

Previsão de público com modelo de Regressão Linear para jogos do Botafogo de Futebol e Regatas no Engenhão

Caio Fernando dos Santos Araujo

Programa Engenharia de Produção, Instituto Alberto Luiz Coimbra de Pós Graduação e Pesquisa em Engenharia, Universidade Federal do Rio de Janeiro.
Centro de Tecnologia, Bloco F, Sala 101, Ilha do Fundão. C.P. 68507, Rio de Janeiro RJ
21941-972
caio.f.s.araujo@gmail.com

Edilson de Arruda Fernandes

Programa Engenharia de Produção, Instituto Alberto Luiz Coimbra de Pós Graduação e Pesquisa em Engenharia, Universidade Federal do Rio de Janeiro.
Centro de Tecnologia, Bloco F, Sala 101, Ilha do Fundão. C.P. 68507, Rio de Janeiro RJ
21941-972
efarruda@po.coppe.ufrj.br

RESUMO

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 Engenhão durante os Campeonatos Brasileiros de 2010, 2011 e 2012. 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). Apesar de ainda estar em fase de desenvolvimento, o modelo permite conclusões sobre o que faz o torcedor assistir a um jogo de futebol no estádio.

PALAVRAS CHAVE: Regressão Linear, Futebol, Estatística.

Área principal: Estatística, PO na Administração & Gestão da Produção.

ABSTRACT

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. Were used financial newsletters from games of Botafogo at the Engenhão stadium during the 2010, 2011 and 2012 Brazilian Championship seasons. A reliable estimate of the public would facilitate, among other things, the dimensioning of the operating facilities to serve games (such as the number of gates to be opened). Although still being perfected, the proposed model helps to explain what makes the fans go watch a game at a stadium.

KEYWORDS. Linear Regression. Football. Statistics.

Main area: Statistics, OR in Administration & Production Management.

1. Introdução

Construído para os Jogos Panamericanos de 2007, o Estádio Olímpico João Havelange é rotineiramente chamado de Engenhão por se situar no bairro do Engenho de Dentro, no Rio de Janeiro. O Botafogo de Futebol e Regatas, clube de futebol da cidade, venceu licitação e passou a administrar o estádio, que abrigou a grande maioria dos jogos em que a equipe tem o mando de campo desde o final de 2007. Com 5 anos indo ao Engenhão, já podemos dizer que a torcida do Botafogo está acostumada a ir ao estádio, o que nos permite tentar construir modelos que expliquem o comportamento deste torcedor.

O estádio é dividido em 4 macro-setores: Norte, Sul, Leste e o Oeste, sendo os dois últimos setores tendo dois pavimentos cada: inferiores e superiores. Norte e Sul abrigam cerca de 4500 torcedores cada, os setores inferiores (Leste e Oeste) têm capacidade de 7334 torcedores, enquanto os superiores (Leste e Oeste) suportam 10775 cada. Normalmente a torcida do Botafogo só não fica na ala Sul, setor reservado aos visitantes. Porém, em clássicos (cada vez mais comuns no estádio após as obras no Maracanã, que está fechado desde 2010) contra Vasco, Flamengo e Fluminense, o Botafogo fica apenas nos setores Norte, Oeste Inferior e Oeste Superior.

Apesar de o estádio ter sido recentemente fechado (em março de 2013) por suposto risco de queda da cobertura, entende-se que o trabalho aqui desenvolvido poder ser replicado para outros estádios (e até para outros clubes, desde que sejam feitas adaptações). Além disso, é esperado que o modelo aqui proposto sirva novamente ao Engenhão, com a reabertura do mesmo.

Como esta é apenas uma primeira versão, a parte final mostra os futuros passos que o trabalho deverá seguir para melhorar os resultados, que inclusive serão discutidos na última seção. Além de se melhorar o modelo de previsão, pretende-se também construir um simulador das filas de entrada dos torcedores, baseado em filas de disciplina M/M/s (com chegadas das pessoas e entradas nas roletas explicadas por distintas distribuições exponenciais), segundo a notação de Kendall. Para mais detalhes sobre modelos e notações de filas, veja Hillier e Lieberman (2010).

2. Breve 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 EUA e no Reino Unido. Há várias tentativas de se buscar o que faz um torcedor ir (ou não) a um estádio de futebol. No entanto, é nova a ideia de se aplicar um modelo apenas para um time, com uma torcida de comportamento e tamanho bem definidos e uma cidade fixa: a enorme maioria dos trabalhos constrói modelos com uma base de dados de campeonatos inteiros, misturando vários times. Este trabalho se propõe a fazer um modelo focado em um único time, mas que possa aproveitar resultados e conclusões de trabalhos anteriores.

Kuypers (1996) tenta buscar 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 de um canal de assinatura brasileiro nos públicos brasileiros.

Assim como Bortoluzzo et. al (2011), Kuypers (1996) faz uso de um modelo Tobit (além do tradicional método dos Mínimos Quadrados Ordinários) para efetuar a regressão. O modelo Tobit é escolhido já que 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).

A incerteza do resultado é apontada como um fator que leva o 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 prévia de vitória dos times. Madalozzo e Villar (2009) e Souza (2004) chegam à mesma conclusão para o caso do futebol brasileiro, variando na medida da incerteza do resultado, que pode ser estimada também pela diferença prévia de pontos entre os dois times envolvidos na partida.

Outra conclusão que é feita por quase todos os trabalhos é que a qualidade da partida tem efeito positivo no público. No entanto, a forma de medir a qualidade varia bruscamente nos trabalhos. García e Rodríguez (2002) estimam a qualidade do jogo pela quantidade de jogadores com passagens por seleções nacionais envolvidos no jogo, medida 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). Já Szymanski (2001) concluiu que a soma das posições dos times envolvidos no jogo é estatisticamente significativa e tem 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 é importante para a presença do público. Bortoluzzo et. al (2011) mostram que as rivalidades locais têm um impacto positivo na presença. Os mesmos autores também comprovam um resultado apontado por Madalozzo e Villar (2009), que mostra que a renda per capita das cidades que sediam os jogos tem impacto negativo na presença do público, um resultado aparentemente contra-intuitivo. 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 é bom, espera-se um impacto positivo no público. E, mais uma vez, há uma série de formas de estimar a fase e a importância do jogo. Neste 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 podemos 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 et. al (2011) explica e discute os tipos de variáveis.

3. Método

Para construir o modelo de Regressão Linear, são necessárias informações que constituam uma base de dados. Tais 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. Neste trabalho foram utilizados dados dos borderôs dos jogos do Botafogo nos Campeonatos Brasileiros de 2010, 2011 e 2012. Sendo assim, o modelo servirá apenas para estimar públicos do Botafogo para jogos do Campeonato Brasileiro, pelo menos nesta 1ª versão. A base de dados tem 62 observações, isto é, 62 jogos: 49 ocorridos em finais de semana (ou feriados) e 13 ocorridos em dias úteis. Os jogos em dias úteis são diferenciados dos outros por meio de uma variável *dummy*.

Construída a base de dados, variáveis são escolhidas para tentar explicar a variável dependente: no nosso caso, o público total de Botafoguenses 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.

4. As Variáveis

São vários os fatores que fazem um torcedor ir (ou não ir) a um estádio de futebol. De

forma rápida, podemos selecionar alguns:

- O adversário a ser enfrentado;
- A data/hora do jogo;
- O preço do ingresso;
- Se estamos 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). 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 supondo que a proporção de torcedores do time da casa e do time visitante é a mesma para o público pagante e paga as gratuidades. Isto é: se em determinado jogo foram registrados, por exemplo, 1000 pagantes do Palmeiras e 9000 pagantes do Botafogo além de 5000 gratuidades, o público presente estimado do Botafogo será de 13500 pessoas, já que o Botafogo teve 90% do público pagante da partida ($9000 + (9000/10000) \times 5000$ gratuidades = $9000 + 0,9 \times 5000 = 9000 + 4500 = 13500$). Chamaremos essa variável (que será a nossa única variável independente) de *PubBot*.

Apesar de ser apenas uma primeira versão do estudo, tentaremos fazer uma análise completa, de modo a captar vários dos possíveis efeitos que influenciam na presença do público. Agora falaremos das 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 passados.

Desta forma, para cada jogo da base de dados, foram computadas as pontuações dos jogos anteriores do Botafogo, mesmo que os jogos anteriores não fossem todos do Campeonato Brasileiro (Campeonato Carioca, Copa do Brasil 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 na tabela 1.

	1º jogo anterior	2º jogo anterior	3º jogo anterior	4º jogo anterior	5º jogo anterior
Correlação com o público	0,471	0,259	0,103	0,132	-0,046

Tabela 1. Correlação do público com a pontuação obtida nos 5 jogos anteriores. Fonte: Os autores.

Como esperado, a memória do torcedor “diminui” conforme nos afastamos temporalmente, e a correlação cai até o 3º jogo anterior. A correlação aumenta para o 4º jogo e fica 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 3 jogos anteriores. Assim, serão utilizadas as variáveis $R1$, $R2$ e $R3$, separadamente, tendo, cada uma, valor 3 quando o resultado for vitória, 1 para empate e 0 para derrota.

• Momentos especiais

As variáveis $R1$, $R2$ e $R3$ não conseguem medir, no entanto, situações ditas especiais como um título conquistado ou a estreia de um grande atleta. Assim, foi criada a variável Mom que será binária: terá valor 1 quando a situação for especial (apresentação ou estreia de um grande jogador, ou até um jogo subsequente a um título) e será 0 quando o jogo for normal. Em 62 observações, apenas 6 jogos tiveram esta variável no valor 1.

• Dia da semana

O dia da partida, acredita-se, 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 (49 jogos de 62 observações) e 0 para jogos em dias úteis (13 observações).

• 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, foi proposta a variável $Hora$, que se comporta de acordo com a tabela 2 abaixo.

Horário	Final de semana/Feriado	Dia útil
16:00	1	-
17:00	1	-
18:00	2	-
18:30	2	-
19:00	2	-
19:30	3	1
20:30	-	2
21:00	3	2
21:50	-	3
22:00	-	3

Tabela 2. Comportamento da variável que mede o horário da partida. Fonte: Os autores.

A 1ª coluna mostra todos os horários de jogos registrados na base de dados. Para jogos em finais de semana ou feriados, foi convencionado que jogos antes das 18h terão a variável valendo 1, valor 2 para jogos entre 18h e 19h e 3 para horário acima das 19h. Em semelhante raciocínio para dias úteis, a variável terá valor 1 quando o jogo for em um dia útil começando antes das 20h, valor 2 para jogos entre 20h e 21h e valor 3 para jogos que começam depois das 21h.

• 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, devemos decidir como passarão os dias desse mês, isto é: o calendário seguirá uma escala linear ou uma exponencial, por exemplo? Se decidirmos 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 gastar com o ingresso (ou o interesse do torcedor em ir ao jogo) cairá linearmente.

Foram testadas 15 variáveis diferentes para esse 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 calendário (1º dia do mês, 5º dia do mês e 5º dia útil do mês).

Sabemos que muitas pessoas não recebem já no 1º dia do mês, e já esperávamos 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.):

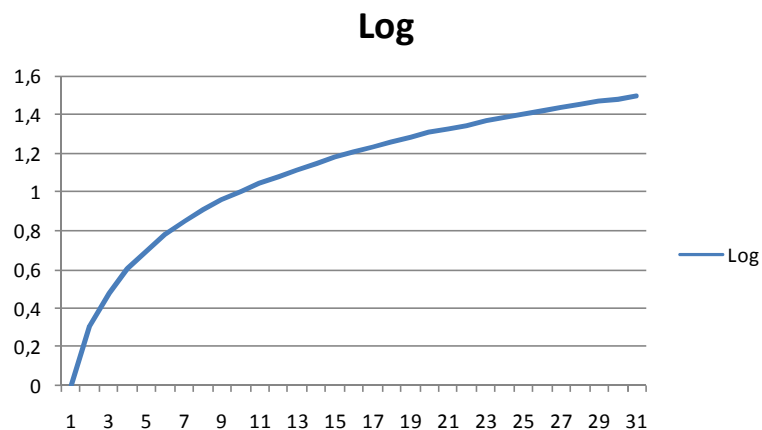


Fig. 1. Comportamento da curva logarítmica de 1 a 31. Fonte: Os autores.

Percebe-se, vendo a figura 1, 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 formos nos afastando do dia 5 nos dias próximos (mas posteriores) ao dia 5, e cairá de maneira mais estável até o dia 4 do mês seguinte. Isto faz sentido, se imaginarmos que há muitas contas a serem pagas nos dias próximos ao recebimento, sobrando cada vez menos do salário para o lazer (futebol).

Como exemplo do comportamento da variável criada (*LogD5c*), a tabela a seguir mostra possíveis valores para determinadas datas.

Data da partida	Variável “Distância dia 5”	<i>LogD5c</i>
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

Tabela 3. Comportamento da variável que mede a distância da data da partida para o dia de pagamento.

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

O preço dos ingressos postos à venda foi ponderado pela quantidade de lugares disponíveis, já que cada setor tem um preço diferente e uma capacidade diferente. Logo, por exemplo, se um setor puder (tiver capacidade de) abrigar 90% do público do Botafogo e o ingresso (preço de inteira) for R\$ 50 e outro setor tiver os 10% restantes de capacidade com preços a R\$ 20, o preço ponderado será de R\$ 47 ($90\% * 50 + 10\% * 20$).

Depois disso, uma variável chamada *PS* foi criada dividindo este preço ponderado (de cada jogo) pelo rendimento médio nominal da região metropolitana do Rio de Janeiro (valores mensais, conseguidos junto à Pesquisa Mensal de Emprego, do IBGE). Os valores foram selecionados para todos os meses de 2010, 2011 e 2012.

O preço foi dividido pelo salário médio a fim de ponderar o peso do ingresso no rendimento do torcedor.

• Clássicos

Por conta da rivalidade local, é esperado que jogos ditos “clássicos” (contra Flamengo, Fluminense e Vasco) atraiam mais torcedores do Botafogo. Assim, foi criada uma variável binária (*dummy*) de nome *Class*, que terá valor 1 quando o jogo for um clássico, e valor 0 para outras partidas.

• Importância do jogo

Um jogo é dito importante quando o time tem objetivos claros e definidos em um campeonato. Isto é: quando ele quer (e tem condições) 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), maior será, espera-se, o público no estádio.

Acreditamos também que os jogos vão se tornando mais importantes conforme o campeonato fica mais “agudo”, isto é, perto do seu final (aqui, devemos lembrar também que o Campeonato Brasileiro tem 38 rodadas). Assim, adaptaremos o uso 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:

- $(PB \times GL)$, onde PB = pontos atrás do líder e GL = número de jogos que faltam.

Quando o time for o líder da competição, o indicador é forçado ao resultado 1 (aqui consideramos que o time será líder sempre que $PB = 0$ mesmo que a liderança não esteja sendo alcançada por critérios de desempate). Quando não há mais chance matemática de título, a fórmula é forçada a 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.

Criamos, assim, a variável *ImpJogo2*, uma adaptação da fórmula descrita acima. Quando o time não tiver mais chances de título, o “alvo” será o 4º lugar, que dá acesso à Taça Libertadores (competição continental) do ano seguinte. Isto é, quando $PB > 3GL$ (pois cada vitória vale 3 pontos), PB irá se referir à 4ª colocação, pois um time sem chances de título nas últimas rodadas, mas com reais chances de classificação para a Taça Libertadores irá motivar a presença de público. Porém, quando também não houver chances matemáticas de atingir a 4ª colocação, a variável será zero (0).

Vale destacar que quanto menor for a variável, mais importante será a partida, e espera-se um impacto positivo no público.

• Qualidade do adversário

Se na variável anterior o foco era o time mandante, agora devemos olhar o time visitante, o adversário.

É esperado 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. Acreditamos que essa é uma realidade ainda distante do futebol brasileiro, devido ao baixo número de jogadores de times brasileiros em seleções nacionais de futebol, e por isso vamos estimar a qualidade do time adversário de outra forma.

Como dito anteriormente, Kuypers (1996) propõe 3 fórmulas para medir a importância de um jogo para determinado time. Aqui, vamos adaptar o uso à terceira fórmula:

- $(PB/3GL) \times (GL^2 + 2GL)$, onde PB = pontos atrás do líder e GL = número de jogos que faltam.

Diferente da segunda fórmula (usada para estimar a importância do jogo para o time mandante), esta dá maior ênfase à diferença de pontos para o líder da competição. Chamaremos essa variável (que poderá mudar bruscamente a cada rodada, dado que serão enfrentados adversários completamente diferentes) de *Qadv*.

Mais uma vez, quando o time (adversário) for o líder da competição, a fórmula dará o resultado 1 (aqui consideramos que o time será líder sempre que $PB = 0$ mesmo que a liderança não esteja sendo alcançada por critérios de desempate).

E assim como foi com a variável anterior (*ImpJogo2*), quanto menor for a variável *Qadv*, mais importante será a partida, e espera-se um impacto positivo no público.

5. O Modelo e os Resultados

As 62 observações com as respectivas 11 variáveis independentes (descritas no item anterior) e a variável independente (*PubBot*) foram inseridas no programa EViews, bastante utilizado para fazer regressões.

Pelo Método dos Mínimos Quadrados Ordinários, chegamos ao seguinte resultado, retirado do programa:

Dependent Variable: PUBBOT
Method: Least Squares
Date: 05/06/13 Time: 15:21
Sample: 1 62
Included observations: 62

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11508.53	5404.199	2.129554	0.0382
FDSFER	4179.207	2010.085	2.079119	0.0428
HORA	-2745.454	1270.368	-2.161149	0.0355
MOM	11075.21	3045.174	3.636972	0.0007
CLASS	-2809.181	2132.809	-1.317128	0.1938
R1	2292.176	642.1075	3.569769	0.0008
R2	248.2397	667.3353	0.371986	0.7115
R3	1149.723	602.4533	1.908401	0.0621
PS	55913.66	168558.3	0.331717	0.7415
QADV	-4.321363	20.59309	-0.209845	0.8346
IMPJOGO2	-10.89403	10.41980	-1.045512	0.3008
LOGD5C	-2204.164	2311.876	-0.953409	0.3450
R-squared	0.588248	Mean dependent var	12835.55	
Adjusted R-squared	0.497662	S.D. dependent var	8407.318	
S.E. of regression	5958.752	Akaike info criterion	20.39510	
Sum squared resid	1.78E+09	Schwarz criterion	20.80680	
Log likelihood	-620.2480	Hannan-Quinn criter.	20.55674	
F-statistic	6.493844	Durbin-Watson stat	1.283399	
Prob(F-statistic)	0.000001			

Tabela 4. Resultados da Regressão Linear utilizando o MQO. Fonte: Os autores.

Vemos na tabela 4 que o coeficiente R^2 (0,588) e o R^2 ajustado (0,498) são razoáveis se levarmos em conta a complexidade do fenômeno estudado.

Sobre os sinais dos coeficientes, todos estão de acordo com o imaginado (por exemplo: o sinal negativo do coeficiente de *Qadv* diz que quanto menor o coeficiente, maior será o público, desde que mantidas as outras variáveis constantes), exceto *PS*. Esperávamos um sinal negativo para o coeficiente *PS*, pois acreditamos que, quanto maior for a razão Preço/Salário, menor será o interesse do torcedor, já que o ingresso ficará mais caro relativo ao salário médio (e consideramos que o torcedor, enquanto consumidor, tem preferência por preços menores).

Quanto à significância das variáveis, são dignas de destaque as variáveis *FDSFer*, *Mom*, *R1* e *Hora*, que são todas estatisticamente significativas ao nível de 5%. A variável *R3* chega perto, com p-valor = 0,0621.

As demais variáveis, no entanto, não são estatisticamente significativas, o que nos leva a já pensar em como evoluir o modelo e melhorar a estimação do público.

6. Próximos passos

Pela dificuldade, pelo fato de ser um trabalho em andamento, e até pelo grau de inovação do tema do trabalho, os resultados foram considerados satisfatórios, ainda que se ressalte a baixa significância de alguns parâmetros do modelo de previsão, assim como a alta variância do modelo.

Alguns pontos devem ser analisados de forma a amadurecer o modelo. Entre eles, podem ser destacados:

- A possibilidade de incluir no modelo jogos de Copa do Brasil e Sul-Americana, que normalmente ocorrem durante a semana (aumentaria a base de dados, que é pequena);

- Estimar o rendimento nominal médio da torcida do Botafogo no Rio de Janeiro, com base em pesquisas de torcidas;
- Tentar aumentar a base de dados com dados referentes a 2009, cujos borderôs não são disponibilizados pelas páginas na internet do Botafogo de Futebol e Regatas ou da Confederação Brasileira de Futebol. A interdição do Engenhão dificulta o aumento da base, já que os jogos do Campeonato Brasileiro de 2013 não serão disputados no estádio, ao menos no início da competição;
- Apesar de parecer uma potencial variável, o fato de o jogo ter transmissão televisiva não deve alterar o modelo, já que o contexto brasileiro é de não serem transmitidos jogos em TV aberta para as cidades em que os jogos ocorrem. Os jogos só são transmitidos para o sistema de sistema de *pay-per-view*, sempre dentro de uma TV por assinatura. E como o sistema de *pay-per-view* passa todos os jogos do campeonato, não há variação e, logo, não há uma variável. Além disso, Kuypers (1996) e Souza (2004), concluíram que a transmissão televisiva não é significativa estatisticamente para explicar o público de jogos de futebol.
- A variável *PS* tenta medir o peso do ingresso no salário médio. Mas falta estimar também o custo médio de vida do cidadão do Rio de Janeiro. Uma variável neste sentido poderia dizer se o preço do ingresso está ou não caro, considerando quanto o cidadão ganha e gasta.
- Construir uma variável que capte os efeitos da variação do tempo (chuva forte, sol forte) pode ajudar a estimar melhor o público. É natural imaginar que, por exemplo, dado uma ocorrência (ou até uma previsão) de chuva forte, alguns torcedores desistam de ir ao estádio de futebol, preferindo ficar em casa.
- Algumas variáveis já descritas aqui devem ser reconstruídas e reavaliadas, dado que muitas não são estatisticamente significativas. A qualidade do adversário (*Qadv*), por exemplo, merece imediata atenção.

Um ponto que merece destaque é o desejo, conforme dito, de evoluir este trabalho também com a construção de um simulador das filas de entradas no estádio. Para cada entrada, há um limite de roletas e de público. A tabela 5 mostra a capacidade por setor tal como o número de roletas.

Setor	Catracas	Capacidade aproximada de público
Norte	12	4500
Sul	12	4500
Oeste	36	18000
Leste	32	18000

Tabela 5: Capacidade e número de catracas por setor. Fonte: Os autores.

É possível notar que não há número exato de roletas para os setores Leste Inferior, Leste Superior, Oeste Inferior e Oeste Superior: as roletas estão alocadas nos setores Leste e Oeste, e podem ser deslocadas.

O simulador faria cálculos de filas M/M/s, com o parâmetro “s” sendo o número de roletas (muda conforme o setor que é simulado).

Desejamos chegar, no futuro, a uma integração do modelo de previsão com o simulador das filas, como esquematizado na figura 2.

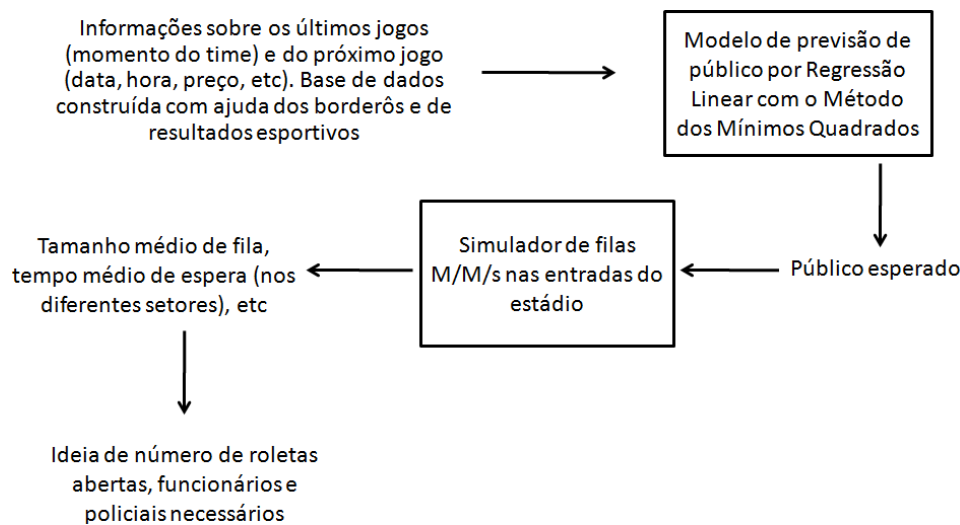


Fig. 2. A futura integração da previsão de público com a simulação das filas de entrada.

Fonte: Os autores.

Conforme vemos na figura 2, as informações contidas nos borderôs e em resultados esportivos servem como *entrada* para o modelo de regressão. O público previsto pela regressão é uma *saída* do modelo de regressão ao mesmo tempo em que seria (de acordo com a integração desejada) uma *entrada* para o simulador das filas, cuja *saída* forneceria informações referentes às filas de entrada. Alterando o número de catracas no simulador, poderíamos chegar a informações e parâmetros como tempo médio e máximo de espera, e assim determinar o número de catracas que serão abertas, bem como o número de operadores de catraca (e até policiais).

Para o simulador, precisaríamos, no entanto, de valores como a taxa de serviço (a taxa de entrada dos torcedores pelas roletas) e a taxa de chegada dos torcedores ao estádio de acordo com o jogo. Se por um lado a interdição do Engenhão dificulta a estimação da taxa de serviço (pois não há jogos ocorrendo), por outro lado o Botafogo fez um estudo que divide o público total de alguns jogos pela faixa de horário em que entraram no estádio (o que facilita a estimação da taxa de entrada).

Em um trabalho posterior, esperamos mostrar o uso do simulador, talvez já acoplado ao modelo de previsão de público, que também sofrerá mudanças, visto que ainda está em desenvolvimento.

Referências

- Alavy, K., Gaskell, A., Leach, S. e Szymanski, S. (2010)**, On the Edge of Your Seat: Demand for Football on Television and the Uncertainty of Outcome Hypothesis, *International Journal of Sport Finance*, 5, p. 75-95.
([http://193.146.160.29/gtb/sod/usu/\\$UBUG/repositorio/10301497_Alavy.pdf](http://193.146.160.29/gtb/sod/usu/$UBUG/repositorio/10301497_Alavy.pdf))
- Bortoluzzo, A. B., Iaropoli, P. T. e Machado, S. J.**, Demand for Brazilian Soccer: A Censored Model Approach, *Working Paper*, n. 237, Insper-SP, 2011
(http://en.insper.edu.br/sites/default/files/2011_wpe237.pdf).
- Falter, J. M. e Pérignon, C. (2000)**, Demand for football and intramatch winning probability: An essay on the glorious uncertainty of sports, *Applied Economics*, 32, 1757-1765.
- Forrest, D. e 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)*, 51, p. 229-241.
- Kuypers, T.**, The beautiful game? An econometric study of audiences, gambling and efficiency in English Football, *Economics PhD Tesis*, University College London, 1996.
- García, J. e Rodríguez, P. (2002)**, The Determinants of Football Match Attendance Revisited : Empirical Evidence From the Spanish Football League, *Journal of Sports Economics*, 3, p. 18-38.
- Greene, W. H.**, *Econometric Analysis*, Prentice Hall, New Jersey, 2003.

Gujarati, D., *Econometria Básica*, Elsevier, Rio de Janeiro, 2006.

Hillier, F. e Lieberman, G.; *Introdução à Pesquisa Operacional*, Mc Graw Hill, São Paulo, 2010.

Madalozzo, R. e Villar, R. B., A Model of Attendance Demand at the Brazilian Football League, *Working Paper*, n. 113, Insper-SP, 2008

(http://www.insper.edu.br/sites/default/files/2008_wpe113.pdf).

Madalozzo, R. e Villar, R. B. (2009), Brazilian football: what brings fans to the game?, *Journal of Sports Economics*, 10 , p. 639-650.

Santana, S. K. S. e Silva, A. S., Fatores determinantes para a demanda por jogos de futebol nos estádios pernambucanos, *XV Congresso de Iniciação Científica da UFPE*, 2007.

([http://www.contabeis.ufpe.br/propesq/images/conic/2007/CONIC_JOIC_2007%20\(E\)/conic/n_pibic/60/60324.pdf](http://www.contabeis.ufpe.br/propesq/images/conic/2007/CONIC_JOIC_2007%20(E)/conic/n_pibic/60/60324.pdf))

Souza, F. A. P., Um estudo sobre a demanda por jogos de futebol nos estádios brasileiros, *Dissertação de mestrado*, FEA-USP, 2004

(<http://www.teses.usp.br/teses/disponiveis/12/12139/tde-21072004-151604/pt-br.php>).

Szymanski, S. (2011), Income inequality, competitive balance and the attractiveness of team sports: some evidence and a natural experiment from English soccer, *The Economic Journal*, 111, p. 69-84 (<http://onlinelibrary.wiley.com/doi/10.1111/1468-0297.00599/pdf>).

http://www.bfr.com.br/futebol/ultimos_borderos.asp, acessado em 05/12/2012

http://www.ibge.gov.br/home/estatistica/indicadores/trabalhoerendimento/pme_nova/defaulttab_hist.shtm, acessado em 15/03/2012

<http://zerohora.clicrbs.com.br/rs/esportes/brasileirao/pagina/historico-de-desempenho.html>, acessado em 20/03/2012