuldade em se estabelecer “leis” que descrevam completamente o comportamento do torcedor.

Neste ponto, deve-se lembrar que é prudente avaliar o grau de explicação da variável dependente (*PubBot*) por meio do coeficiente R^2 ajustado, que penaliza a inclusão de variáveis. Poder-se-ia aqui incluir uma série de variáveis, que por mais que fossem correlacionadas com demais variáveis ou nada tivessem a ver com o futebol brasileiro, certamente aumentariam o coeficiente R^2 , gerando uma falsa impressão de melhor explicação. A questão é que isso pode gerar relações espúrias entre as novas

variáveis e a variável explicada, o que certamente não é objetivo. Assim, é prudente basear-se no coeficiente R^2 ajustado.

Sobre os sinais dos coeficientes, vários estão de acordo com o imaginado (por exemplo: o sinal positivo do coeficiente de $R1$ diz que quanto maior o coeficiente, maior será o público, desde que mantidas as outras variáveis constantes – o que faz sentido, considerando que é esperado maior público com o time ganhando o jogo anterior). No entanto, alguns coeficientes estão com sinal diferente do esperado:

$PMINJAN15$ está com sinal positivo: esperava-se um sinal negativo, pois convém interpretar que o torcedor, enquanto consumidor, tenha preferência por preços menores. Assim, pode-se supor que o aumento de preço fosse algo que afastasse o torcedor e a diminuição de preço algo que o aproximasse do estádio. Todavia, o sinal pode ser esperado em função da correlação do preço com outras variáveis do modelo. Por exemplo, clássicos contra rivais locais costumam atrair maiores públicos e são historicamente mais caros.

Quanto à significância das variáveis, são dignas de destaque as variáveis $FDSFer$, $HB1$, $HB2$, $MomP$, $Class$, $RB1$, $RB2$, $R1$, $R3$ e $R4$, todas estatisticamente significativas ao nível de 5%, pois seus coeficientes têm respectivos valores-p menores do que 0,05. As demais variáveis, no entanto, não alcançaram significância estatisticamente suficiente.

Pelo fato de a base, apesar de não ser pequena (90 observações em 5 anos), ter jogos com pouca variação entre si no preço mínimo do ingresso, como visto anteriormente e explicitado na figura 7, a variável que mede o efeito do preço ($PMINJAN15$) apresentou valor-p alto (0,477), e não será utilizada nos próximos passos do modelo. Esse resultado já era esperado, como mencionado anteriormente, devido à baixa variabilidade do parâmetro na base amostral.

E, mais uma vez, pelo Método dos Mínimos Quadrados Ordinários (MQO), chega-se ao seguinte resultado para o modelo do Fluminense, retirado do programa EViews®:

Tabela 13 - Resultados da Regressão Linear para o modelo do Fluminense utilizando o MQO.

Fonte: O autor.

Dependent Variable: PUBFLU

Method: Least Squares

Date: 04/02/18 Time: 01:09

Sample: 1 33

Included observations: 33

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|--------|
| C | 18953.07 | 8323.397 | 2.277084 | 0.0339 |
| FDSFER | 1853.832 | 3382.162 | 0.548120 | 0.5897 |
| HB1 | -1172.571 | 3537.381 | -0.331480 | 0.7437 |
| HB2 | -929.9497 | 4271.156 | -0.217728 | 0.8298 |
| MOMP | 24665.90 | 6088.907 | 4.050957 | 0.0006 |
| RB1 | -7389.895 | 5857.483 | -1.261616 | 0.2216 |
| RB2 | -1854.122 | 5321.650 | -0.348411 | 0.7312 |
| R1 | 2998.151 | 1217.492 | 2.462564 | 0.0230 |
| R2 | 765.9763 | 1115.634 | 0.686584 | 0.5002 |
| R3 | -74.33504 | 1131.247 | -0.065711 | 0.9483 |
| PMINJAN15 | -158.0569 | 130.8236 | -1.208168 | 0.2411 |
| IMPJOGO1 | -7.846283 | 13.32582 | -0.588803 | 0.5626 |
| LOGD5C_1 | 7278.510 | 4189.123 | 1.737478 | 0.0977 |
| R-squared | 0.760391 | Mean dependent var | 21225.72 | |
| Adjusted R-squared | 0.616626 | S.D. dependent var | 11041.02 | |
| S.E. of regression | 6836.288 | Akaike info criterion | 20.78498 | |
| Sum squared resid | 9.35E+08 | Schwarz criterion | 21.37451 | |
| Log likelihood | -329.9522 | Hannan-Quinn criter. | 20.98334 | |
| F-statistic | 5.289124 | Durbin-Watson stat | 2.098953 | |
| Prob(F-statistic) | 0.000552 | | | |

Observa-se na Tabela 13 que o coeficiente R^2 (0,760) e o R^2 ajustado (0,617) são bons levando-se em conta a complexidade do fenômeno estudado, como no caso anterior. Neste modelo, na verdade, os resultados são ainda melhores.

Em relação aos sinais dos coeficientes, poder-se-ia esperar que o coeficiente de *LogD5c_1* fosse um número negativo: isto é, com as demais variáveis constantes, a variável dessa teria efeito negativo no público conforme crescesse. Por outro lado, esperava-se sinais positivos no coeficiente de *R3*, que está com sinais negativos.

A 5% de significância, apenas as variáveis *Momp* e *R1* são significativas. A variável *LOGD5C_1* é significativa a 10% de significância (valor-p de 0,098), e merece destaque.

Comparando os 2 resultados, observa-se que, apesar de o modelo do Fluminense ter encontrado R^2 maior (e R^2 ajustado também), indicando que o modelo conseguiu explicar mais a variação (o modelo se ajustou melhor à amostra), o modelo do Botafogo

teve muito mais variáveis significativas. Ou seja: o modelo do Botafogo tende a nos dar maior noção sobre os fatores que importam na decisão do torcedor de comparecer ou não ao estádio.

Para avaliar a eficácia dos modelos nas estimações, cada uma das bases será dividida em 2 grupos: 80% dos dados serão selecionados para compor uma nova base de dados, enquanto os 20% restantes serão jogos a serem previstos. Isto é: dos 90 jogos observados do Botafogo, os primeiros 72 (exatamente 80% da base) servirão de base para uma nova equação de estimação, e o resultado da previsão será comparado com os dados efetivamente observados nos 18 jogos restantes (20% da base original de 90). Ao final, deve-se observar e avaliar os erros de previsão. Para a base do Fluminense, 26 das 33 observações (78,79%) servirão de parâmetros no intuito de prever o público dos 7 jogos restantes (21,21%).

Tabela 14 - Resultados da Regressão Linear para o Botafogo com apenas 80% da base utilizando o MQO.

Fonte: O autor

Dependent Variable: PUBBOT
 Method: Least Squares
 Date: 04/02/18 Time: 01:08
 Sample: 1 72
 Included observations: 72

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|--------|
| C | 13088.72 | 4216.493 | 3.104171 | 0.0029 |
| FDSFER | 2954.005 | 1694.830 | 1.742951 | 0.0864 |
| HB1 | -6112.571 | 2305.724 | -2.651042 | 0.0102 |
| HB2 | -5098.925 | 2365.943 | -2.155134 | 0.0351 |
| MOMP | 10068.48 | 2827.470 | 3.560951 | 0.0007 |
| RB1 | -3718.427 | 1914.981 | -1.941756 | 0.0568 |
| RB2 | -3309.182 | 1926.172 | -1.718010 | 0.0909 |
| R1 | 2362.489 | 541.9811 | 4.358989 | 0.0001 |
| R2 | 608.6851 | 581.7502 | 1.046300 | 0.2996 |
| R3 | 1174.616 | 534.3694 | 2.198134 | 0.0317 |
| R4 | 1012.717 | 581.3424 | 1.742031 | 0.0865 |
| R-squared | 0.566144 | Mean dependent var | 13228.42 | |
| Adjusted R-squared | 0.495020 | S.D. dependent var | 8231.065 | |
| S.E. of regression | 5849.155 | Akaike info criterion | 20.32575 | |
| Sum squared resid | 2.09E+09 | Schwarz criterion | 20.67357 | |
| Log likelihood | -720.7268 | Hannan-Quinn criter. | 20.46422 | |
| F-statistic | 7.959963 | Durbin-Watson stat | 1.303762 | |
| Prob(F-statistic) | 0.000000 | | | |

É importante deixar claro que apenas as variáveis significativas (ou as dignas de destaque) serão utilizadas nessas novas equações. Ou seja: para o modelo do Botafogo, serão utilizadas as variáveis FDSFer, HB1, HB2, MomP, RB1 RB2, R1, R2, R3 e R4. Apesar da variável R2 ter apresentado coeficiente com valor-p de apenas 0,410, não faria sentido usar R1, R3 e R4, imaginando que o torcedor lembra-se mais de R3 ou R4 do que R2.

Para o Fluminense, serão usadas as variáveis *MomP*, *R1* e *LOGD5C_1*. Neste momento, já se pode notar que as variáveis escolhidas para o modelo do Fluminense são menos significativas que as do modelo do Botafogo. O resultado da nova regressão com a nova base (72 jogos) do Botafogo está na Tabela 14. A Tabela 15 traz os resultados referentes aos jogos do Fluminense.

Tabela 15 - Resultados da Regressão Linear para o Fluminense com apenas 80% da base utilizando o MQO.

Fonte: O autor

Dependent Variable: PUBFLU
 Method: Least Squares
 Date: 04/02/18 Time: 01:20
 Sample: 1 26
 Included observations: 26

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|--------|
| C | 10533.31 | 6267.644 | 1.680586 | 0.1070 |
| MOMP | 21320.98 | 6486.845 | 3.286802 | 0.0034 |
| R1 | 5720.771 | 1351.433 | 4.233114 | 0.0003 |
| LOGD5C_1 | 4290.447 | 5463.837 | 0.785244 | 0.4407 |
| R-squared | 0.555984 | Mean dependent var | 22711.75 | |
| Adjusted R-squared | 0.495436 | S.D. dependent var | 11425.92 | |
| S.E. of regression | 8116.135 | Akaike info criterion | 20.98173 | |
| Sum squared resid | 1.45E+09 | Schwarz criterion | 21.17529 | |
| Log likelihood | -268.7625 | Hannan-Quinn criter. | 21.03747 | |
| F-statistic | 9.182572 | Durbin-Watson stat | 2.340923 | |
| Prob(F-statistic) | 0.000394 | | | |

É possível, agora, confrontar os resultados obtidos pela estimativa dos modelos construídos com a realidade. A Tabela 16 mostra, para cada um dos 18 jogos testados, o público real (PUBBOT), o público estimado através dos coeficientes obtidos (mostrados na Tabela 14), a diferença absoluta e a diferença percentual.

Analisando os resultados obtidos, chega-se a alguns resultados bastante satisfatórios. 7 dos 18 resultados (aproximadamente 39%) estimados têm erros menores que 10% frente ao público real (em módulo). A média dos erros é de -17,82%, e o fato de ser negativa mostra que o modelo tende a errar o público para baixo (subestimando-o). Enquanto isso, a média dos erros absolutos (ou o módulo dos erros) é de 28,18%. Isto é: em média, o modelo errou 28,18% (para mais ou para menos).

A Figura 9 ilustra o bom desempenho do modelo de estimação construído frente ao público observado nos jogos do Botafogo.

Tabela 16 - Diferença nos resultados do modelo de estimativa do Botafogo para a realidade

Fonte: O autor

| Data da Partida | PUBBOT | Estimado | Diferença | Erro % | Erro Absoluto |
|-----------------|-----------|-----------|------------|---------|---------------|
| 05/10/2013 | 12.990,54 | 9.748,83 | -3.241,71 | -24,95% | 24,95% |
| 13/10/2013 | 8.665,81 | 11.164,62 | 2.498,82 | 28,84% | 28,84% |
| 20/10/2013 | 8.893,13 | 8.252,44 | -640,68 | -7,20% | 7,20% |
| 26/10/2013 | 8.747,78 | 9.858,56 | 1.110,78 | 12,70% | 12,70% |
| 13/11/2013 | 10.223,54 | 7.190,81 | -3.032,73 | -29,66% | 29,66% |
| 16/11/2013 | 13.517,61 | 6.922,68 | -6.594,93 | -48,79% | 48,79% |
| 08/12/2013 | 33.984,28 | 22.848,35 | -11.135,93 | -32,77% | 32,77% |
| 27/04/2014 | 8.728,35 | 8.238,09 | -490,26 | -5,62% | 5,62% |
| 10/05/2014 | 4.910,47 | 2.130,73 | -2.779,75 | -56,61% | 56,61% |
| 27/07/2014 | 13.836,55 | 14.202,77 | 366,23 | 2,65% | 2,65% |
| 02/08/2014 | 8.338,01 | 9.050,49 | 712,48 | 8,54% | 8,54% |
| 23/08/2014 | 18.136,88 | 9.459,74 | -8.677,14 | -47,84% | 47,84% |
| 31/08/2014 | 13.321,83 | 12.089,57 | -1.232,26 | -9,25% | 9,25% |
| 17/09/2014 | 4.660,73 | 1.606,19 | -3.054,54 | -65,54% | 65,54% |
| 25/09/2014 | 6.978,20 | 7.043,10 | 64,90 | 0,93% | 0,93% |
| 28/09/2014 | 10.689,36 | 14.921,52 | 4.232,16 | 39,59% | 39,59% |
| 08/10/2014 | 7.988,49 | 7.309,51 | -678,98 | -8,50% | 8,50% |
| 15/11/2014 | 5.823,69 | 1.327,46 | -4.496,23 | -77,21% | 77,21% |

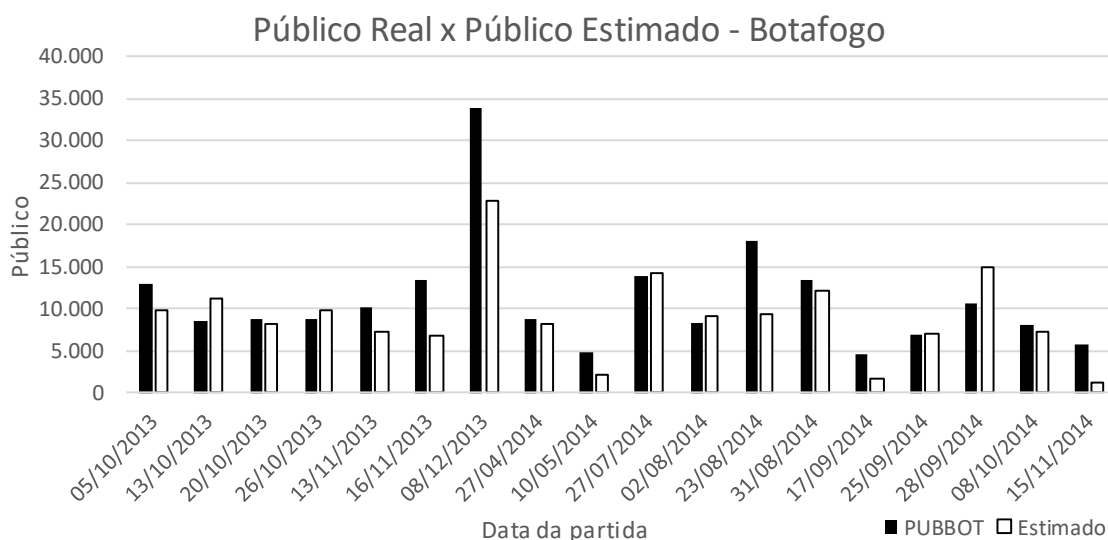


Figura 9 - Comparação entre o público estimado e o público real para 18 jogos do Botafogo.

Fonte: O autor.

A próxima tabela, de número 17, mostra, para cada um dos 7 jogos na amostra de validação, o público real dos jogos de validação do Fluminense (PUBFLU), o público

estimado com os coeficientes obtidos (mostrados na Tabela 15), a diferença absoluta e a diferença percentual.

Percebe-se grande diferença de qualidade na estimativa para o Fluminense. Talvez pela base menor, mas certamente influenciada pela baixa significância estatística dos coeficientes obtidos, apenas 1 dos 7 (14,29%) públicos estimados teve erro menor que 10% em relação ao público real. A média dos erros é de 66,69% (sendo positiva, mostra que este modelo tendo a errar os públicos para mais, superestimando-os), enquanto a média dos erros absolutos (a média dos módulos dos erros) é de 67,39%. Isto é: em média, o modelo errou 67,39% (para mais ou para menos).

Tabela 17 - Diferença nos resultados do modelo de estimativa do Fluminense para a realidade.

Fonte: O autor

| Data da partida | PUBFLU | Estimado | Diferença | Erro % | Erro Absoluto |
|-----------------|-----------|-----------|-----------|---------|---------------|
| 24/09/2014 | 5.967,79 | 22.096,51 | 16.128,71 | 270,26% | 270,26% |
| 09/10/2014 | 11.139,78 | 18.301,15 | 7.161,37 | 64,29% | 64,29% |
| 18/10/2014 | 13.165,65 | 14.627,44 | 1.461,79 | 11,10% | 11,10% |
| 25/10/2014 | 22.273,90 | 32.861,84 | 10.587,94 | 47,54% | 47,54% |
| 15/11/2014 | 16.784,31 | 16.375,73 | -408,57 | -2,43% | 2,43% |
| 20/11/2014 | 29.186,15 | 33.836,82 | 4.650,67 | 15,93% | 15,93% |
| 30/11/2014 | 11.425,76 | 18.301,15 | 6.875,39 | 60,17% | 60,17% |

A Figura 10 mostra a diferença do público estimado e do público real para cada um dos 7 jogos testados do Fluminense.

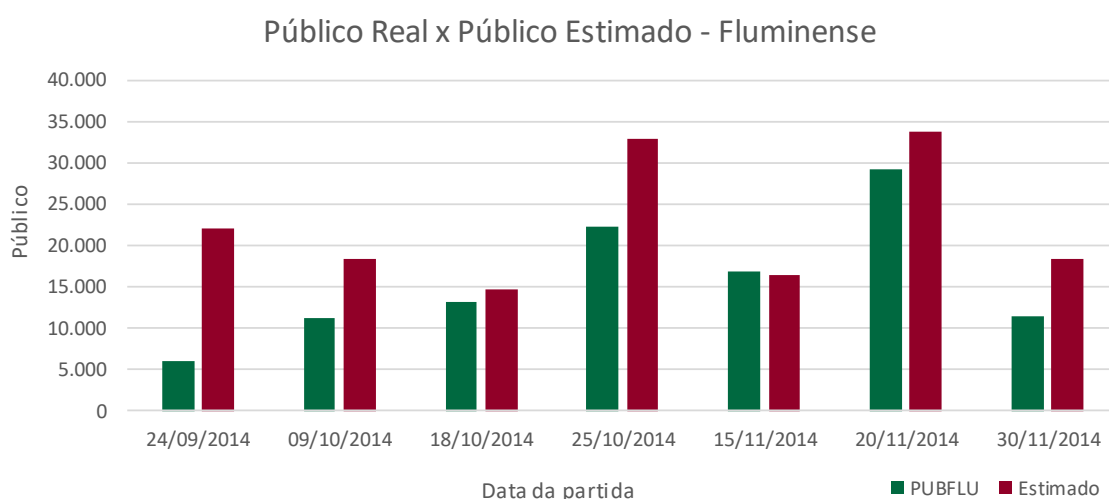


Figura 10 - Comparação entre o público estimado e o público real para os 7 jogos do Fluminense.

Fonte: O autor

4. Discussão

Conseguir explicar mais da metade da variação de público nos jogos é deveras satisfatório. Sendo um fenômeno social, pouco explorado pela matemática e pela estatística, conseguir construir um modelo que chegou a prever resultados quase exatos (para o modelo do Botafogo), é possível acreditar que este trabalho pode ter utilidade prática no futebol. A construção de uma base maior pode trazer resultados ainda mais surpreendentes.

Há de se encarar os resultados do trabalho não apenas em termos do modelo e das equações de estimação. O modelo pode também ser visto como uma ferramenta que pode nos mostrar a sensibilidade do torcedor frente a diversos fatores relevantes. Foi possível entender, por exemplo, que o botafoguense se motiva bastante com os resultados anteriores, que o horário da partida tem grande influência na decisão de comparecer ao estádio, bem como o dia da semana. Além disso, clássicos regionais também se apresentam como chamarizes de público.

Ainda assim, precisamos deixar claro que não será possível, nunca, estimar a presença apenas com base em uma equação ou modelo matemático, por mais preciso que este seja. Fatores que influenciam a decisão de ir ou não a um estádio como a (falta de) sensação de segurança ou problemas em transporte público são dificilmente captados com base de dados, a não ser que se capture dados qualitativos como em pesquisas de opinião. Existem, além disso, fatores imprevisíveis, como as condições de tempo e trânsito na hora da partida, por exemplo.

Outro fator que certamente tem peso, mas que não pode ser estimado (até pelo fato de ter estado ausente em todas as observações) é a venda de bebida alcoólica, que esteve proibida dentro dos estádios do Rio de Janeiro entre 2008 e outubro de 2015, embora tivesse sido liberada durante os eventos e jogos da Copa das Confederações (2013) e Copa do Mundo (2014). Em pesquisa feita com botafoguenses em 2013 (no grande portal esportivo brasileiro globoesporte.com)¹⁶, de 3.299 entrevistados, 2.219 (67,26%) moravam na Região Metropolitana do Rio de Janeiro. Destes, 19,6% afirmaram que o fato de não poderem consumir bebidas alcoólicas dentro do estádio tinha muita ou média influência na decisão de ir ao estádio. De fato, é comum observar muitos torcedores vendo jogos (que são no Rio de Janeiro) em bares, consumindo álcool. Observando os arredores dos estádios, nota-se também muitos torcedores consumindo álcool até os minutos que antecedem a partida, entrando no estádio a poucos minutos do

¹⁶ <http://globoesporte.globo.com/rj/torcedor-botafogo/platb/2013/11/01/pesquisa-o-botafoguense-no-estadio/>

apito inicial (ou até depois). A figura 11 mostra como foi dada a entrada dos torcedores, em espaços de 10 minutos, em um jogo do Botafogo em final de semana (Botafogo x Luverdense, sábado - 01/08/2015 às 16:30).

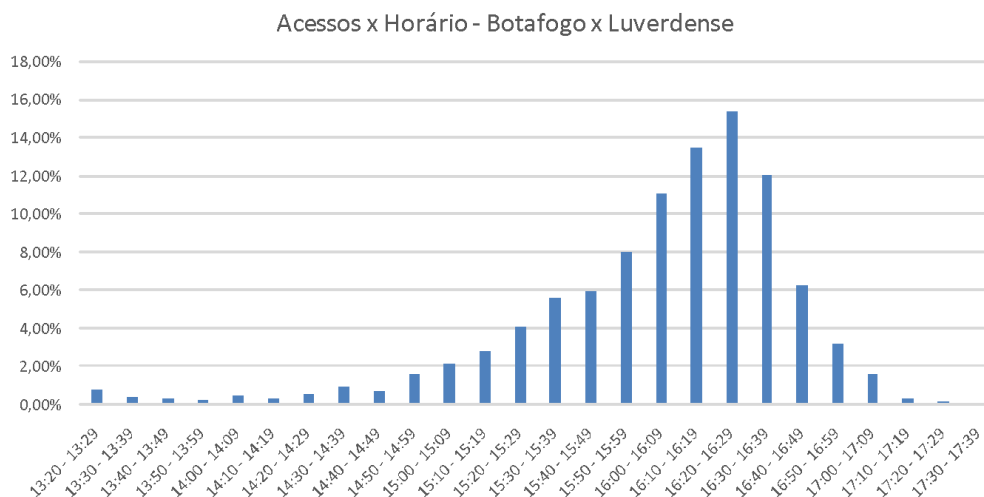


Figura 11 - Entrada de torcedores por faixa de horário – Botafogo x Luverdense (01/08/2015).

Fonte: O autor

Analisando a fundo os números, observa-se que 27,56% dos torcedores entram com 10 minutos de antecedência do início da partida ou nos 10 minutos iniciais do jogo. Além disso, incríveis 24,09% dos torcedores passaram nas roletas de acesso após o horário de início (16:30). Sendo um jogo em dia não útil, e sendo esse comportamento facilmente verificado em outras partidas, não é difícil imaginar que a falta de bebida alcoólica dentro dos estádios (assim como outras atrações) possam efeito na decisão de ir ao estádio (e inclusive no momento de entrar no estádio).

Estudo publicado por Pearson e Sale (2011) mostrou, como resultado de 15 anos de pesquisa e entrevistas com autoridades policiais britânicas e italianas (além de torcidas inglesas), que restringir o consumo de bebida alcoólica não é o fator determinante para reduzir índices de violência nos estádios (outro fator que afasta a torcida, mas que não se consegue estimar com exatidão por meio do modelo proposto). Entre outros problemas causados pela restrição à bebida alcoólica, o estudo aponta:

- Torcedores aumentam a quantidade de bebida ingerida fora do estádio, e passam a tomar bebidas mais fortes;
- Entrada “em cima da hora” do jogo, dificultando esquemas de segurança;

- Tumulto gerado pela entrada em conjunto e concentrada perto do início do jogo, aumentando utilização de catracas e dificultando o trânsito nos arredores;
- Perda de receitas para os estádios, não só para bares e restaurantes, como para lojas;
- Com mais pessoas nas ruas antes do jogo, além de se dificultar o acesso geral, aumenta-se a chance de confronto entre torcedores rivais nas ruas.

Com relação aos aspectos operacionais de jogo, é interessante analisar o impacto operacional e financeiro que o modelo, ao menos do Botafogo, traz para os jogos reais.

É de se esperar que uma partida com público grande tenha custos de operação maiores que uma partida com público menor. É de praxe, em estádios grandes e/ou modernos, haver uma tabela padrão de custos operacionais para cada expectativa de público. Itens de custo como serviço de manutenção de elevadores ou suporte de rede constituem praticamente custos fixos. No entanto, gastos com segurança, limpeza, postos médicos e socorristas são nitidamente variáveis, aumentando de acordo com o tamanho do jogo.

No Estádio Nilton Santos, a administração trabalha com uma tabela de custos que conta com 9 faixas de público, e obviamente o custo operacional total da partida vai aumentando gradativamente, embora exista ganho de escala: o custo por pessoa é cada vez menor, de acordo com o aumento do público.

A tabela 18, a seguir, mostra os itens de custo no Estádio Nilton Santos considerados pela administração em março de 2017. Como são dados sigilosos, os valores foram mascarados e multiplicados por um fator único, a fim de manter as proporções válidas.

Tabela 18 - Itens de custo, por faixa esperada máxima de público, no Estádio Nilton Santos em março/2017 (valores originais foram normalizados e multiplicados por um fator único).

Fonte: Botafogo FR / Estádio Nilton Santos

| | Faixa de público (Máximo) | | | | | | |
|--|---------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | 5.000 | 9.999 | 15.000 | 19.999 | 29.999 | 39.999 | 45.000 |
| Itens de custo | Valor (R\$) | Valor (R\$) | Valor (R\$) | Valor (R\$) | Valor (R\$) | Valor (R\$) | Valor (R\$) |
| Apoio de trânsito | 0,12 | 0,12 | 0,12 | 0,12 | 0,14 | 0,14 | 0,14 |
| Equipe Operacional | 0,25 | 0,25 | 0,38 | 0,38 | 0,63 | 0,63 | 0,63 |
| Gerador | 0,40 | 0,40 | 0,40 | 0,40 | 0,40 | 0,40 | 0,40 |
| Grade | 0,30 | 0,30 | 0,30 | 0,30 | 0,30 | 0,40 | 0,40 |
| Iluminação | 0,50 | 0,50 | 0,50 | 0,50 | 0,50 | 0,50 | 0,50 |
| Lanche Operacional | 0,03 | 0,04 | 0,05 | 0,06 | 0,09 | 0,11 | 0,13 |
| Limpeza | 0,29 | 0,29 | 0,29 | 0,36 | 0,36 | 0,54 | 0,54 |
| Manutenção Elétrica/Hidráulica/Ar Condicionado | 0,17 | 0,17 | 0,17 | 0,17 | 0,17 | 0,17 | 0,17 |
| Manutenção Elevador | 0,04 | 0,04 | 0,04 | 0,04 | 0,04 | 0,04 | 0,04 |
| Material Limpeza | 0,16 | 0,16 | 0,16 | 0,16 | 0,19 | 0,20 | 0,20 |
| Orientadores | 0,07 | 0,07 | 0,17 | 0,17 | 0,18 | 0,23 | 0,23 |
| Posto Médico | 0,15 | 0,23 | 0,30 | 0,37 | 0,50 | 0,62 | 0,73 |
| Segurança | 0,21 | 0,30 | 0,43 | 0,77 | 0,77 | 0,98 | 0,98 |
| Socorristas | 0,05 | 0,07 | 0,10 | 0,12 | 0,17 | 0,22 | 0,25 |
| Som e Telão | 0,63 | 0,63 | 0,63 | 0,63 | 0,63 | 0,63 | 0,63 |
| Suporte Rede/CCO | 0,04 | 0,04 | 0,04 | 0,06 | 0,06 | 0,06 | 0,06 |

A tabela 19, no entanto, mais uma vez normalizando valores, mostra a evolução percentual dos custos total e por pessoa de acordo com a faixa de público anterior.

Tabela 19 - Evolução dos custos totais e por pessoa, por faixa esperada máxima de público, no Estádio Nilton Santos em março/2017 (valores originais foram normalizados e multiplicados por um fator único).

Fonte: Botafogo FR / Estádio Nilton Santos

| Item | Faixa de público (Máximo) | | | | | | |
|---|---------------------------|--------|--------|--------|--------|--------|--------|
| | 5.000 | 9.999 | 15.000 | 19.999 | 29.999 | 39.999 | 45.000 |
| | R\$ | R\$ | R\$ | R\$ | R\$ | R\$ | R\$ |
| Custo Operacional Total | 3,39 | 3,59 | 4,05 | 4,60 | 5,09 | 5,85 | 6,00 |
| Evolução (%) – Custo Total | - | 5,82% | 12,91% | 13,38% | 10,85% | 14,89% | 2,52% |
| Evolução (%) – Custo Operacional/Pessoa | - | 47,09% | 24,73% | 14,96% | 26,10% | 13,83% | 8,88% |

A tabela acima deixa claro o ganho de escala: o custo por pessoa chega a cair 47% quando passamos de um jogo de até 5.000 presentes para um de até 9.000

presentes. No entanto, a mesma tabela revela algo preocupante: o custo total pode aumentar até quase 15% de uma faixa para outra (como ocorre na passagem de um jogo de até 29.999 presentes para a faixa seguinte, com limite máximo de 39.999). Ou seja: superestimar o público pode, eventualmente, aumentar o custo operacional em dois dígitos percentuais de maneira desnecessária. Por outro lado, como mencionado anteriormente, subestimar o público não acarretará despesas operacionais desnecessárias (embora possa complicar a operação com estresse em filas de compra de ingressos), mas ocasionará ao clube e à administração do estádio renúncia a receitas como as de estacionamento, bares, lojas oficiais e, claro, o próprio valor do ingresso.

Para avaliar o impacto de custos/economia do modelo estimador de público, podemos analisar qual o peso da diferença do público estimado para o realizado no que diz respeito aos custos operacionais do estádio.

De maneira mais prática, pode-se cruzar os dados da tabela 16 (Diferença nos resultados do modelo de estimativa do Botafogo para a realidade) com os da tabela 18 (custos operacionais por público presente no Estádio Nilton Santos).

Caso o número estimado de torcedores presentes e o público observado estejam na mesma faixa de público, o estimador não trará custo operacional extra ou perda de receitas, uma vez que será aberta partida com público limitado à faixa que contenha o público estimado. Caso contrário:

- Se o público estimado for maior que o público observado e estiver em faixa de custo operacional superior (deixando claro que será alcançada pelo menos uma faixa desnecessária de custos), trará custos operacionais desnecessários, calculados pela diferença de custo entre as faixas;
- Se o público estimado for menor que o público observado e estiver em faixa de custo operacional inferior (deixando claro que o limite de público esperado é inferior ao público observado), teremos perdas de receitas diretas, pois está claro que haveria demanda reprimida. A perda de receita será calculada com valores de ingressos apurados na partida além de um valor fixo para consumos com bares e restaurantes, descontados, no entanto, do aumento de custo operacional que deveria haver (para admitir a entrada desses torcedores).

Tabela 20 - Incremento de custos ou perda de receitas pelo uso do estimador de público do Botafogo (valores originais foram normalizados e multiplicados por um fator único).

Fonte: O autor

| Data | PUBBOT | Estimado | Custo Operacional PubBot | Custo Operacional Estimado | Incremento de custo operacional (erro para mais) | Perda de receitas diretas (erro para menos) |
|------------|-----------|-----------|--------------------------|----------------------------|--|---|
| 05/10/2013 | 12.990,54 | 9.748,83 | R\$ 4,05 | R\$ 3,59 | R\$ 0,00 | R\$ 1,91 |
| 13/10/2013 | 8.665,81 | 11.164,62 | R\$ 3,59 | R\$ 4,05 | R\$ 0,46 | R\$ 0,00 |
| 20/10/2013 | 8.893,13 | 8.252,44 | R\$ 3,59 | R\$ 3,59 | R\$ 0,00 | R\$ 0,00 |
| 26/10/2013 | 8.747,78 | 9.858,56 | R\$ 3,59 | R\$ 3,59 | R\$ 0,00 | R\$ 0,00 |
| 13/11/2013 | 10.223,54 | 7.190,81 | R\$ 4,05 | R\$ 3,59 | R\$ 0,00 | -R\$ 0,33 |
| 16/11/2013 | 13.517,61 | 6.922,68 | R\$ 4,05 | R\$ 3,59 | R\$ 0,00 | R\$ 1,02 |
| 08/12/2013 | 33.984,28 | 22.848,35 | R\$ 5,85 | R\$ 5,09 | R\$ 0,00 | R\$ 1,02 |
| 27/04/2014 | 8.728,35 | 8.238,09 | R\$ 3,59 | R\$ 3,59 | R\$ 0,00 | R\$ 0,00 |
| 10/05/2014 | 4.910,47 | 2.130,73 | R\$ 3,39 | R\$ 3,39 | R\$ 0,00 | R\$ 0,00 |
| 27/07/2014 | 13.836,55 | 14.202,77 | R\$ 4,05 | R\$ 4,05 | R\$ 0,00 | R\$ 0,00 |
| 02/08/2014 | 8.338,01 | 9.050,49 | R\$ 3,59 | R\$ 3,59 | R\$ 0,00 | R\$ 0,00 |
| 23/08/2014 | 18.136,88 | 9.459,74 | R\$ 4,60 | R\$ 3,59 | R\$ 0,00 | R\$ 2,79 |
| 31/08/2014 | 13.321,83 | 12.089,57 | R\$ 4,05 | R\$ 4,05 | R\$ 0,00 | R\$ 0,00 |
| 17/09/2014 | 4.660,73 | 1.606,19 | R\$ 3,39 | R\$ 3,39 | R\$ 0,00 | R\$ 0,00 |
| 25/09/2014 | 6.978,20 | 7.043,10 | R\$ 3,59 | R\$ 3,59 | R\$ 0,00 | R\$ 0,00 |
| 28/09/2014 | 10.689,36 | 14.921,52 | R\$ 4,05 | R\$ 4,05 | R\$ 0,00 | R\$ 0,00 |
| 08/10/2014 | 7.988,49 | 7.309,51 | R\$ 3,59 | R\$ 3,59 | R\$ 0,00 | R\$ 0,00 |
| 15/11/2014 | 5.823,69 | 1.327,46 | R\$ 3,59 | R\$ 3,39 | R\$ 0,00 | R\$ 0,55 |

Percebe-se, na tabela 20 acima, que apenas uma vez (jogo do dia 13/10/2013) o estimador erra o público de forma que isso signifique um aumento de custo operacional. Isso porque foi estimado um público (11.164,62) que está numa faixa de público superior à que melhor se adequa ao público observado (8.665,81). Isto é: abre-se uma faixa superior de público de maneira desnecessária, pois o limite superior de 9.999 pessoas (vide tabela 18 ou 19) comportaria o público observado.

Por outro lado, em 6 vezes o modelo errou o público “para baixo”, de modo que o público observado estivesse em faixa superior de público em relação ao público estimado. Isso significa que, caso fosse implementado o limite superior da faixa em que está inserido o público estimado, este limite seria inferior ao público observado da partida. Assim, essa diferença (entre público observado e o limite da faixa do público estimado) seriam pessoas que poderiam ir ao estádio, mas não iriam por repressão de demanda (baixa oferta de ingressos), deixando de comprar entradas e consumir dentro do estádio. Em uma dessas situações, no entanto (partida do dia 13/11/2013), o público-extra que

poderia ir à partida ($10.223,54 - 9.999 = 224,54$ pessoas), mesmo consumindo produtos alimentícios e pagando ingressos, não compensaria o esforço e custo operacional de abrir o jogo para uma faixa superior. Por isso, a perda “negativa” diz que o erro do modelo é benéfico neste caso.

Marcados em laranja, na tabela 20, são observados jogos ditos “clássicos” (contra Flamengo, Fluminense ou Vasco), em que a estrutura de custos do estádio (principalmente nas linhas de segurança) muda completamente. Além disso, como muitas vezes o público é dividido, simplesmente estimar a presença de público do Botafogo não é tão relevante para se pensar no custo global operacional do estádio, já que a outra torcida tem tamanho relevante, por ser do Rio de Janeiro. De fato, torcidas visitantes, quando são de fora do Rio, têm comportamento mais estável, sendo mais fácil pensar no estimador do Botafogo como um bom parâmetro para prever o público total do estádio (e então o custo geral). Serão retirados, assim, estes jogos das análises futuras.

Sabe-se que, na prática, é usual “superestimar” públicos na hora de planejar uma partida. Como mencionado anteriormente, na maioria das vezes, errar o público “para menos” é mais custoso do que errar “para mais”, na medida em que os torcedores (pagando ingressos e consumindo dentro do estádio) compensam os custos operacionais maiores.

Assim, é possível comparar as perdas de receitas e custos atrelados ao uso do estimador com os custos relacionados ao fato de se operar sempre com uma faixa de segurança. Exemplificando, caso o público real seja de 17.500 (faixa até 19.999), considerar-se-ia que o jogo foi aberto para até 29.999 (uma faixa acima). Os resultados estão na próxima abaixo (com valores mais uma vez normalizados):

Tabela 21 - Comparativo do prejuízo acumulado abrindo partida na faixa de público com limite superior imediatamente acima à do público estimado com prejuízo acumulado abrindo partida na faixa de público superior à do público observado.

Fonte: O autor

| Data | PUBBOT | Estimado | Diferença | Perda abrindo faixa do Estimado | Perda abrindo 1 faixa acima do Observado |
|------------------|-----------|-----------|------------|---------------------------------|--|
| 05/10/2013 | 12.990,54 | 9.748,83 | -3.241,71 | R\$ 1,91 | R\$ 0,54 |
| 13/10/2013 | 8.665,81 | 11.164,62 | 2.498,82 | - | - |
| 20/10/2013 | 8.893,13 | 8.252,44 | -640,68 | - | - |
| 26/10/2013 | 8.747,78 | 9.858,56 | 1.110,78 | R\$ 0,00 | R\$ 0,46 |
| 13/11/2013 | 10.223,54 | 7.190,81 | -3.032,73 | -R\$ 0,33 | R\$ 0,54 |
| 16/11/2013 | 13.517,61 | 6.922,68 | -6.594,93 | R\$ 1,02 | R\$ 0,54 |
| 08/12/2013 | 33.984,28 | 22.848,35 | -11.135,93 | R\$ 1,02 | R\$ 0,15 |
| 27/04/2014 | 8.728,35 | 8.238,09 | -490,26 | R\$ 0,00 | R\$ 0,46 |
| 10/05/2014 | 4.910,47 | 2.130,73 | -2.779,75 | R\$ 0,00 | R\$ 0,20 |
| 27/07/2014 | 13.836,55 | 14.202,77 | 366,23 | - | - |
| 02/08/2014 | 8.338,01 | 9.050,49 | 712,48 | R\$ 0,00 | R\$ 0,46 |
| 23/08/2014 | 18.136,88 | 9.459,74 | -8.677,14 | R\$ 2,79 | R\$ 0,50 |
| 31/08/2014 | 13.321,83 | 12.089,57 | -1.232,26 | R\$ 0,00 | R\$ 0,54 |
| 17/09/2014 | 4.660,73 | 1.606,19 | -3.054,54 | R\$ 0,00 | R\$ 0,20 |
| 25/09/2014 | 6.978,20 | 7.043,10 | 64,90 | R\$ 0,00 | R\$ 0,46 |
| 28/09/2014 | 10.689,36 | 14.921,52 | 4.232,16 | R\$ 0,00 | R\$ 0,54 |
| 08/10/2014 | 7.988,49 | 7.309,51 | -678,98 | R\$ 0,00 | R\$ 0,46 |
| 15/11/2014 | 5.823,69 | 1.327,46 | -4.496,23 | - | - |
| Total | | | | R\$ 6,41 | R\$ 6,07 |
| Diferença | | | | 5,59% | |

Apesar de todos os valores estarem normalizados, percebe-se que o uso do estimador é cerca de 5,6% mais “caro” do que estar sempre com uma faixa de público aberta a mais. Consideramos bom este resultado, uma vez que superestimar o público em “apenas” uma faixa é o melhor dos casos, sendo possível errar 2 ou 3 faixas acima, o que já pioraria o resultado da última coluna da tabela 21, comprovando que o resultado do estimador é bom. Observe, também, que foram descartados os jogos “clássicos” (grifados em laranja), como explicado anteriormente.

Pode-se, no entanto, melhorar o resultado do modelo ao mitigar os erros de público “para baixo”. Ao se selecionar sempre uma faixa acima do público estimado, diminui-se o risco de perder receitas com ingressos e consumo dentro do estádio, como se pode ver na tabela a seguir.

Tabela 22 - Comparativo do prejuízo acumulado (aumento de custos ou perda de receita) abrindo partida na 1ª faixa de público acima do público estimado com prejuízo acumulado abrindo partida na faixa de público superior à do público observado.

Fonte: O autor

| Data | PUBBOT | Estimado | Diferença | Perda abrindo 1 faixa acima do Estimado | Perda abrindo 1 faixa acima do Observado |
|------------------|-----------|-----------|------------|---|--|
| 05/10/2013 | 12.990,54 | 9.748,83 | -3.241,71 | R\$ 0,00 | R\$ 0,54 |
| 13/10/2013 | 8.665,81 | 11.164,62 | 2.498,82 | - | - |
| 20/10/2013 | 8.893,13 | 8.252,44 | -640,68 | - | - |
| 26/10/2013 | 8.747,78 | 9.858,56 | 1.110,78 | R\$ 0,46 | R\$ 0,46 |
| 13/11/2013 | 10.223,54 | 7.190,81 | -3.032,73 | R\$ 0,00 | R\$ 0,54 |
| 16/11/2013 | 13.517,61 | 6.922,68 | -6.594,93 | R\$ 0,00 | R\$ 0,54 |
| 08/12/2013 | 33.984,28 | 22.848,35 | -11.135,93 | R\$ 0,00 | R\$ 0,15 |
| 27/04/2014 | 8.728,35 | 8.238,09 | -490,26 | R\$ 0,46 | R\$ 0,46 |
| 10/05/2014 | 4.910,47 | 2.130,73 | -2.779,75 | R\$ 0,20 | R\$ 0,20 |
| 27/07/2014 | 13.836,55 | 14.202,77 | 366,23 | - | - |
| 02/08/2014 | 8.338,01 | 9.050,49 | 712,48 | R\$ 0,46 | R\$ 0,46 |
| 23/08/2014 | 18.136,88 | 9.459,74 | -8.677,14 | R\$ 0,92 | R\$ 0,50 |
| 31/08/2014 | 13.321,83 | 12.089,57 | -1.232,26 | R\$ 0,54 | R\$ 0,54 |
| 17/09/2014 | 4.660,73 | 1.606,19 | -3.054,54 | R\$ 0,20 | R\$ 0,20 |
| 25/09/2014 | 6.978,20 | 7.043,10 | 64,90 | R\$ 0,46 | R\$ 0,46 |
| 28/09/2014 | 10.689,36 | 14.921,52 | 4.232,16 | R\$ 0,54 | R\$ 0,54 |
| 08/10/2014 | 7.988,49 | 7.309,51 | -678,98 | R\$ 0,46 | R\$ 0,46 |
| 15/11/2014 | 5.823,69 | 1.327,46 | -4.496,23 | - | - |
| Total | | | | R\$ 4,72 | R\$ 6,07 |
| Diferença | | | | -22,30% | |

Abrindo a partida uma faixa acima do público estimado faz com que se reduza em 22,30% o prejuízo comparado ao estar sempre uma faixa acima do público observado.

Isso mostra o bom resultado do estimador construído que, combinado com o conhecimento prático (operar com margem para cima, e não para baixo), faz com que sejam alcançados bons resultados práticos.

5. Conclusão

Finalizando este trabalho e concluindo com base nos resultados, pode-se explicar boa parte da variação de um fenômeno social e, como sabido, movido muitas vezes mais pela paixão do que pela razão. Assim, foi possível atingir o principal objetivo deste trabalho, a saber construir um modelo de previsão de público para jogos do Botafogo.

Na discussão dos resultados que o modelo, também foi observado que o modelo estatístico de previsão de público pode ser usado de modo a minimizar tanto os riscos de desperdício de receitas (prever público muito menor do que a demanda) quanto as chances de custos operacionais desnecessários (prever público muito maior do que a demanda). Desta forma, vê-se que um dos objetivos paralelos deste trabalho foi alcançado.

Além de conseguir extrair uma equação ou formulação matemática, entender o comportamento do torcedor com apoio numérico era mais uma meta deste trabalho. E, assim como as outras metas, ela também foi atingida. Observou-se, por exemplo, o grande peso que horário e o dia da semana representam para o torcedor botafoguense ir ao estádio, parâmetros esses muitas vezes determinados pela tabela ou Televisão. Será que o produto da Televisão não teria maior valor agregado com mais público no estádio e, para tornar isso factível, os horários tenham que ser diferentes?

Como sugestão de trabalhos futuros, pode ser interessante analisar os impactos da oferta (ou ausência) de transporte público, da segurança pública e também da distância do estádio para as moradias dos torcedores. A importância que o torcedor do Botafogo, como visto, dá para o horário de início da partida ou o dia em que a mesma é jogada pode ser o indício de uma sensação de insegurança (agravada pelo horário) ou mesmo pela dificuldade no retorno ao lar com transporte público.

Outra sugestão de análise posterior, dado que o resultado das partidas anteriores tem enorme influência na decisão de ir ao estádio, é trabalhar com técnicas de análise de sentimento dos torcedores, principalmente utilizando-se dados da internet e das redes sociais. Neste trabalho, os resultados anteriores foram medidos com as pontuações usuais de um campeonato de pontos corridos (3 pontos para vitória, 1 ponto para empate e 0 pontos para derrota), além das variáveis binárias para momento positivos e negativos. É possível que estimar o sentimento da torcida por outras maneiras possa apresentar resultados melhores.

É importante também deixar claro que o estudo e o método aqui apresentados não abandonam nem devem substituir o conhecimento prático e acumulado por gestores

e profissionais da área. Com eles, por exemplo, as ferramentas podem ser aprimoradas e adaptadas a cada clube, com novas variáveis e novas modelagens.

Por fim, espera-se que este trabalho estimule maior produção científica e intelectual para o futebol (esporte) brasileiro. O Brasil é pentacampeão mundial (até o momento, faltando pouco para a Copa do Mundo de 2018) e não há dúvidas de que nos sobra talento em campo: aperfeiçoar a prática fora de campo, entre gestores e dirigentes, deve ser um constante caminho para melhorar a qualidade e a organização do futebol jogado aqui.

Todo torcedor sabe que a bola não entra por acaso e este trabalho mostra também que o torcedor não vai ao estádio sem motivo.

Referências

ALAVY, K., GASKELL, A., LEACH, S. et al, 2010, "On the Edge of Your Seat: Demand for Football on Television and the Uncertainty of Outcome Hypothesis", *International Journal of Sport Finance*, v. 5, n. 2, pp. 75-95.

BORTOLUZZO, A. B., IAROPOLI, P. T., MACHADO, S. J., 2011, *Demand for Brazilian Soccer: A Censored Model Approach*. Working Paper, n. 237, Insper, São Paulo, SP, Brasil. Disponível em: <http://en.insper.edu.br/sites/default/files/2011_wpe237.pdf>. Acesso em 12 de março de 2017.

BORTOLUZZO, A. B., BORTOLUZZO, M. M., MACHADO, S. J. et al, 2017, "Ticket consumption forecast for Brazilian championship games", *Revista de Administração*, v. 52, n. 1, pp. 70-80.

DEMMERT, H. G., 1973, *The Economics of Professional Team Sports*. 1 ed. Lexington, Lexington Books.

DREVER, P., MCDONALD, J., 1981, "Attendances at South Australian Football Games", *International Review for the Sociology of Sport* v. 16, n. 2, pp. 103–113.

FALLS, G. A., NATKE, P. A., 2016, "College football attendance: A panel study of the Football Championship Subdivision", *Managerial and Decision Economics* v. 37, n. 8, pp. 530–540.

FALTER, J. M., PÉRIGNON, C., 2000, "Demand for football and intramatch winning probability: An essay on the glorious uncertainty of sports", *Applied Economics*, v. 32, n. 13, pp. 1757-1765.

FORREST, D., SIMMONS, R., 2002, "Outcome uncertainty and attendance demand in sport: the case of English soccer", *Journal of the Royal Statistical Society: Series D (The Statistician)*, v. 51, n.2, pp. 229–241.

GARCÍA, J., RODRÍGUEZ, P., 2002, "The Determinants of Football Match Attendance Revisited : Empirical Evidence From the Spanish Football League", *Journal of Sports Economics*, v. 3, n.1, pp. 18-38.

HART, R. A., HUTTON, J., SHAROT, T., 1975, "A Statistical Analysis of Association Football Attendances", *Journal of the Royal Statistical Society - Series C Applied Statistics*, v. 24, n. 1, pp. 17-27.

KINGSTRAD, M., SOLBERG, H. A., JAKOBSEN, T. G., 2018, "Does live broadcasting reduce stadium attendance? The case of Norwegian football", *Sport, Business and Management: An International Journal*, v. 8, n. 1, pp. 67-81.

KUYPERS, T., 1996, *The beautiful game? An econometric study of audiences, gambling and efficiency in English Football*. Economics PhD Tesis, University College London, London, United Kingdom.

LEMKE, R. J., LEONARD, M., & TLHOKWANE, K., 2010, "Estimating Attendance at Major League Baseball Games for the 2007 Season", *Journal of Sports Economics*, v. 11, n. 3, pp. 316-348.

LIMA, M., 2015, *Economia Futebol e Paixão: O Comportamento do Torcedor do Esporte Clube Bahia*. Trabalho de conclusão de curso de graduação, Universidade Federal da Bahia, Faculdade de Economia, Salvador, BA, Brasil.

MADALOZZO, R., VILLAR, R. B., 2008, *A Model of Attendance Demand at the Brazilian Football League*. Working Paper, n. 113, Insper, São Paulo, SP, Brasil. Disponível em: https://www.insper.edu.br/wp-content/uploads/2013/12/2008_wpe113.pdf. Acesso em 10 de março de 2017.

MADALOZZO, R., VILLAR, R. B., 2009, "Brazilian football: what brings fans to the game?", *Journal of Sports Economics*, v. 10, n. 6, pp. 639-650.

MEIER, H. E., KONJER, M., LEINWATHER, M, 2015, "The demand for women's league soccer in Germany", *European Sport Management Quarterly*, v. 16, n. 1, pp. 1-19.

MIRABILE, M. D., 2015, "The Determinants of Attendance at Neutral Site College Football Games", *Managerial and Decision Economics* v. 36, n. 3, pp. 191–204.

NOLL, R. G., 1974, "Attendance and Price Setting". In: NOLL, R.G. (ed.), *Government and The Sports Business*, 1 ed., Washington, DC, Brookings Institution, pp. 115-158.

PEARSON, G., SALE, A., 2011, "'On the Lash': revisiting the effectiveness of alcohol controls at football matches", *Policing and Society* v. 21, n. 2, pp. 150-166.

SOUZA, F. A. P., 2004, *Um estudo sobre a demanda por jogos de futebol nos estádios brasileiro*. Dissertação, Universidade de São Paulo, São Paulo, SP, Brasil.

STRNAD, D., NERAT, A., KOHEK, S., 2017, "Neural network models for group behavior prediction: a case of soccer match attendance", *Neural Computing and Applications* v. 28, n. 2, pp. 287–300.

SZYMANSKI, S., 2001, "Income inequality, competitive balance and the attractiveness of team sports: some evidence and a natural experiment from English soccer", *The Economic Journal*, v. 111, n. 469 (Fev.), pp. F69-F84.

SZYMANSKI S., 2013, "Wages, Transfers and The Variation of Team Performance in The English Premier League". In: RODRÍGUEZ, P., KÉSENNE, S., GARCÍA, J. (eds), *The Econometrics of Sport*, 1 ed, chapter 3, Northampton, MA, Edward Elgar.

SZYMANSKI S., 2014, *Stefan Szymanski on the business of football*. Working paper, Disponível em: <http://www.open.edu/openlearn/money-management/management/business-studies/stefan-szymanski-on-the-business-football>>. Acesso em: 10 de dezembro de 2017.

SZYMANSKI S., HALL, S., 2003, *Making money out of football*. Working paper, Imperial College London, London, United Kingdom.

VILLA, G., MOLINA, I., FRIED, R., 2011, "Modeling attendance at Spanish professional football league", *Journal of Applied Statistics* v. 38, n. 6, pp. 1189–1206.

WELKI, A. M., ZLATOPER, T. J., 1999, "U.S. Professional Football Game-Day", Attendance", *Atlantic Economic Journal*, v. 27, n. 3 (Set.), pp. 285-298.