



PAÍS DO FUTEBOL? A PORCENTAGEM DE OCUPAÇÃO NOS ESTÁDIOS E A SUA VARIABILIDADE NO CAMPEONATO BRASILEIRO 2018¹

Luan Azevedo² (UEPG), luanlim@gmail.com
Lucas Marchioro³ (PUCPR), lucas_marchioro93@gmail.com
Luis Rodrigo Troyner⁴ (PUCPR), rodrigotroyner@gmail.com

RESUMO: O artigo tem como problema de pesquisa explicar quais as relações entre as principais variáveis que determinam a porcentagem de ocupação nos estádios em 2018. Partindo de um problema empírico circunscrito ao Campeonato Brasileiro - baixa adesão do público nos estádios de futebol no principal campeonato nacional - pretendemos investigar as variáveis via metodologia quantitativa que indicam a ocupação nas arenas, apresentando os motivos da variação de público nas arquibancadas no 'Brasileirão 2018'. Destaca-se como resultado positivo da qual foi possível conhecer a porcentagem de 66% do público que compareceu ao campeonato de acordo com os parâmetros propostos nesse estudo

Palavras-Chave: Futebol. Ocupação dos estádios. Campeonato Brasileiro 2018. Métodos Quantitativos.

ABSTRACT: The article has as a research problem to explain the relations between the main variables that determine the percentage of occupation in the stadiums in 2018. Starting from an empirical problem circumscribed to the Brazilian Championship - low adhesion of the public in the soccer stadiums in the main national championship - we intend to investigate the variables through quantitative methodology that indicate the occupation in the arenas, presenting the reasons of the variation of the public in the bleachers in the 'Brasileirão 2018'. It stands out as a positive result of which it was possible to unveil the percentage of 66% of the public that attended the championship according to the parameters proposed in this study.

Key-Words: Soccer. Occupation of the stadiums. Campeonato Brasileiro 2018. Quantitative methods.

ÁREA 8: Métodos Quantitativos

JEL (classificação JEL): C150 Statistical Simulation Methods: General

¹ Trabalho final apresentado como requisito parcial para obtenção do título de Especialista em Estatística com Ênfase em Pesquisa PUC-PR.

² Graduando em Economia (UEPG) e Mestre em Ciência Política (UFPR).

³ Bacharel em Economia (UFPR) e Especialista e Estatística (PUC-PR).

⁴ Bacharel em Economia (Faculdades Santa Cruz), MBA em Gestão Estratégica (UFPR) e Especialista e Estatística (PUC-PR).

1. INTRODUÇÃO

1.1 O PARADOXO DO FUTEBOL NO BRASIL

O Brasil não só é conhecido como país do futebol, tal como suas referências o declaram, como também carrega o prestígio de melhor do mundo com sua seleção que apresenta um brasão com cinco estrelas ao lado esquerdo de sua camisa amarela, a mais vitoriosa nas competições quadrienais.

Fora das quatro linhas do campo graminado, o futebol no país representa também uma significativa importância social pelos números e cifras que são envolvidas nesse esporte (GASTALDO, 2009). O tema também é omnipresente nas conversas diárias e nos jornais com editoriais exclusivas para o debate público, (COELHO & TIESLER, 2006) tamanho se trata da atenção de uma nação voltada para o esporte.

Outros argumentos que justificam em termos materiais a particularidade do que o futebol representa no contexto brasileiro seria a grade de canais que falam sobre o assunto, o crescimento das cifras envolvidas e a visão social do esporte. O primeiro fundamento que releva o tema de estudo pode ser encontrado nas horas dedicadas nas transmissões televisivas. Com dados recuperados do IBOPE, até 2012, a modalidade representava 54% do total das transmissões em comparação com todos os outros esportes, que se somava num total anual de 29 mil horas nos canais abertos e fechados⁵. No que se refere ao dinheiro que gira em torno do Esporte, só a CBF (Confederação Brasileira de Futebol), tem um orçamento maior que o Ministério dos Esportes no Brasil, em torno de meio bilhão de reais⁶. Outros clubes da série A também já ultrapassam esses valores.

Por fim, diversos estudos apresentam que o futebol se torna a única forma encontrada como possibilidade de ascensão social para as camadas menos privilegiadas na sociedade (GUEDES, 1982; DAMO, 2005; SOARES 2003; SOUZA et al, 2008).

Apesar dessa popularidade, vista em várias dimensões, televisivos e na linguagem cotidiana, o futebol pouco se caracteriza presente na realidade dos estádios, ou seja, parafraseando Nelson Rodrigues, no palco onde o drama, a tragédia, o horror e a compaixão⁷ reinam, temos ainda um número reduzido de espectadores nas arquibancadas para acompanhar os jogos, tendo tal importância no país que se assume como principal no esporte mais praticado no mundo.

Para fins de comparação, temos as taxas de ocupação de 2012 nos campeonatos nacionais também mais assistidos do mundo com uma média muito próxima da ocupação total. A Bundesliga, campeonato nacional da Alemanha, conta com a maior média de espectadores, atingindo 95% de ocupação. Em seguida está a Inglaterra com também 95% de ocupação, contando com estádios um pouco menores que os alemães. Em terceiro se coloca o campeonato espanhol, com 83% de ocupação⁸. Porém, em 2018, o maior campeonato nacional brasileiro, bateu seu recorde de ocupação nos estádios desde que passou para pontos corridos (2003). Todavia, paradoxalmente, isso representa só 43% da porcentagem de público que esteve nos estádios.

Esse estudo ainda revela que o Campeonato Brasileiro fica apenas na 15ª posição no que se refere a ocupação dos estádios em comparação com outras ligas, atrás das segundas divisões da Alemanha e Inglaterra e de países que possuem uma tradição muito menor no futebol, como Estados Unidos, China, Japão e Turquia (AGUIAR e de SÁ RIBEIRO, 2017).

⁵ Dados retirados de: <http://www.ibope.com/pt-br/noticias/Paginas/Aumenta-o-total-das-horas-de-esportes-transmitidos-na-TV.aspx>

⁶ A Pasta perdeu a legitimidade de Ministério em 2019, mas tem seu orçamento previsto em 478 milhões, enquanto os gastos da CBF giram em torno de 545 milhões. Fonte: <https://oglobo.globo.com/politica/orcamento-do-ministerio-do-esporte-quintuplica-em-oito-anos-movimenta-5-bilhoes-3080037>

CBF: <https://epoca.globo.com/esporte/epoca-esporte-clubes/noticia/2018/06/cbf-fatura-alto-com-selecao-brasileira-eis-de-onde-vem-e-para-onde-vai-o-dinheiro.html>

⁷ O autor faz uma alusão em seu livro da teoria clássica do Drama de Aristóteles na obra Poética. RODRIGUES, Nelson. **À Sombra das Chuteiras Imortais**. São Paulo: Cia das Letras, 1993.

⁸ Estudo realizado pela consultoria Esporte Pluri. Disponível em: <https://gq.globo.com/Essa-e-nossa/noticia/2013/07/os-20-campeonatos-com-maior-media-de-publico-no-mundo.html>

Podemos também ver na tabela abaixo a crescente média anual da ocupação de público nos últimos dez anos, da qual pouco variou nos números atuais:

TABELA 1 – MÉDIA DE OCUPANTES ANUAL DO CAMP. BRASILEIRO

2008	16.992
2009	17.807
2010	14.800
2011	14.976
2012	13.148
2013	14.969
2014	16.537
2015	17.051
2016	15.219
2017	15.961
2018	18.821

Fonte: Autores

Dado o entendimento dessa problemática e os fatos apresentados sobre o futebol brasileiro, esse estudo debruça-se em explorar os motivos da variação do público e caracterizar a realidade da ocupação dos torcedores nos estádios durante o campeonato Brasileiro de 2018. Dessa forma, temos como problema principal de nossa pesquisa: Quais as relações entre as principais variáveis que determinam a % de ocupação nos estádios no Campeonato Brasileiro 2018?

No próximo tópico, apresentaremos quais serão os objetivos desse estudo. Em seguida, justificaremos a abordagem dado ao tema, a forma como os dados foram estudados e quais as delimitações do trabalho que qualificam a pesquisa.

OBJETIVO DO ESTUDO

O objetivo principal dessa pesquisa é determinar as variáveis que explicam a oscilação de público de todos os estádios utilizados pelas equipes de futebol ao longo da 1ª divisão do Campeonato Brasileiro da série A 2018.

OBJETIVOS ESPECÍFICOS

Tendo o objetivo principal como norteador, os objetivos específicos servirão para indicar as etapas a serem seguidas de nossa investigação:

- Levantamento dos dados sobre os jogos e torcedores em 2018;
- Explorar o comparecimento aos estádios;

- Analisar as relações entre as principais variáveis;
- Comparar a presença dos torcedores nos estádios entre as equipes;

2. DELIMITAÇÃO: O FOCO NOS ESTÁDIOS

Este estudo tem como foco a ocupação dos estádios de futebol utilizados pelos times na 1ª Divisão do Campeonato Brasileiro em 2018. Dada a pouca produção científica de estudos sobre o tema (COELHO & TIESLER, 2006; BECCARINI & FERRAND, 2006), e a escassez do diagnóstico que vá além do jornalismo esportivo com números absolutos, raros são os esforços que buscam novas respostas dedicadas ao assunto (SOUZA, 2004; AGUIAR, 2017;). Com isso, dentro dessa análise bibliográfica, esta investigação se torna exploratória, da qual atribuímos a tarefa de relacionar alguns aspectos relevantes da participação do público nos estádios com outras dimensões como elenco, valores e tipos de arenas⁹, portanto, relacionando com o objetivo balizador desse artigo. De acordo com ventura:

São úteis também na exploração de novos processos ou comportamentos, novas descobertas, porque têm a importante função de gerar hipóteses e construir teorias. Ou ainda, pelo fato de explorar casos atípicos ou extremos para melhor compreender os processos típicos (pp. 386).

Os dados foram levantados ao final do campeonato, tendo a principal base de informações, no que diz respeito a taxa de ocupação e valores médios dos ingressos, o site da Globo Esporte¹⁰. Um segundo importante banco de dados foi o site português transfmarket¹¹, especializado nos valores monetários que circula nesse esporte. Por fim, uma última informação que foi considerada importante para explorar o comparecimento de torcedores nas arenas de futebol foram as condições climáticas no dia e no momento do jogo¹².

Quanto às justificativas, podemos elencar algumas acerca da questão. A primeira se resume na importância da viabilidade financeira dos estádios, dos quais os profissionais esportivos responsáveis pelas arenas utilizam com frequência dados para viabilizar e diminuir os prejuízos financeiros do custo operacional do estádio. Desta forma, compreender algumas causas da ocupação dos torcedores permite que os encarregados possam analisar meios e ações midiáticas tomadas na frequência de público nos estádios brasileiros.

A opção por estudar apenas o Campeonato Brasileiro de 2018, sem considerar outros anos, e trabalhando apenas com a primeira divisão se baseia em alguns critérios. Um deles seria a falta de embasamento teórico que nos possibilite entender quais variáveis teriam maior peso para abarcar a porcentagem de ocupação dos estádios no Brasil. Assim, selecionamos diversas possibilidades dentro do universo do tema para ter um “pontapé inicial” na construção de respostas para o problema de pesquisa, tipificando como um estudo exploratório.

Delimitamos os resultados apenas para o ano de 2018, pois conseguiríamos aumentar o índice de explicação para o comparecimento de torcedores nos estádios, diminuindo fatores externos se optássemos por fazer uma pesquisa prospectiva por exemplo (GIL, 2008).

A escolha pelo Campeonato Brasileiro foi feita por ter mais rodadas com intensidade de público semelhante para cada time em todas as rodadas, permitindo uma distribuição de jogos ao longo de dez meses que podem ser melhor avaliada. Todavia, incluímos algumas possíveis variáveis externas ao campeonato, que é capaz de alterar a frequência do comparecimento do torcedor nos estádios, do mesmo modo que a equipe poderia ter jogado outro campeonato no meio de semana por exemplo.

⁹ As variáveis e parâmetros construídos serão melhor apresentados no próximo tópico.

¹⁰ <http://www.globoesporte.com/>

¹¹ transfermarket.pt

¹² weather.com

Outro argumento condiz com as equipes que disputam o campeonato, já que boa parte da renda financeira arrecadada pelos times tem sua origem no público que comparece aos jogos.

Por fim, uma última e relevante justificativa, é do interesse dos patrocinadores, pois com um alto percentual de ocupação média, mais suas marcas serão vistas por seu público, e visando melhorar a qualidade e reconhecimento ao diagnosticar essas nuances com a constatação da variação do público presente nos estádios frente suas placas de propagandas (TREIN & BARCELLOS, 2006)

No próximo tópico, iremos apresentar a hipótese de trabalho e a metodologia utilizada para a coleta dos dados. Além disso, todas as variáveis que foram criadas e obtidas, serão descritas para o entendimento da pesquisa.

3. MÉTODOS E VARIÁVEIS:

O método de pesquisa escolhido para investigação é o quantitativo. Utilizamos como ferramenta a estatística inferencial. Mas fica a pergunta, por que utilizar a estatística? Em nosso caso, foi necessário utilizar a teoria das probabilidades para produzir resultados mais contundentes quanto a pergunta de nosso problema. Como já mencionado antes, como temos como objetivo entender a variação do público torcedor ao longo do campeonato, lidamos com altos números e muitas variáveis, e se quisermos determinar o peso explicativo do que corrobora na ocupação dos estádio, torna-se necessário a exploração das diversas relações entre os fenômenos que circunscrevem o motivo da presença dos torcedores aos jogos (GIL, 2008).

3.1 HIPÓTESE

A partir desse caminho, lançamos a seguinte hipótese, baseada em nossa pesquisa, da qual busca entender quais das variáveis tem maior peso para explicar a taxa de ocupação nos estádios. Acreditamos que conseguiremos um retorno esperado com foco no valor do elenco e número de vitórias seguidas, pois espera-se que a taxa média de ocupação aumente conforme o time ganhe ao longo do campeonato.

3.2 UNIVERSO E VARIÁVEIS

Visando o entendimento acerca dos motivos que buscamos para explicar a presença do público nos estádios de futebol, pretendemos descrever o universo da pesquisa e as variáveis criadas e/ou coletadas para determinar esse fenômeno.

Nosso universo abrange todo o campeonato Brasileiro de 2018. As 20 equipes participaram do processo, tendo em conta todas as 38 rodadas (jogos). Destaca-se que todo o método de coleta foi automatizado, diminuindo o erro humano que poderia ocorrer caso tivéssemos que fazer manualmente.

Cabe dizer, que antes de todo esforço feito para a coleta, fora necessário pensar em como a pesquisa iria ser produzida a partir da relação entre as variáveis. Por meio disso, foram levantadas as diversas informações a respeito da porcentagem de torcedores para entender qual o percentual de cada estádio em um jogo do Brasileirão de 2018. Assim sendo, as variáveis foram determinadas da seguinte maneira:

- *Perc Ocup:* A variável dependente. Continua. Trata-se do Percentual de Ocupação do estádio
- *Aproveitamento:* Variável contínua. Trata-se do aproveitamento do time mandante na competição. É calculada através do número de pontos ganhos no campeonato sobre o número de pontos disputados até a data do jogo em questão.
- *Chuva:* Variável dummy. caso seja 1, estava chovendo tres horas antes do jogo. Caso seja 0, não chveu.

- *Clássico Regional*: Variável dummy. Caso o jogo em questão seja de dois times do mesmo estado, a variável é marcada como 1.
- *Colocação Mandante*: Variável contínua. Colocação do mandante na competição até a data do jogo.
- *Colocação Visitante*: Variável contínua. Colocação do visitante na competição até a data do jogo.
- *Condição*: Variável categórica. Trata-se da condição climática em três horas anteriormente ao jogo. As variáveis podem ser:
 - "Ameno" "Chuva" "Chuva forte" "Chuva leve" "Limpo" "Neblina" "Nublado"
 - "Pós-chuva" "Sol"
- *Derrota Cum*: Variável contínua. Quantidade de derrotas do time na competição até a data do jogo.
- *Derrota Cum T*: Variável contínua. Quantidade de derrotas do time considerando outros campeonatos até a data do jogo.
- *Derrota Emp Seguida*: Variável contínua. Quantidade de derrotas ou empates seguidos do time mandante dentro da competição que o time teve até a data do jogo. Caso o time vença, a variável é zerada.
- *Derrota Emp Seguida T*: Variável contínua. Quantidade de derrotas ou empates seguidos do time mandante considerando outros campeonatos que o time teve até a data do jogo. Caso o time vença, a variável é zerada.
- *Fim de semana*: Variável dummy. Caso o jogo tenha ocorrido em um sábado ou domingo, a variável é apresentada como 1.
- *Dia Do Mes*: Variável contínua. Refere-se a qual dia do mês o jogo ocorreu.
- *Empate Cum*: Variável contínua. Quantidade de empates do time mandante na competição até a data do jogo.
- *Empate Cum T*: Variável contínua. Quantidade de empates do time mandante considerando outros campeonatos até a data do jogo.
- *Feriado*: Variável dummy. Caso o jogo tenha ocorrido em um feriado nacional, a variável é apresentada como 1.
- *Graus*: Variável contínua. Quantos graus celsius estava a temperatura em três horas antes do jogo
- *Hora Var*: Variável contínua. Horário em que ocorreu o jogo.
- *Inau Estadio*: Variável contínua. Idade do estádio em que ocorreu o jogo
- *Jogou Outro Camp Na Semana*: Variável dummy. Caso o time mandante tenha jogado um jogo de outro campeonato na mesma semana, é apresentada como valor 1.
- *Mandante*: Variável categórica. Qual era o time mandante.
- *Mes*: Variável contínua. Qual mes ocorreu o jogo.
- *Pib Cid Estadio*: Variável contínua. Valor do PIB da cidade em que se realizou o jogo.
- *Qtd Sócios Mandante*: Variável contínua. Quantidade de sócios do time mandante.
- *Ranking Torcida Mandante*: Variável contínua. Ranking do time da torcida mandante. O time com maior torcida tem o valor de 1 e com menor torcida tem o valor de 20.

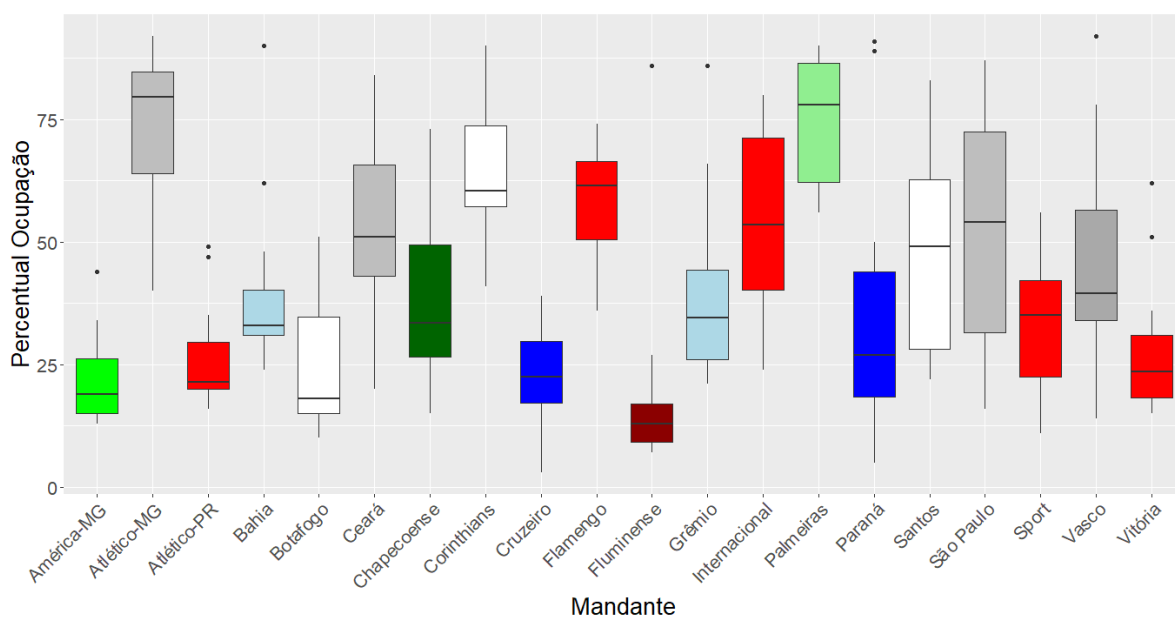
- *Ranking Torcida Visitante*: Variável contínua. Ranking do time da torcida visitante. O time com maior torcida tem o valor de 1 e com menor torcida tem o valor de 20.
- *Renda Media Cid*: Variável contínua. Renda média da cidade em que se realizou o jogo.
- *Resultado Anterior*: Variável categórica. Os valores representam o resultado do último jogo na competição. Podem ser:
 - “Vitória” “Empate” “Derrota”
 - Na primeira rodada, todos os valores desta variável estão como “0”
- *Rodada*: Variável contínua. Qual rodada da competição o jogo está acontecendo.
- *Semana Ano*: Variável contínua. Mostra em qual semana do ano ocorreu o jogo.
- *Semana Do Mes*: Variável contínua. Mostra em qual semana do mês ocorreu o jogo.
- *Sg Contra*: Variável contínua. Saldo de gols contra do time mandante considerando apenas a competição Campeonato Brasileiro 2018.
- *Sg Contra T*: Variável contínua. Saldo de gols contra do time mandante considerando outras competições.
- *Sg Pro*: Variável contínua. Saldo de gols a favor do time mandante considerando apenas a competição Campeonato Brasileiro 2018.
- *Sg Pro T*: Variável contínua. Saldo de gols a favor do time mandante considerando outras competições.
- *Ticket*: Variável contínua. Valor médio da receita bruta do estádio.
- *Vitoria Seguida*: Variável contínua. Considera a quantidade de vitórias seguidas do time mandante considerando apenas a competição Campeonato Brasileiro 2018. Em caso de derrota ou empate, a variável é zerada.
- *Vitoria Seguida T*: Variável contínua. Considera a quantidade de vitórias seguidas do time mandante considerando outras competições. Em caso de derrota ou empate, a variável é zerada.
- *Vitorias Cum*: Variável contínua. Considera a quantidade de vitórias acumuladas do time mandante considerando apenas a competição Campeonato Brasileiro 2018.
- *Vitorias Cum T*: Variável contínua. Considera a quantidade de vitórias acumuladas do time mandante considerando outras competições
- *Vlr Elenco mandante M*: Variável contínua. Valor do elenco mandante em milhões de dólares.

Através da criação de variáveis e parâmetros utilizados para análise, foram obtidos alguns resultados que nos levam a entender a variação da ocupação média nos estádios brasileiros de futebol. Desta forma, o próximo tópico ilustra os dados obtidos, assim como o seu resultado final.

4. RESULTADOS

A variável dependente possui alta dispersão por time, conforme se verifica no gráfico abaixo:

GRÁFICO 1 – DISPERSÃO DO PERCENTUAL DE OCUPAÇÃO DO ESTÁDIO POR TIME



Fonte: Autores.

TABELA 2 – DISTRIBUIÇÃO DO PERCENTUAL DE OCUPAÇÃO DOS TIMES

Mandante	Minimo	Q1	Mediana	Q3	Maximo	Diff_Q3_Q1	Diff_Max_Min
América-MG	13%	15%	19%	26%	44%	11%	31%
Atlético-MG	40%	64%	80%	85%	92%	21%	52%
Atlético-PR	16%	20%	22%	30%	49%	10%	33%
Bahia	24%	31%	33%	40%	90%	9%	66%
Botafogo	10%	15%	18%	35%	51%	20%	41%
Ceará	20%	43%	51%	66%	84%	23%	64%
Chapecoense	15%	27%	34%	50%	73%	23%	58%
Corinthians	41%	57%	61%	74%	90%	17%	49%
Cruzeiro	3%	17%	23%	30%	39%	13%	36%
Flamengo	36%	51%	62%	67%	74%	16%	38%
Fluminense	7%	9%	13%	17%	86%	8%	79%
Grêmio	21%	26%	35%	44%	86%	18%	65%
Internacional	24%	40%	54%	71%	80%	31%	56%
Palmeiras	56%	62%	78%	87%	90%	24%	34%
Paraná	5%	19%	27%	44%	91%	26%	86%
Santos	22%	28%	49%	63%	83%	35%	61%
São Paulo	16%	32%	54%	73%	87%	41%	71%
Sport	11%	23%	35%	42%	56%	20%	45%

Vasco	14%	34%	40%	57%	92%	23%	78%
Vitória	15%	18%	24%	31%	62%	13%	47%

Fonte: Autores.

A diferença média entre o mínimo e máximo de cada time é de 54,5%, enquanto a diferença média interquartil é de 19,9%.

A dispersão entre os times também é grande. Enquanto alguns times, como Atlético-MG e Palmeiras, tem percentual de ocupação mediano por jogo acima de 75%, muitos times tiveram percentual de ocupação mediano abaixo de 25% durante o campeonato Brasileiro de 2018.

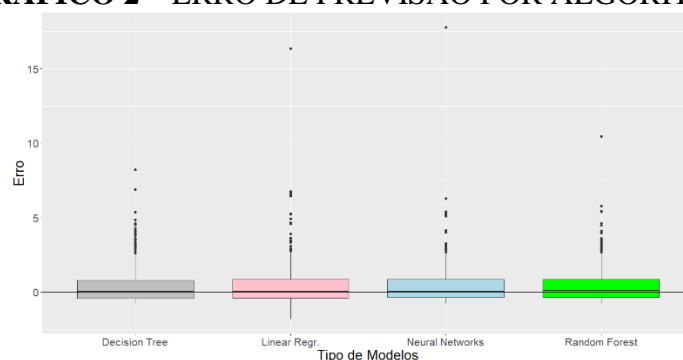
Para entender o quanto e quais variáveis importam mais determinar o percentual de ocupação de cada jogo do Campeonato brasileiro 2018, foram utilizados 4 algoritmos: regressão linear, árvores de decisão, random forest com 500 repetições e redes neurais de tamanho 3 e com 100 repetições.

Para determinar o erro que temos ao tentar estimar e ver quais variáveis são mais importantes, foi utilizado o método de cross-validation. Este método consiste em quebrar a base em várias partes, utilizar uma grande porcentagem para treino e o resto do percentual usar como teste, buscando prever os resultados.

Neste caso, dividimos a base e aplicamos os modelos dez vezes, treinando com 90% da base e depois testando nos 10% restantes da base. Desta forma, após as dez vezes de treinamento e teste, temos previsão para todos os jogos.

O erro das previsões fica da seguinte forma:

GRAFICO 2 – ERRO DE PREVISÃO POR ALGORITMO



Apesar de ter um erro na mediana muito próximo de 0, o erro dos modelos ficou enviesado com os outliers do modelo, com MAPE próximo de 80%. Os maiores outliers são em decorrência dos jogos com baixo percentual de ocupação, sendo o ponto de maior erro o jogo no Mineirão em que o Cruzeiro teve apenas 3% do percentual de ocupação, e todos os jogos com mais de 300% de erro foram jogos que tiveram no máximo 14% de ocupação do estádio.

Em relação aos modelos, os 4 algoritmos tiveram performance muito parecida. Conforme mostra tabela abaixo:

TABELA 3 – PERFORMANCE POR MODELO

	Modelo			
	LR	DT	RF	RN
Média	49,67%	46,23%	47,24%	50,28%

Q1		-38,74%	-42,66%	-35,17%	-35,12%
Mediana		5,24%	4,33%	4,58%	4,63%
Q3		86,86%	78,89%	85,62%	86,08%
MAPE	5%				
trim		79,48%	76,77%	72,16%	73,24%

Fonte: Autores.

Ao gerar a previsão, porém, foi alertada a presença de multicolineariedade entre as variáveis. Por isso, fizemos um quadro de correlação e retiramos as variáveis que possuem correlação acima de 0,9 entre si.

TABELA 4 – CORRELAÇÕES ENTRE VARIÁVEIS

Var1	Var2	value
semana_ano	RODADA	0,99
RODADA	Mes	0,98
semana_ano	Mes	0,99
semana_do_mes	dia_do_mes	0,97
vitorias_cum	SG_PRO	0,95
SG_PRO_T	SG_PRO	0,93
SG_CONTRA_T	SG_CONTRA	0,97
vitorias_cum_T	vitorias_cum	0,94
RODADA	empate_cum	0,91
Mês	empate_cum	0,91
semana_ano	empate_cum	0,92
empate_cum_T	empate_cum	0,93
SG_CONTRA	derrota_cum	0,94
SG_CONTRA_T	derrota_cum	0,91
derrota_cum_T	derrota_cum	0,97
derrota_emp_seguida_T	derrota_emp_seguida	0,96
vitorias_cum	SG_PRO_T	0,93
vitorias_cum_T	SG_PRO_T	0,97
SG_CONTRA	derrota_cum_T	0,93
SG_CONTRA_T	derrota_cum_T	0,95

Fonte: Autores.

TABELA 5 – VARIÁVEIS SELECIONADAS

var1	var2	var3	var4	var5
Semana_ano	semana_do_mes	vitorias_cum	empate_cum	SG_CONTRA_T
Mês	dia_do_mes	SG_PRO	empate_cum_T	SG_CONTRA
RODADA		SG_PRO_T		derrota_cum_T
		vitorias_cum_T		derrota_cum

Fonte: Autores.

As variáveis em amarelo foram as selecionadas para se manter no modelo, pois foram as que tiveram maior grau de significância na regressão linear treinada acima.

Ao retirar as variáveis e rodar o modelo de novo, os resultados são os seguintes:

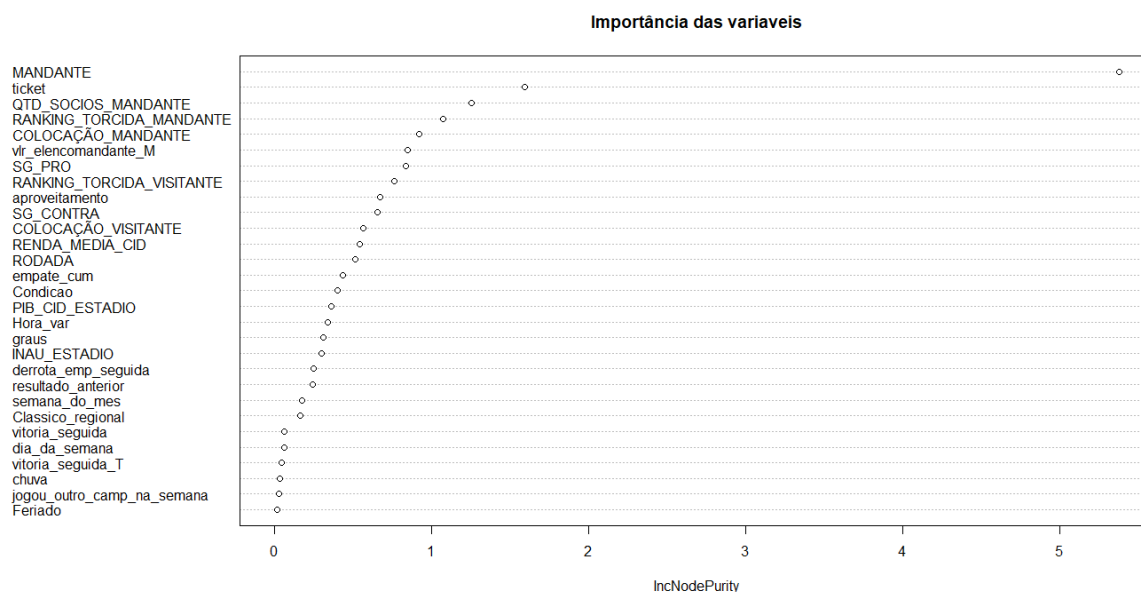
TABELA 6 – RESULTADOS APÓS RETIRADA DE VARIÁVEIS COM MULTICOLINEARIDADE:

	Modelo			
	LR	DT	RF	RN
Média	49,71%	44,82%	46,93%	50,29%
Q1	-38,89%	-42,43%	-34,96%	-34,71%
Mediana	3,51%	5,01%	4,70%	6,01%
Q3	88,25%	77,63%	86,42%	87,60%
MAPE (5% trim)	78,31%	74,55%	71,62%	73,15%

Fonte: Autores.

Como Random Forest continua sendo o modelo com menor MAPE, este modelo será utilizado como base para determinarmos quais são as variáveis mais importantes.

GRÁFICO 3 – IMPORTÂNCIA DAS VARIÁVEIS



Fonte: Autores.

Conforme vemos no gráfico, as variáveis mais importantes são: quem é o time mandante, qual o ticket, a quantidade de sócios do time mandante, qual o ranking da torcida mandante, o valor do elenco mandante, a quantidade de gols a favor, o ranking da torcida visitante, o aproveitamento, a colocação do visitante, a renda média da cidade e em qual rodada foi o jogo.

Muitas das variáveis são estáticas de acordo com o time mandante. Dessa forma, rodamos um modelo por time apenas com as variáveis selecionadas acima e o clássico regional. O R² de uma regressão linear buscando determinar o percentual por time através das variáveis descritas acima é a seguinte:

TABELA 7 – R² POR MODELO DE CADA TIME

Time	R ²
América-MG	0,5994364
Atlético-MG	0,7448472
Atlético-PR	0,7875274
Bahia	0,9938048
Botafogo	0,5152151
Ceará	0,7952503
Chapecoense	0,5143482
Corinthians	0,588054
Cruzeiro	0,7025539
Flamengo	0,3681814
Fluminense	0,9561606
Grêmio	0,7684868

Internacional	0,6804985
Palmeiras	0,9319813
Paraná	0,8666023
Santos	0,6917921
São Paulo	0,820022
Sport	0,5551371
Vasco	0,8496148
Vitória	0,863246

Fonte: Autores.

Apesar de muitos modelos terem R^2 acima de 0,75, alguns times não tiveram a variação bem explicada por estas variáveis, como é o caso do Flamengo, Corinthians, Botafogo e Chapecoense. Talvez outras variáveis possam explicar melhor o interesse da torcida em ir ao estádio para estes times, mas de uma maneira geral, as variáveis selecionadas explicam boa parte do comportamento para a maior parte dos clubes que disputaram o Brasileiro de 2018.

CONSIDERAÇÕES FINAIS

A partir dos resultados obtidos é possível concluir alguns fatos acerca dos principais fatores que levam à variação da porcentagem de ocupação média dos estádios brasileiros de futebol. Primeiramente, cabe deixar como observação que por se tratar de um modelo construído com cunho exploratório, podemos afirmar que foram positivos os resultados alcançados. Sabe-se que existem diversos fatores, que combinados, que podem incrementar mais o modelo sugerido para prospecção da ocupação nos estádios.

Os primeiros testes feitos para entender se havia um padrão entre a ocupação percentual e sua variação ao longo do campeonato se mostrou pouco compreensivo em termos absolutos. Em outras palavras, apenas visualizando a variação da média ao longo do ano com as frequências, sem relacionar as variáveis, enxergou-se que a dispersão entre os times é alta. Porém, ao aplicarmos os algoritmos de regressão com repetições para averiguar os erros que podem ser decorrentes dos testes, conseguimos separar os outliers, ou seja, entender por exemplo que jogos com um baixo percentual de ocupação se torna mais difícil de se explicar.

Respondendo à pergunta de pesquisa, quais as relações entre as principais variáveis que determinam a % de ocupação nos estádios no Campeonato Brasileiro 2018? Concluímos que, em geral, a nossa hipótese foi parcialmente aceita, pois é possível estabelecer que as variáveis determinantes que revelam os estádios com maiores públicos têm relação em sua maioria com a equipe que está jogando, como vitórias, gols, valor do elenco, aproveitamento e o ranking da torcida mandante. Outras variáveis ainda corroboram para a nossa resposta: colocação visitante, renda média da cidade e o ticket médio.

No comparativo entre os times, alguns não foram possíveis de se explicar a variação por meio das determinantes selecionadas, como Flamengo, Corinthians, Botafogo e Chapecoense. Cabe salientar que podemos futuramente aperfeiçoar o nosso modelo para conseguir esclarecer melhor o interesse dessas torcidas em ir ao estádio.

Outros fatores, pouco considerados nesse estudo por estarem atrelados a qualidade de serviços oferecidos ao público que frequenta os estádios, podem ser abrangidos a partir da iniciativa dos gestores das arenas esportivas, ponderando também a dificuldade na compra de ingressos, a violência em torno do estádio, a transmissão da TV e a baixa qualidade nos serviços oferecidos podem evidenciar mais a variação da frequência dos torcedores nos estádios.

Deixamos registrado nesse artigo também que, além de ter seus objetivos de investigação alcançados, apresentamos indiretamente medidas de pesquisa aplicadas que podem contribuir para evitar o despovoamento dos estádios brasileiros. Assim, salienta-se o futebol como parte significativa da sociedade brasileira, e a presença e convivência de torcedores de diversas equipes nas arquibancadas cheias recupera a popularidade¹³ desse esporte.

REFERÊNCIAS BIBLIOGRÁFICAS

AGUIAR, Lucas; de SÁ RIBEIRO, Marco Aurélio. **Um estudo sobre a taxa de ocupação média dos estádios brasileiros de futebol**. Rio de Janeiro, 2017. 40p. Trabalho de Conclusão de Curso – Departamento de Administração, Pontifícia Universidade Católica do Rio de Janeiro.

BECCARINI, C; FERRAND, A. **Factors affecting soccer club season ticket holders' satisfaction: The influence of club image and fans' motives**. *European Sport Management Quarterly*, 6(1):1–22, 2006.

COELHO, João Nuno & TIESLER, Nina Clara. **O paradoxo do jogo português: a omnipresença do futebol e a ausência de espectadores dos estádios**. *Anál. Social [online]*. 2006, n.179 - pp.519-551

DAMO, A. S. **Do dom à profissão: uma etnografia do futebol de espetáculo a partir da formação de jogadores no Brasil e na França**. Tese (Doutorado em Antropologia Social) –Programa de Pós-Graduação em Antropologia Social, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005.

GASTALDO, Édison. **“O país do futebol” mediatizado: mídia e Copa do Mundo no Brasil**. *Sociologias*, Porto Alegre, ano 11, nº 22, jul./dez. 2009, p. 352-369

GIL, A. C. **Métodos e técnicas de pesquisa social**. 2008

GUEDES, S. L. **Subúrbio: celeiro de craques**. In: DAMATTA, R. (Org.). *Universo do futebol: esporte e sociedade brasileira*. Rio de Janeiro: Pinakotheke, 1982. p. 59-74.

SOARES, A. J. G.; LOVISOLO, H. R. **Futebol: a construção histórica do estilo nacional**. *Revista Brasileira de Ciências do Esporte*, Campinas, v. 25, n. 1, p. 129-144, 2003.

SOUZA, Camilo Araújo Máximo de et al. **Difícil reconversão: futebol, projeto e destino em meninos brasileiros**. *Horiz. antropol.*, Porto Alegre, v. 14, n. 30, p. 85-111, Dec. 2008.

SOUZA, Fabio Augusto Pera de. **Um estudo sobre a demanda por jogos de futebol nos estádios brasileiros**. 2004. Dissertação (Mestrado em Administração) - Faculdade de Economia, Administração e Contabilidade, University of São Paulo, São Paulo, 2004. doi:10.11606/D.12.2004.tde-21072004-151604. Acesso em: 2019-01-11

¹³ Condição do que é conhecido e amado por um grande número de pessoas; glória, renome, celebridade.

TREIN, F.; BARCELLOS, P. **Qualidade dos serviços em estádios de futebol.** Anais do Encontro Nacional de Engenharia de Produção, Fortaleza, CE, Brasil, 26, 2006.

VENTURA, M. M. **O Estudo de Caso como Modalidade de Pesquisa.** REVISTA Pedagogia Médica SOCERJ. 2007, pp. 383-386.