



PROPOSIÇÃO DE MODELO DE GERAÇÃO DE VIAGENS PARA ESTÁDIOS DE FUTEBOL BRASILEIROS BASEADOS EM DADOS SECUNDÁRIOS

João Guilherme da Costa Braga França

João Victor de Oliveira Miranda

Universidade Federal de Minas Gerais

Leise Kelli de Oliveira

Universidade Federal de Minas Gerais

Universidade Federal de Pernambuco

Isabela Kopperschmidt de Oliveira

Universidade Federal de Pernambuco

RESUMO

Os estádios de futebol são grandes polos geradores de viagem, principalmente no contexto brasileiro em que o esporte tem grande repercussão. Por isso, este artigo teve como objetivo estimar modelos de geração de viagens para os estádios de futebol brasileiros baseados em dados secundários. O público pagante foi usado como *proxy* do número de viagens e os resultados indicaram que a capacidade dos estádios e o valor do *ticket* médio tem influência positiva na geração de viagens. Como nem todos os modelos estimados pelo método dos mínimos quadrados ordinários atenderam os pressupostos de linearidade, pela identificação de heterocedasticidade (erros não são constantes), alternativamente, foram estimados modelos pela regressão robusta. Apesar do uso de dados secundários, os resultados contribuem com a literatura referente à esses polos e com o planejamento de transportes.

ABSTRACT

Soccer stadiums are major trip generating centres, especially in the Brazilian context where the sport has great repercussion. Based on this, this article aimed to estimate trip generation models for Brazilian stadiums based on secondary data. The paying public was used as a proxy for the number of trips and the results indicated that the capacity of the stadiums and the average ticket value have a positive influence on the trip generation. As not all models estimated by the ordinary least squares method met the linearity assumptions, by identifying heteroscedasticity (errors are not constant), alternatively, models were estimated by robust regression. Despite the use of secondary data, the results contribute to the literature about trip generation and transport planning.

1. INTRODUÇÃO

A maneira como grandes empresas e empreendimentos geram viagens (atraem e produzem) tem sido fundamento de estudos por diversos pesquisadores, centros de pesquisas internacionais e agências governamentais, desde a segunda metade do século XX (Hooper, 2017). Para exemplificar, o *Institute of Transportation Engineering* (ITE), dos Estados Unidos, iniciou, em 1970, um programa para coletar informação das viagens geradas pelos mais diversos usos de solo, incluindo grandes polos geradores de viagens (PGVs), que possuem efeito significativo no tráfego das cidades (Portugal, 2012). Desde então, diversos modelos têm sido estimados para melhor entender a geração de viagem desses PGVs e, assim, mensurar seus impactos, com o objetivo de ajudar no planejamento do sistema de transportes. No Brasil, esses PGVs são compreendidos como locais que oferecem produtos ou serviços, que atraem grande número de viagens - majoritariamente realizadas por automóveis - e impactam o sistema rodoviário nos quesitos de acessibilidade e segurança viária (CET-SP, 1983; Goldner, 1986; DENATRAN, 2001; Portugal e Goldner, 2003; Oliveira *et al.*, 2017a).

Em sua obra, Portugal (2012) elencou as atividades mais estudadas como PGVs e, mesmo após 10 anos de sua publicação, os conceitos apresentados pelo autor ainda são importantes para estimar a geração de viagens em várias cidades brasileiras. Para Portugal (2012), os



estabelecimentos considerados PGVs podem ser residenciais (unifamiliares e multifamiliares), comerciais (*shopping centers*, supermercados, bares e restaurantes) e de serviços (hotéis, escritórios, hospitais, universidades, escolas, academias, centros esportivos, estádios, cinemas, teatros, estacionamentos, bancos, postos de gasolina, terminais rodoviários, aeroportos, áreas industriais e armazéns).

As principais externalidades diretas que a instalação destes estabelecimentos traz para seu entorno são associadas aos aumentos (i) na frota de veículos; (ii) no tempo de viagem; (iii) no congestionamento; (iv) no número de acidentes; e (v) na demanda por estacionamentos. As externalidades indiretas, por sua vez, estão relacionadas com a mudanças no uso do solo e na forma urbana do seu entorno, com o potencial para (i) criação de novas centralidades; (ii) aumento da renda média local; (iii) agravar problemas ambientais, como aumento das emissões e dos ruídos; e (iv) danos ao patrimônio histórico (Portugal, 2012).

Os PGVs do tipo residencial foram estudados no contexto brasileiro. Cervero (2008) e Marcolini (2011) concluíram que a taxa de geração de viagens por cada unidade residencial é menor que a estimada pelo manual do ITE para o contexto estadunidense. Fatores como densidade habitacional, tipologia construtiva, desenho urbano do entorno das áreas residenciais e acessibilidade ao sistema de transporte do entorno tem grande influência na taxa de geração de viagem (Marcolini, 2011). Torquato e Raia Junior (2014) identificaram que o número de unidades habitacionais ocupadas apresenta modelos mais satisfatórios do que quando analisada a área das residências. Os autores apontaram, ainda, indícios que condomínios horizontais possam gerar mais viagens que os verticais.

Oliveira *et al.* (2017b), em uma análise para diversos segmentos de PGVs da cidade de Belo Horizonte, concluíram que para condomínios residenciais, os resultados são inferiores quando confrontados com os obtidos na literatura, ou seja, as variáveis utilizadas “número de quartos”, “área construída” e “unidades residenciais” revelam menor geração de viagem do que o outrora estudado. Oliveira *et al.* (2020a) analisaram edifícios em construção e concluíram que os modelos utilizando o “número de funcionários” tem mais capacidade explicativa da geração de viagens de carga do que o “número de unidades” construídas ou a “quantidade de andares” do edifício.

Em relação aos PGVs do tipo serviço, os principais trabalhos disponíveis avaliaram o impacto que bares, restaurantes e supermercados possuem no tráfego local e as mudanças que a sua implantação traz para a microeconomia local. CET-SP (1983), Macedo *et al.* (2002), Nishimori e Raia Junior (2011), Gontijo (2014) e Oliveira *et al.* (2017b) estimaram modelos de geração de viagens para hospitais enquanto Nunes (2005) e Oliveira *et al.* (2017b) analisaram os modelos de geração de viagem para escolas e centros educacionais. Goldner (1986), CET-SP (2000), DENATRAN (2001), Jacobsen *et al.* (2010), Portugal (2012) e Oliveira *et al.* (2017b) avaliaram os modelos de geração de viagem para shopping centers enquanto Oliveira *et al.* (2020b) estimaram modelos de geração de viagem de carga para armazéns.

As análises de PGVs do tipo serviço são múltiplas e englobam variáveis diversas e distintas de acordo com a natureza de cada um dos estabelecimentos e o serviço ofertado. Em síntese, os modelos desses PGVs revelam que estabelecimentos com maiores capacidades de operação,



analisadas em termos de “áreas”, “vagas” e “número de funcionários/alunos/leitos/lojas”, tendem a produzir mais viagens, variando também quando a periodicidade e objetivo dos deslocamentos.

Ainda se tratando dos PGVs do tipo serviço, Portugal (2012) indica que os estádios de futebol possuem grandes dimensões e grande potencial de atração de viagens. Apesar disso, a literatura não apresenta trabalhos que os avaliem como PGVs, até onde os autores deste artigo tem conhecimento. Mesmo no Brasil, país que presenciou um aumento no número de estádios com a Copa do Mundo de 2014, não existem trabalhos acadêmicos específicos sobre os efeitos de suas construções e operações, nas gerações de viagem. Considerando, ainda, que o futebol é o esporte mais popular no Brasil, e atrai grandes públicos (Almeida *et al.*, 2019), a avaliação dessas arenas esportivas torna-se mais importante e relevante.

No contexto desse artigo, são consideradas as partidas de futebol ocorridas em 2019, em um cenário anterior à pandemia da COVID-19. Naquele ano, ocorreram 2.293 jogos de 19 diferentes campeonatos, de cunho local/regional, nacional e internacional. Estas partidas atraíram cerca de 19 milhões de pessoas, obtendo estádios com quase 100% de ocupação e somaram quase R\$ 600 milhões de renda bruta às 60 equipes ranqueadas (Lemos *et al.*, 2019). Além de partidas de futebol, muitos destes estádios e arenas são cartões postais de suas cidades e recebem concertos, exibições artísticas, festivais e outros eventos culturais, aumentando o seu potencial como geradores viagens.

Percebe-se, assim, que os estádios de futebol brasileiros possuem grande potencial de atração de viagens e, desse modo, é importante conhecer suas taxas de geração de viagens para estimar o impacto delas no entorno dos empreendimentos, além de permitir analisar as condições de acesso. Ainda, as taxas de geração de viagem fornecem subsídios para traçar estratégias de planejamento do sistema viário, de dimensionamento de transporte público e de projetos de circulação viária, tanto para dias de eventos quando para dias fora de eventos.

Assim, neste trabalho, são estimados modelos de geração de viagens para estádios de futebol brasileiros, baseado em dados secundários. A pergunta norteadora desta pesquisa é: “é possível estimar modelos de geração de viagens para estádios brasileiros baseados em dados secundários?”. Três são as hipóteses de pesquisa: a primeira (H1) é que o valor do *ticket* médio influencia negativamente na geração de viagens, e a segunda (H2) é que o tamanho do estádio (capacidade) influencia positivamente na geração de viagens. Por fim, a terceira hipótese (H3) é que é possível estimar modelos de geração de viagens para estádios brasileiros baseado em dados secundários.

Este trabalho está dividido em cinco seções. Após esta primeira seção introdutória, que também oferece panorama geral do contexto e estudos sobre os PGVs, a seção 2 conta com a apresentação dos dados utilizados nesta pesquisa. Na seção 3 está apresentado o método de pesquisa, cujos resultados são indicados na seção 4. Por fim, a seção 5 contém as principais conclusões deste trabalho.



2. DESCRIÇÃO DOS DADOS

Para realização deste estudo, foram obtidos dados do valor do *ticket* médio dos ingressos e do público pagante de cada jogo de futebol ocorrido no Brasil no ano de 2019, cenário pré-pandemia da COVID-19 (Lemos *et al.*, 2019). Devido ao grande número de partidas com uma grande variação de suas características, foi feita triagem dos dados, adotando-se para as análises os 13 principais campeonatos, sendo 9 locais/regionais, 2 nacionais e 2 internacionais. Essa seleção resultou em 1.260 partidas (55% dos jogos totais), com audiência superior a 15 milhões de pessoas (cerca de 80% do público pagante total de 2019). O resumo dos dados encontra-se na Tabela 1.

Tabela 1: Dados dos campeonatos avaliados.

Campeonato	Tipo	Partidas	Audiência
Libertadores	Internacional	36	1.386.815
Sul-Americana	Internacional	18	447.000
Brasileirão Série A	Nacional	380	8.070.188
Copa do Brasil	Nacional	119	1.443.827
Paulista	Local/Regional	121	1.152.658
Carioca	Local/Regional	97	777.223
Mineiro	Local/Regional	76	410.353
Gaúcho	Local/Regional	80	420.931
Baiano	Local/Regional	51	221.997
Cearense	Local/Regional	62	233.968
Paranaense	Local/Regional	74	232.409
Pernambucano	Local/Regional	53	163.689
Catarinense	Local/Regional	93	269.541
Total		1.260	15.230.599

No que diz respeito aos estádios de futebol, foram obtidos dados da Confederação Brasileira de Futebol (CBF) sobre o número de arenas distribuídas no Brasil. Os estados com maior concentração de estádios são São Paulo e Rio Grande do Sul, seguidos pelo Rio de Janeiro, Bahia, Santa Catarina, Minas Gerais, Pernambuco e Piauí (CBF, 2016). Grande parte destes estádios é de pequena capacidade, sendo que mais da metade deles suporta público inferior a 5 mil pessoas, e apenas 4% têm capacidade de mais de 30 mil pessoas (CBF, 2016).

Para as análises desse artigo, foram selecionados os 45 maiores estádios - em termos de capacidade. Esta escolha foi feita cruzada às informações das partidas (Lemos *et al.*, 2019), de forma a analisar (i) os estádios que mais receberam partidas em 2019, ou seja, aqueles que mais geraram viagens dos torcedores pagantes; e (ii) mais de 90% do público presente nos 1.260 jogos (aqueles determinados na Tabela 1). A Tabela 2 resume a amostra analisada (768 jogos após triagem de dados e seleção dos estádios), enquanto a Figura 1 ilustra a localização dos 45 estádios em um panorama geral da concentração de arenas por estado brasileiro.

Tabela 2: Resumo das partidas analisadas.

Tipo do Campeonato	Total da Amostra		Amostra Analisada [45 estádios]			
	Partidas	Audiência	Partidas	%	Audiência	%
Internacional	54	1.833.815	54	100	1.833.815	100
Nacional	499	9.514.015	441	88	9.052.204	95
Local/Regional	707	3.882.769	273	39	3.088.604	80
Total	1.260	15.230.599	768	61	13.974.623	92

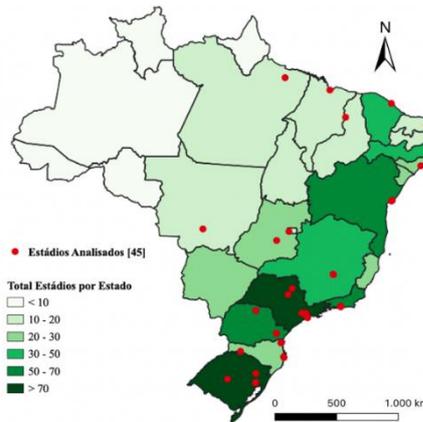


Figura 1: Distribuição dos estádios nos estados brasileiros e estádios analisados.

As Figuras 2 e 3 apresentam a estatística descritiva dos dados usados na modelagem e a Tabela 3 descreve-os, em termos de valor médio. Nota-se que, em relação ao público pagante, os campeonatos internacionais atraem mais pessoas, enquanto os campeonatos nacionais apresentam maior dispersão dos dados. No que diz respeito à capacidade dos estádios, observa-se que os campeonatos locais/regionais acontecem em arenas de menor capacidade. Por fim, verifica-se que os campeonatos internacionais têm o maior ticket médio.

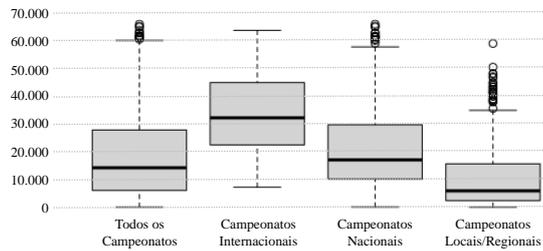


Figura 2: Boxplot do público pagante.

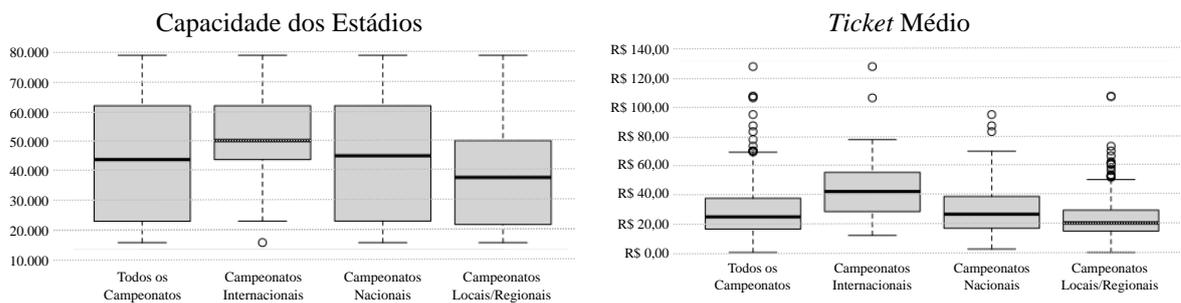


Figura 3: Boxplot da capacidade dos estádios e ticket médio.

Tabela 3: Resumo das variáveis da amostra analisada.

Tipo do Campeonato	Partidas	Valores Médios		
		Público Pagante	Capacidade	Ticket Médio
Internacionais	54	33.960	52.923	R\$ 44,50
Nacionais	441	20.527	44.986	R\$ 29,94
Locais/Regionais	273	11.314	38.951	R\$ 24,18
Todos	768	18.196	43.399	R\$ 28,92



Os dados de capacidade e *ticket* médio foram selecionados por entender que podem influenciar diretamente na geração de viagens (medida pelo público pagante). Ainda, em testes prévios de correlação de Pearson, foi identificada correlação fraca ou moderada entre estas variáveis, propiciando a modelagem. No entanto, outras variáveis referentes a infraestrutura do estádio e comportamento das partidas estão disponíveis em CBF (2016) e Lemos *et al.* (2019).

3. MÉTODO DE PESQUISA

O método de pesquisa descrito nesta seção seguiu etapas para estimar e testar a acurácia dos modelos estimados, com o objetivo de minimizar os erros de previsão e prover estimativas mais confiáveis (Oliveira *et al.*, 2022). A análise foi feita considerando os dados descritos na Seção 2. A variável dependente (y) dos modelos foi o número de pagantes nas partidas de futebol, como *proxy* das viagens atraídas em cada partida. As variáveis dependentes foram a capacidade dos estádios (x_1) e o valor do *ticket* médio pago pelo público (x_2).

Inicialmente, foi verificado o erro de especificação dos dados pelo teste Ramsey *Regression Equation Specification Error* (teste *RESET*). O teste *RESET* é um teste de especificação geral para o modelo de regressão linear que testa as combinações não lineares dos valores para aferir a explicação da variável de resposta. Valores sem significância estatística indicam que o modelo de regressão linear está mal especificado e poderia ser aproximado por outra função (Elian *et al.*, 2020).

Identificado a correta especificação do modelo, verificou-se a existência de *outliers* utilizando a técnica de distância de Cook. Essa técnica mede a influência de uma observação na amostra, a partir do resíduo de saída, combinando com a alavancagem para estimar o nível de influência, isto é, combinado com a relação entre a extremidade da variável em relação ao seu valor médio. Com a remoção dos *outliers*, o teste *RESET* foi novamente aplicado para confirmar a especificação do modelo.

Com um conjunto de dados sem *outliers* e corretamente especificado, a normalidade foi testada usando os testes de normalidade Kolmogorov-Smirnov e Shapiro-Wilk. Estes testes identificam se a variável dependente (y) apresenta distribuição aproximada a uma distribuição normal, necessário para a aplicação de uma regressão linear. Por fim a correlação de Pearson foi calculada para estimar o nível de associação entre as variáveis, sendo que valores acima de 0,70 indicam que é esperado que as variáveis tenham relação entre si.

Confirmada a normalidade, estimou-se os modelos de regressão linear (OLS) pelo método dos mínimos quadrados (MMQ). MMQ é uma técnica usual usada para estimar modelos de geração de viagem, que minimiza a diferença entre os valores observados e previstos (soma dos erros quadrados). A forma de uma função OLS considera uma variável dependente (y), variáveis explicativas (x), estimadores (β) e erro (e), para um número n de variáveis, conforme equação (1).

$$Y_n = \beta_1 + \beta_2 X_1 + \beta_2 X_2 + (\dots) + \beta_n X_n + e_n \quad (1)$$

Os modelos foram avaliados pelo teste-t, teste f e pelo coeficiente de determinação (R^2). Ainda, o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC) foram



utilizados para comparar e selecionar o modelo mais adequado para explicar o fenômeno estudado, sendo que menores valores indicam melhores modelos. Por fim, foram verificados os principais pressupostos do MMQ, por meio das seguintes técnicas:

- Linearidade dos parâmetros: teste *RESET*
- Média dos erros: gráfico dos resíduos *versus* ajustes
- Homoscedasticidade: teste de Breusch-Pagan
- Não autocorrelação dos erros: teste de Durbin-Watson
- Erros normalmente distribuídos: teste de Shapiro-Wilk
- Multicolinearidade: fator de inflação da variância (VIF)

Um usual problema na estimação por MMQ é a heterocedasticidade (ausência de homoscedasticidade) que ocorre quando os erros não são constantes. Uma alternativa a esta violação é estimar modelos pela regressão robusta. A regressão robusta, proposta por Tabatabai *et al.* (2012), é uma técnica que estimada modelos reduzindo a influência dos pontos que reduzem a qualidades dos parâmetros utilizados no modelo. Os modelos estimados pela regressão robusta devem produzir estimativas tão boas quanto outras técnicas de regressão (por exemplo, pelo MMQ) à anuência das variáveis de resposta. A regressão robusta considera a mesma equação (1) da regressão linear, sendo os estimadores (β) denominado (β') e estimados por interação de mínimos quadráticos ponderados, denominados “M-Estimadores”.

Por fim, para os modelos que se mostram mais adequados, em termos de significância e critérios de informação, foram avaliados os poderes de predição por meio de validação cruzada (k-fold), sendo que apenas os modelos validados pela validação cruzada podem ser considerados preditivos e podem ser utilizados para estimativa da geração de viagens. Caso isto não ocorra, os modelos são apenas explicativos do fenômeno.

Todas as análises desse artigo foram realizadas no ambiente R (R-Core Team, 2013) usando os pacotes lmtest (Zeileis e Hothorn, 2002), stats (R-Core Team, 2013), gvlma (Peña e Slate, 2006), MASS (Venables e Ripley, 2002) e caret (Kuhn *et al.*, 2022).

4. RESULTADOS E DISCUSSÕES

Após a remoção dos *outliers* pela distância de Cook, foram considerados dados de 740 jogos em diferentes campeonatos. Na Tabela 4 são apresentados os resultados do teste *RESET*, já considerando a exclusão dos *outliers*, que permite identificar a transformação de dados que melhor ajusta os modelos. Observa-se que a transformação logarítmica é a que melhor se adequa a todos os conjuntos de dados, retratado pelos diferentes tipos de campeonatos. Esta transformação foi usada para a estimativa dos modelos de geração de viagens.

Tabela 4: Resultados do *RESET test*

Tipo de Transformação da Variável (y)	Campeonatos			
	Todos	Internacionais	Nacionais	Locais/Regionais
linear	9,714 ***	2,442 *	13,063 ***	4,268 ^{SS}
logarítmica	27,237 ***	3,292 *	6,333 **	6,644 **
quadrática	2,354 *	2,660 *	1,556 ^{SS}	1,336 ^{SS}

Significância: ***0.001 **0.01 *0.1 ^{SS}Sem Significância (>0,1)



Na Tabela 5 estão apresentados os resultados do teste da normalidade, para verificar se a regressão linear é adequada para os dados. Os resultados indicam que a amostra apresenta comportamento normal com significância estatística de 90% para todos os cenários no teste de Shapiro-Wilk, o que não ocorre para os campeonatos internacionais e locais/regionais no teste de Kolmogorov-Smirnov. Como um todo, o pressuposto de normalidade indica que a regressão linear pode ser utilizada para estimar modelos de geração de viagens dos estádios de futebol. No caso dos campeonatos internacionais, os resultados do modelo precisam ser cuidadosamente verificados pelo não atendimento do teste de Kolmogorov-Smirnov. Os testes de correlação de Pearson mostram que as variáveis possuem correlações fracas e/ou moderadas entre si.

Tabela 5: Resultados dos testes de normalidade.

Campeonato	Teste (p-valor)	
	Kolmogorov-Smirnov	Shapiro-Wilk
Todos	0,000 ***	0,000 ***
Internacionais	0,607 ^{SS}	0,015 *
Nacionais	0,027 *	0,000 ***
Locais/Regionais	0,554 ^{SS}	0,000 ***

Nível de significância do teste: ***0.001 **0.01 *0.1 ^{SS}Sem Significância (> 0,1)

Na Tabela 6 estão apresentados os resultados dos modelos estimados pelo MMQ. Observa-se que todos os parâmetros são estatisticamente significativos (p-valor < 0,001) e os valores ajustados de R² são estão na média de 46,67%.

Tabela 6: Resultados dos modelos OLS.

Modelo	Variáveis	Parâmetros	R ² Ajustado	AIC	BIC
Modelo 1	Intercepto	7,27319 ***			
Todos os campeonatos	Capacidade	0,00003 ***	45,68%	1.982,202	2.000,777
	<i>Ticket</i> médio	0,02750 ***			
Modelo 2	Intercepto	9,20518 ***			
Campeonatos internacionais	Capacidade	0,00001 ***	50,73%	45,616	53,571
	<i>Ticket</i> médio	0,01224 ***			
Modelo 3	Intercepto	8,01649 ***			
Campeonatos nacionais	Capacidade	0,00002 ***	45,11%	857,188	873,544
	<i>Ticket</i> médio	0,01726 ***			
Modelo 4	Intercepto	6,51969 ***			
Campeonatos locais/regionais	Capacidade	0,00003 ***	45,14%	785,167	799,604
	<i>Ticket</i> médio	0,03948 ***			

Nível de significância do teste-t: *** p-valor < 0.001

Os resultados dos testes para validar os pressupostos do MMQ estão apresentados na Tabela 7. O Modelo 1, que considera todos os jogos, e o Modelo 3, que considera os campeonatos nacionais, têm todos os pressupostos do MMQ atendidos e, desta forma, os resultados estimados por MMQ podem ser utilizados. Como o valor do AIC e do BIC foram menores para o Modelo 3 (quando comparado com o Modelo 1), sugere-se utilizar este modelo para o caso de considerar apenas jogos do campeonato nacional.



Tabela 7: Verificação dos pressupostos do MMQ.

Pressupostos	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Linearidade dos parâmetros	Atende	Atende	Atende	Atende
Média dos erros	Atende	Atende	Atende	Atende
Homoscedasticidade	Atende p-valor = 0,000	Não atende p-valor = 0,170	Atende p-valor = 0,000	Não atende p-valor = 0,240
Multicolinearidade	Atende VIF = 1,031	Atende VIF = 1,104	Atende VIF = 1,000	Atende VIF = 1,075
Autocorrelação dos erros	Atende p-valor = 0,000	Atende p-valor = 0,000	Atende p-valor = 0,000	Atende p-valor = 0,000
Erros normalmente distribuídos	Atende p-valor = 0,000	Atende p-valor = 0,000	Atende p-valor = 0,000	Atende p-valor = 0,000

Para os Modelos 2 e 4 foi observada a presença de heterocedasticidade. Desta forma, conforme indicado na descrição do método, técnicas alternativas ao MMQ são necessárias. Assim, estimou-se a regressão robusta, cujos resultados estão apresentados na Tabela 8. Observa-se que os coeficientes estimados são similares ao da regressão pelo MMQ e todos tem validade estatística. Desta forma, e em atendimento à literatura, sugere-se a utilização dos modelos estimados para a regressão robusta para dados provenientes de campeonatos internacionais e locais/regionais para melhor estimativa dos modelos.

Tabela 8: Resultados dos modelos estimados por regressão robusta

Modelo	Variáveis	Parâmetros	AIC	BIC
Modelo 2'	Intercepto	9,3309 ***		
Campeonatos internacionais	Capacidade	0,0000 ***	49,611	57,567
	<i>Ticket</i> médio	0,0092 ***		
Modelo 4'	Intercepto	6,7108 ***		
Campeonatos locais/regionais	Capacidade	0,000 ***	790,407	804,844
	<i>Ticket</i> médio	0,0374 ***		

Nível de significância do teste-t: *** p-valor < 0.001

Na Tabela 9 estão ilustrados os resultados da validação cruzada. Observa-se que apesar do baixo poder de predição de alguns modelos (R^2) os valores do Erro Absoluto Médio (MAE) e da Raiz do Erro Quadrático Médio (RMSE) são considerados baixos, indicando poucos erros na predição. Desta forma, os modelos estimados e apresentados neste estudo podem ser utilizados para predição do público pagante e, conseqüentemente, do número de viagens geradas para os estádios de futebol.

Tabela 9: Resultados da validação cruzada.

Modelo	Método da Estimativa	R^2	MAE	RMSE
Modelo 1	Regressão Linear	0,46156	0,64712	0,87275
Modelo 2'	Regressão Robusta	0,63918	0,26347	0,33952
Modelo 3	Regressão Linear	0,46034	0,45704	0,63026
Modelo 4'	Regressão Robusta	0,47852	0,75949	1,01691

Os resultados apresentados neste artigo permitem comprovar as hipóteses de estudo estabelecidas. Em todos os foi verificado a influência positiva do valor do *ticket* médio na



geração de viagens, contrariando a hipótese H1 que esperava que mais viagens fossem atraídas quando o valor desembolsado para assistir à partida fosse menor.

A hipótese H2 foi comprovada, indicando que estádios com maior capacidade atraem mais público pagante, ou seja, mais viagens. Este resultado parece ser óbvio, no entanto é de fundamental importância conforme discutido abaixo. Por fim, a hipótese H3 foi confirmada comprovando que é possível estimar modelos de geração de viagens para estádios brasileiros por meio de dados secundários.

No que tange aos resultados, os modelos estimados por MMQ apresentaram altos valores de intercepto. Esse resultado indica que outras variáveis latentes não foram usadas na modelagem, mas influenciam na geração de viagem aos estádios de futebol. Estas variáveis podem ser: a facilidade de acesso aos estádios por transporte público, promoções de ingressos, condições de estacionamento para veículos privados, dia da semana e horário das partidas, dentre outras.

Por fim, usar dados secundários para a estimativa de geração de viagens é uma alternativa interessante quando não se pode fazer coleta de dados. Este estudo foi desenvolvido durante a pandemia da COVID-19, quando não era permitido a presença de público nos estádios e/ou não ocorriam jogos, motivo pelo qual foram usados os dados de 2019. Desta forma, a maneira usual para coleta de dados, sugerida em Portugal (2012), não pode ser utilizada neste estudo, desafiando os pesquisadores a encontrar alternativas a coleta de caso usual.

No que tange estádios de futebol, a CBF tem detalhado as informações sobre cada jogo ocorrido no país. Estas informações podem ser utilizadas pelos municípios para identificar os frequentadores dos estádios e permitir identificar melhores estratégias para, principalmente, reduzir o uso do veículo privado nestes eventos. Apesar da maioria dos estádios brasileiros ter uma elevada capacidade de estacionamento, deve-se sempre sugerir o uso do transporte público, garantindo viagens por modos mais sustentáveis. Por fim, para os estádios que não tem acesso por transporte público, os dados deste estudo permitem traçar estratégias para estimular o uso deste modo nos eventos promovidos nos estágios de futebol.

5. CONCLUSÃO

Sendo o futebol um esporte tão presente na cultura brasileira, os estádios nacionais são locais centrais que atraem público para as partidas esportivas e outros eventos culturais. Apesar disso, não se identifica na literatura modelos de geração de viagens para essas arenas. Neste sentido, este artigo apresenta resultados que vêm contribuir com a literatura sobre Polos Geradores de Viagem e com o planejamento de transportes dos eventos realizados nestes empreendimentos.

Seis modelos foram estimados considerando dados secundários das partidas realizadas no ano de 2019 em 45 dos maiores estádios brasileiros: público pagante (variável dependente), a capacidade dos estádios e o valor do *ticket* médio (variáveis explicativas). Os modelos para todo o conjunto de dados (englobando campeonatos internacionais, nacionais, locais/regionais) e o modelo para campeonatos nacionais tiveram seus pressupostos validados quando estimados por MMQ. Para o conjunto de dados de campeonatos internacionais e locais/regionais, que não tiveram pressupostos validados por MMQ, foram estimados modelos por regressão robusta.



Os resultados obtidos permitem responder positivamente à pergunta de pesquisa “é possível estimar modelos de geração de viagens para estádios brasileiros usando dados secundários?”. Ainda, foi refutada a hipótese H1, em que o *ticket* médio tem influência negativa na geração de viagens e comprovada a hipótese H2, em que a capacidade do estádio tem influência positiva.

Apesar da validação das hipóteses, os altos valores dos interceptos dos modelos estimados indicam que outras variáveis influenciam no processo e podem ser incorporadas no modelo. Essa é uma das sugestões para trabalhos futuros, incluindo esses fatores na estimativa de modelos de geração de viagens. Ainda, sugere-se pesquisas com os frequentadores de estádios, visto que os dados obtidos podem auxiliar no melhor entendimento da geração de viagens ao trazer a perspectiva do usuário para os modelos de geração de viagem.

Agradecimentos

Os autores agradecem o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo incentivo a esta pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

- Almeida, M. I. S.; R. L. F. Coelho; D. S. Oliveira; A. Camargo e P. Savioli (2019). Sales-Based Brand Equity as a Performance Driver in ‘The Country of Soccer’. *Revista de Administração Contemporânea*, v. 24, p.134–50. <https://doi.org/10.1590/1982-7849rac2020180284>
- CBF (2016). Cadastro Nacional de Estádios do Brasil. Disponível em: <https://pt.slideshare.net/cassiozipa/cadastro-nacional-de-estádios-do-brasil-2016>. (acessado em 20 de junho de 2022).
- Cervero, R. e G. B. Arrington (2008). Vehicle Trip Reduction Impacts of Transit-Oriented Housing. *Journal of Public Transportation*, v.11, n. 3, p.1-17. <http://doi.org/10.5038/2375-0901.11.3.1>
- CET-SP (1983). Polos Geradores de Tráfego (Boletim Técnico nº 32). Boletim Técnico 32. São Paulo: Companhia de Engenharia de Tráfego CET-SP. Disponível em: <http://www.cetsp.com.br/media/65486/bt32-%20polos%20geradores%20de%20trafego.pdf>. (acessado em 20 de junho de 2022).
- CET-SP (2000). Polos Geradores de Tráfego (Boletim Técnico no 36). Boletim Técnico 36. São Paulo: Companhia de Engenharia de Tráfego CET-SP. Disponível em: <http://www.cetsp.com.br/media/66593/bt36-%20polos%20geradores%20de%20trafego%20ii.pdf>. (acessado em 20 de junho de 2022).
- DENATRAN (2001). Manual de Procedimentos para o Tratamento de Polos Geradores de Tráfego. Departamento Nacional de Trânsito". Brasília: Departamento Nacional de Trânsito - Ministério da Justiça, Disponível em: <http://www.peasistemas.com.br/freeaspupload/suma/DENATRAN%20Polos%20Geradores.pdf>. (acessado em 20 de junho de 2022).
- Elian, M.; N. Sawalha e A. Bani-Mustafa (2020). Revisiting the FDI-Growth Nexus: ARDL Bound Test for BRICS Standalone Economies. *Modern Applied Science*, v.14, n. 6, p.1-20. <https://doi.org/10.5539/mas.v14n6p1>
- Goldner, L.G. (1986). A interferência dos polos geradores de tráfego no sistema viário: análise e contribuição metodológica para Shopping Centers. Dissertação de Mestrado, Rio de Janeiro: Universidade Federal do Rio de Janeiro.
- Gontijo, G. A. S. (2014). Modelos e taxas de atração de viagens para PGVs—Hospitais públicos localizados em cidades de médio porte do interior do estado de São Paulo. Tese de Doutorado, São Carlos: Universidade Federal de São Paulo. p.159. Disponível em: <https://repositorio.ufscar.br/bitstream/handle/ufscar/4191/5765.pdf?sequence=1&isAllowed=y>. (acessado em 20 de junho de 2022).
- Hooper, K. G. (2017). Trip Generation Handbook: An ITE Recommended Practice. ITE: Washington D.C.
- Jacobsen, A. C.; H. B. B. Cybis; L. A. Lindau e A. B. Pinto (2010). Modelos de geração e variabilidade no volume diário de veículos em shopping centers. *Transportes*, v. 18, n. 1, p. 105-113. <https://doi.org/10.14295/transportes.v18i1.388>
- Kuhn, M.; J. Wing; S. Weston; A. Williams; C. Keefer; A. Engelhardt; T. Cooper; Z. Mayer; B. Kenkel; M. Benesty; R. Lesarbeau; A. Ziem; L. Scrucca; Y. Tang; C. Candan e T. Hunt (2022). caret: Classification and Regression Training (versão 6.0-92). Disponível em: <https://CRAN.R-project.org/package=caret>. (acessado em 20 de junho de 2022).
- Lemos, C.; R. Breves, R. e L. Silva (2019). Público nos estádios do Brasil em 2019 | globoesporte.com. 2019.



- Disponível em: <http://app.globoesporte.globo.com/futebol/publico-no-brasil/2019/index.html> (acessado em 20 de junho de 2022).
- Macedo, M. H.; I. M. Filizola e E.A. Souza (2002). Polos geradores de tráfego: Estudo de agrupamento de clínicas médicas. Em Anais do XVI Congresso de Pesquisa e Ensino em Transportes. Natal.
- Marcolini, S. (2011) Ambiente urbano e geração de viagens: Niterói, um estudo de caso. Dissertação de Mestrado, Rio de Janeiro. Universidade Federal do Rio de Janeiro. p.194.
- Nishimori, F. T. I e A.A. Raia Junior (2011). Modelos de geração de viagens a prontos-socorros públicos. Em Anais do XXV Congresso de Pesquisa e Ensino em Transportes. Belo Horizonte. <http://redpgv.coppe.ufrj.br/index.php/es/produccion/articulos-cientificos/2011-1/587-modelos-de-geracao-de-viagens-a-prontos-socorros-publicos/file>.
- Nunes, J. L. (2005). Estudo da demanda por estacionamento em instituições de ensino superior. Dissertação de Mestrado, Brasília: Universidade de Brasília. p.173. Disponível em: <https://docplayer.com.br/57222228-Estudo-da-demanda-por-estacionamento-em-instituicoes-de-ensino-superior.html>. (acessado em 20 de junho de 2022).
- Oliveira, L. K.; R. A. A. Nóbrega; D. G. Ebias e B.S.G. Corrêa (2017a). Analysis of Freight Trip Generation Model for Food and Beverage in Belo Horizonte (Brazil). *Region*, v. 4, n. 1, 17–30. <https://doi.org/10.18335/region.v4i1.102>
- Oliveira, L. K.; L. C. M. Stubbs; N. T. Gontijo e R.L.M. Oliveira (2017b). Proposição de modelos de geração de viagens para Belo Horizonte. *Transportes*, v. 25, n. 2, p. 137–55. <https://doi.org/10.14295/transportes.v25i2.1243>
- Oliveira, L. K.; R. T. Herédia; B. V. Bertoncini e R. L. M. Oliveira (2020a). Geração de viagens de carga para edifícios em construção: uma análise comparativa entre regressão linear e regressão linear generalizada. *Transportes*, v. 28, n. 5, p. 28–42. <https://doi.org/10.14295/transportes.v28i5.1885>
- Oliveira, L. K.; G. P. Lopes; R. L. M. Oliveira e L.S.F.P. Bracarense (2020b). Locational Context for Warehouse Facilities in Urban Areas: A Case Study in Belo Horizonte (Brazil). *Transportation Research Procedia*, v. 48, p. 401–15. <https://doi.org/10.1016/j.trpro.2020.08.048>.
- Oliveira, L. K.; G. G. F. Araújo; B. V. Bertoncini; C. D. Pedrosa e F. G. F. Silva (2022). Modelling Freight Trip Generation Based on Deliveries for Brazilian Municipalities. *Sustainability*, v. 14, p. 10300. <https://doi.org/10.3390/su141610300>.
- Peña, E. A. e E. H. Slate (2006). Global Validation of Linear Model Assumptions. *Journal of the American Statistical Association*, v. 101, n. 473, p.341-354. <https://doi.org/10.1198/016214505000000637>
- Portugal, L. S. e L.G. Goldner (2003). Estudo de Pólos Geradores de Tráfego e de seus Impactos nos Sistemas Viários e de Transportes. Editora Blucher: Rio de Janeiro.
- Portugal, L. (2012). Polos geradores de viagens orientados a qualidade de vida e ambiental: Modelos e taxas de geração de viagens. Editora Interciência: Rio de Janeiro.
- R-Core Team. (2013). Diagnostic Checking in Regression Relationships. <https://www.r-project.org/>. (acessado em 20 de junho de 2022).
- Tabatabai, M. A.; W. M. Eby; H. Li; S. Bae e K. P. Singh (2012) TELBS robust linear regression method. *Open Access Medical Statistics*, p. 65-84. <https://doi.org/10.2147/OAMS.S37395>
- Torquato, T. L. L e A.A. Raia Júnior (2014). Modelos de geração de viagem para condomínios residenciais horizontais. *Transportes*, v. 22, n. 1, p. 56-64. <https://doi.org/10.14295/transportes.v22i1.708>
- Venables, B. e B. Ripley (2002). Statistical Analysis of Financial Data in S-Plus. Springer Texts in Statistics. Editora Springer: Nova Iorque.
- Zeileis, A. e T. Hothorn (2002). Diagnostic Checking in Regression Relationships, Disponível em: <https://cran.r-project.org/web/packages/lmtest/vignettes/lmtest-intro.pdf>. (acessado em 20 de junho de 2022).