



UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE ESTATÍSTICA

Vinicius Gabriel de Aquino

Métodos de Inferência Causal Aplicados ao Contexto de Mobilidade Urbana

**CURITIBA
2025**

Vinicius Gabriel de Aquino

Métodos de Inferência Causal Aplicados ao Contexto de Mobilidade Urbana

Trabalho de Conclusão de Curso apresentado à disciplina Laboratório de Estatística do Curso de Graduação em Estatística da Universidade Federal do Paraná, como exigência parcial para obtenção do grau de Bacharel em Estatística.

Orientador: Prof. Dr. Pedro Henrique Toledo de Oliveira Sousa

**CURITIBA
2025**

Aos meus irmãos e minhas sobrinhas

E a todos os sonhos que vocês ainda hão de ter!

AGRADECIMENTOS

Um agradecimento a minha mãe, Rosemary Mulhman, pela insistência e garra com que sempre apoiou esse sonho. Gostaria de pontuar, mãe, que sim, suas ex-patroas estavam certas: os alunos cotistas costumam reprovar muito e tendem a abandonar a formação, mas aqui estamos, mesmo que com 3 anos de atraso, 2 cursos abandonados e mais de 15 reprovações.

Um agradecimento, também, ao meu padrasto, José Aparecido, que me buscou em todos os meus ENEM's e foi uma referência diária de hombridade.

Agradeço também a minha companheira, Juliana, por todo apoio nos meus últimos semestres de graduação e por deixar a vida bem mais leve todos os dias.

Finalmente, um agradecimento especial aos professores do Departamento de Estatística da UFPR, especialmente ao meu orientador Pedro Henrique Toledo de Oliveira Sousa, por toda compreensão e paciência.

“Triunfo, se não for coletivo, é do sistema”.

Emicida

RESUMO

O presente estudo analisou o impacto causal da construção da Estação São Paulo-Morumbi no público do Estádio Cícero Pompeu de Toledo, nos jogos mandados pelo São Paulo Futebol Clube durante o Campeonato Paulista entre 2015 e 2024. Foi feita uma revisão literária dos conflitos sociais acerca da construção de uma Estação de Metrô em uma das regiões mais nobres de São Paulo, além de um levantamento prévio acerca da literatura dos públicos nos estádios, bem como uma revisão sobre os métodos estatísticos utilizados. O estudo se valeu do método diferenças nas diferenças e concluiu que a construção da Estação teve um efeito causal positivo no público do Morumbi, possibilitando que nos últimos 5 anos cerca de 350 mil torcedores pudessem ter acessado o estádio.

Palavras-chave: Diferenças nas Diferenças. Efeito Causal. Morumbi.

LISTA DE FIGURAS

Figura 1 -	Caminho do Estádio Morumbi as Estações de Metrô São Paulo-Morumbi e Butantã.....	9
Figura 2 -	Caminho do Estádio Neo-Química Arena até a Estação de Metrô Corinthians-Itaquera.....	10
Figura 3 -	Caminho do Estádio Allianz Parque até a Estação Palmeiras-Barra Funda.....	10
Figura 4 -	Método Diferenças nas Diferenças.....	17
Figura 5 -	Bilheteria Antes e Após a Construção da Estação São Paulo-Morumbi – Análise de Tendência.....	21
Figura 6 -	Bilheteria Antes e Após a Construção da Estação São Paulo-Morumbi.....	22
Figura 7 -	Bilheteria em Jogos em Dia de Semana e Fim de Semana.....	22
Figura 8 -	Bilheteria em Clássicos.....	23
Figura 9 -	Bilheteria em Jogos Eliminatórios.....	23
Figura 10 -	Resíduo dos modelos Diff-in-Diff com Covariáveis e Diff-in-Diff com Efeitos Fixos no Tempo.....	29
Figura 11 -	Resíduo dos modelos Diff-in-Diff com Covariáveis e Diff-in-Diff com Efeitos Fixos no Tempo ao longo do tempo.....	29
Figura 12 -	Resíduos x Peso com Distância de Cook.....	31

LISTA DE TABELAS

Tabela 1 - Saída do Modelo Diff-in-Diff.....	24
Tabela 2 - Saída do Modelo Diff-in-Diff com Covariáveis.....	25
Tabela 3 - Saída do Diff-in-Diff com Efeitos Fixos no Tempo.....	27
Tabela 4 - Efeitos Fixos Estimados do Diff-in-Diff com Efeitos Fixos no Tempo.....	27
Tabela 5 - Saída do Modelo Diff-in-Diff com Efeitos Fixos no Tempo com p-valores.....	28
Tabela 6 - Resultados do Teste Shapiro-Wilk.....	30

SUMÁRIO

1 INTRODUÇÃO	8
2 REVISÃO DE LITERATURA	11
2.1 Regressão Linear.....	11
2.2 Modelo de Rubin e Inferência Causal.....	13
2.3 Efeito de Tratamentos e Resultados Potenciais.....	14
2.4 Método Diferenças nas Diferenças.....	15
2.5 Demanda pelo Público nos Estádios de Futebol.....	18
2.6 Estação Morumbi e Conflitos Sociais.....	19
3 MATERIAL E MÉTODOS	20
3.1 Material.....	20
3.2 Método.....	21
4 RESULTADOS E DISCUSSÃO	27
5 CONCLUSÕES E COMENTÁRIOS FINAIS	33
REFERÊNCIAS	34

1 INTRODUÇÃO

A construção de novas infraestruturas de transporte público, como estações de metrô, é fundamental para o desenvolvimento sustentável das grandes cidades. Essas obras visam não apenas melhorar a mobilidade urbana, mas também promover a integração social e estimular o desenvolvimento econômico e a redução das desigualdades sociais.

Ostrensky, Porsse & Frota (2022) concluíram que a construção de estações de metrô podem estar também associadas ao processo de gentrificação, uma vez que valorizam o preço do m² da região, o que acaba por “expulsar” a população vulnerável da região. Todavia, quando se fala de regiões nobres, especialmente em São Paulo, essa conversa torna-se mais complicada, uma vez que há uma defesa por parte das elites na manutenção das características dos seus bairros (Mendes & Marino, 2012).

No meio deste conflito social, está localizado no bairro do Morumbi, o Estádio Cícero Pompeu de Toledo, conhecido como Morumbi, ou por direitos de nome, como Morumbis, casa do São Paulo Futebol Clube. O clube, apesar de, segundo a AtlasIntel (2024), ter a terceira maior torcida do país, até o ano de 2018 não tinha uma estação de metrô próximo ao seu estádio, dificultando o acesso do seu torcedor nos dias de jogos.

Antes da inauguração da estação São Paulo - Morumbi, a estação de metrô mais próxima do Estádio Morumbis era a Estação Butantã, situada a considerável distância de 4,2 km e construída em 2011. Essa longa distância dificultava o acesso dos torcedores do São Paulo Futebol Clube ao estádio, especialmente em jogos em dias de semana.

A construção da nova estação, a apenas 1,4 km do estádio, foi debatida entre moradores durante décadas e buscava não somente atender a essa demanda reprimida, mas principalmente facilitar o acesso de trabalhadores da região. A Figura 1 apresenta tanto a antiga rota, quanto a nova rota que os torcedores podem fazer.

Para avaliar o impacto da nova estação no público do Morumbi e na dinâmica urbana da região, este estudo utiliza o método diferenças nas diferenças, comparando a evolução do público do São Paulo Futebol Clube durante o Campeonato Paulista com a de outros dois grandes clubes de futebol de São Paulo: Palmeiras e Corinthians.

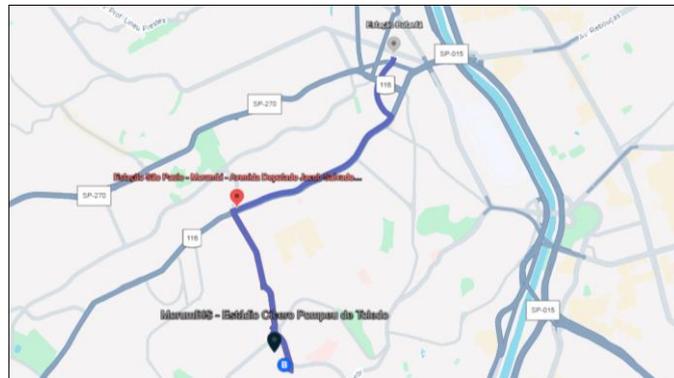
O método diferenças nas diferenças têm sido amplamente utilizado nas

últimas décadas na avaliação do impacto de políticas públicas. O método permite uma estimativa do efeito de intervenções em uma determinada variável, aqui o público nos estádios durante o paulistão, a partir da comparação entre um grupo que recebeu essa intervenção, aqui a torcida são paulina no estádio do Morumbi, com grupos controle que não receberam essa intervenção, aqui a torcida palmeirense e corinthiana nos seus respectivos estádios.

Os estádios destes clubes, Allianz Parque e Neo Química Arena, respectivamente, já possuíam estações de metrô próximas antes da construção da estação São Paulo-Morumbi, conforme apresentam as Figuras 2 e 3. A estação Palmeiras – Barra Funda e a estação Corinthians – Itaquera foram inauguradas ainda em 1988 e estão, respectivamente, a 1.3km e a 1.1km de distância dos respectivos estádios.

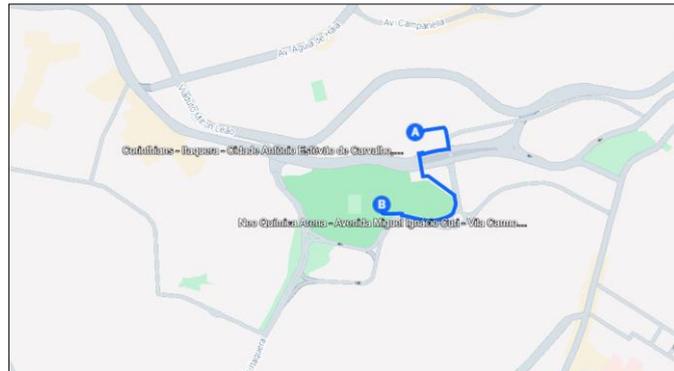
Essa característica torna os clubes Palmeiras e Corinthians grupos de controle ideais para a análise, pois permite isolar o efeito da melhoria na mobilidade proporcionada pela nova estação sobre o público do Estádio Morumbi.

Figura 1 - Caminho do Estádio Morumbi as estações de Metrô São Paulo-Morumbi e Butantã



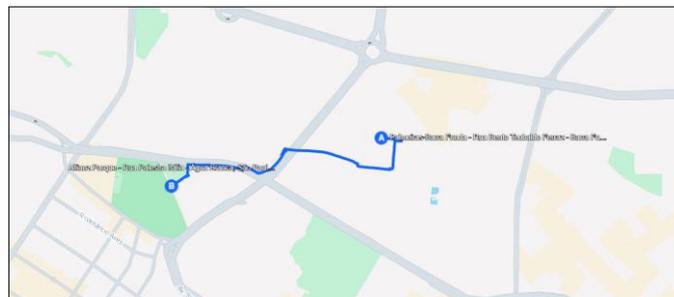
Fonte: Google Maps. 2024.

Figura 2 - Caminho do Estádio Neo-Química Arena até a Estação de Metrô Corinthians-Itaquera



Fonte: Google Maps. 2024.

Figura 3 - Caminho do Estádio Allianz Parque até a Estação Palmeiras-Barra Funda



Fonte: Google Maps. 2024.

2 REVISÃO DE LITERATURA

2.1 Regressão Linear

A Regressão Linear é um modelo estatístico onde modela-se uma variável dependente (ou de resposta) a partir de um conjunto de variáveis independentes X . O objetivo é entender como essas variáveis independente se relacionam com a uma dita parte determinística da variável dependente (CHARNET, 1999). O modelo da Regressão Linear é dado por

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon$$

Tal que:

- β_0 representa o parâmetro de intercepto, o valor esperado da variável y quando todas as variáveis dependentes valem 0;
- β_j representa o valor esperado incrementado em y quando incrementado uma unidade na covariável x_j ;
- $\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$ em conjunto representa uma quantidade de y “determinística”, que pode ser explicada pelo vetor de covariáveis X
- ϵ representa a parte estocástica de y e assume-se que ϵ tem distribuição normal, com média igual a 0 e variância igual a σ^2 .

O Modelo da Regressão Linear pode ser também expressado na forma matricial $Y = \beta X + \epsilon$, onde o β pode ser estimado via Mínimos Quadrados Ordinários por $X_{n \times p}$ e $y_{1 \times n}$ como

$$\hat{\beta} = (X)^{-1} X' y.$$

Sob as seguintes suposições (HILL, 2003):

- $\epsilon \sim N(0, \sigma^2)$
- $cov(\epsilon_i, \epsilon_j) = 0$
- $cov(x_{pi}, \epsilon_i) = 0$
- Os valores de x_{ik} não são aleatórios e nem funções lineares exatas das demais covariáveis

Pode-se garantir que $\hat{\beta}$ é o estimador não enviesado de menor variância existente.

O Modelo de Regressão Linear corriqueiramente recebe também o uso de variáveis dummies (ou binárias), que são covariáveis que recebem o valor 1 ou 0. Essas variáveis permitem modelar condições que possuem um caráter dicotômico, como ser brasileiro ou não ser, por exemplo.

Elas podem ser implementadas no modelo de regressão linear associadas ao intercepto, como

$$y = \beta_0 + \beta_1 D_1 + \beta_2 x_2 + \epsilon$$

Neste contexto, o valor esperado da variável y quando a covariável binária D_1 vale 1 é dado por

$$E(y) = E(\beta_0 + \beta_1(1) + \beta_2 x_2 + \epsilon) = (\beta_0 + \beta_1) + \beta_2 x_2$$

E, de forma similar, quando D_1 tem valor 0, então o valor esperado para y é dado por

$$E(y) = E(\beta_0 + \beta_1(0) + \beta_2 x_2 + \epsilon) = \beta_0 + \beta_2 x_2$$

E, dessa forma, a presença da covariável D_1 , ser brasileiro, por exemplo, gera um “prêmio” no intercepto de y .

Uma outra forma de se implementar *variáveis dummies* no Modelo de Regressão Linear é usando as como artifícios de mudanças de inclinação, como a equação

$$y = \beta_0 + \beta_1 x_1 + \beta_2 D_1 x_1 + \epsilon$$

Nesse contexto, o valor esperado da variável y quando a covariável binária D_1 vale 1 é dado por

$$E(y) = E(\beta_0 + \beta_1 x_1 + \beta_2(1)x_1 + \epsilon) = \beta_0 + (\beta_1 + \beta_2)x_1$$

De forma similar, quando D_1 tem valor 0, então o valor esperado para y é dado pela equação

$$E(y) = E(\beta_0 + \beta_1 x_1 + \beta_2(0)x_1 + \epsilon) = \beta_0 + \beta_1 x_1$$

Desta forma, a presença da covariável D_1 gera um incremento diferente a y sob um novo incremento em uma unidade de x_1 . Na prática, o grupo de observações que possuem a característica D_1 reagem de uma forma diferente a covariável x_1 .

Variáveis dummies são artifícios amplamente utilizados no contexto de Inferência Causal, especialmente no Método das Diferenças nas Diferenças (Card, 1992).

2.2 Modelo de Rubin e Inferência Causal

A Inferência Causal, isso é, a capacidade de determinar que um evento Y aconteceu por conta de um outro evento X , é debatido desde o começo do Século XX. Ainda na década de 20, Pearson argumentava que a atribuição de causa só era possível quando havia atribuição aleatória nos tratamentos, algo que num contexto de experimentação laboratorial é perfeitamente aceitável, mas num contexto de Ciência Sociais é impensável.

A dificuldade de se definir causa existe decorrente do fato de que não é possível observar a mesma unidade amostral sob o tratamento e sob o controle para comparar as saídas. Isso é, imagine que o estado do Paraná implemente um conjunto de leis antifumo a fim de reduzir o consumo de tabaco. Após alguns anos pode ser que se observe uma queda no consumo, mas jamais se saberá exatamente como teria sido esse consumo sem esse conjunto de leis. Não há um "Paraná 2", onde todas as

variáveis são idênticas, mas não há aplicação destas leis.

Este problema é o que se chama de Problema Fundamental da Inferência Causal. Ainda na década de 1920, Pearson argumentava que a atribuição aleatória de tratamentos era o suficiente para a atribuição de causa. Evidentemente isso não é possível num contexto de ciências sociais. Não há como "sortear" quais estados aplicarão quais leis e, mesmo que seja, os estados não seriam réplicas perfeitas uns dos outros.

Rubin (2005) propõe um framework para resolver esse problema: apesar de não ser possível observar a unidade de tratamento sob o tratamento e sob o controle, pode-se estimar qual seria a saída potencial dessa observação e, a partir daí, ter uma inferência causal. Para isso, Rubin se vale da suposição de não haver vazamento dos tratamentos (SUTVA - stable unit treatment value).

Na prática, isso quer dizer que o tratamento de uma observação não pode afetar a saída observada de outra. Essa é uma suposição difícil de se garantir no contexto de políticas públicas. Pensando, por exemplo, no contexto de leis armamentistas, se em um estado A aprova uma lei que aumenta a circulação de armas (tratamento) e o estado vizinho B mantém suas leis (controle), então é preciso supor que as armas circulantes no estado A não vai impactar o número de crimes violentos no estado B.

No contexto de mobilidade urbana isso se torna ainda mais desafiador. Para este estudo, foi necessário supor que a construção da Estação São Paulo – Morumbi não impactou o público palmeirense e corinthiano nos seus respectivos estádios. Na prática, isso é impossível de se assumir, uma vez que o torcedor palmeirense ou corinthiano que mora próximo à Estação São Paulo – Morumbi se beneficiou do ganho de mobilidade urbana e pôde acessar o estádio do seu time a um custo menor. Todavia, é necessário considerar esse tipo de efeito desprezível pela viabilidade do estudo (Ostrensky, Porsse & Frota, 2022).

2.3 Efeito de Tratamentos e Resultados Potenciais

A partir do Modelo de Rubin, isso é, do conceito de resultado potencial, é possível definir algumas medidas que são interessantes na discussão de Inferência

Causal. A primeira delas é o ATE (*Average Treatment Effect – Efeito Médio do Tratamento*), que é a diferença esperada entre os indivíduos que receberam o tratamento e as suas respectivas saídas potenciais. Esse é o dito efeito causal. Sua fórmula é dada por

$$ATE = E(Y^0 - Y^1) \quad (2.3.1)$$

Onde Y^0 são as unidades amostrais sob o grupo controle e Y^1 as unidades amostrais sob o tratamento. Além do ATE, existe também o ATT (*Average Treatment Effect on the Treated – Efeito Médio do Tratamento nos Tratados*), que é similar ao primeiro, mas tem sua esperança calculada condicionada ao tratamento. Sua fórmula é dada por

$$ATT = E(Y^0 - Y^1 | T = 1) \quad (2.3.2)$$

Ele é o impacto causal em quem recebeu o tratamento, o que permite entender o impacto da intervenção, e por isso $T = 1$. Enquanto o ATE estima o efeito de se construir uma estação no público dos estádios vizinhos, o ATT vai estimar o efeito que foi observado por terem construído a Estação Morumbi.

Outra medida de interesse no estudo da Inferência Causal é o ATC (*Average Treatment Effect on the Control – Efeito Médio do Tratamento nos Tratados*), que é similar a medida anterior, mas tem sua esperança condicionada ao grupo controle. Isso é, calcula-se o impacto causal que poderia ter sido observado, mas não foi. Ela é dada pela equação

$$ATC = E(Y^0 - Y^1 | T = 0) \quad (2.3.3)$$

E está fora do escopo da análise, uma vez que todos os estádios dentro do grupo controle já possuem uma estação nas suas redondezas.

2.4 Método Diferenças nas Diferenças

O método Diferenças nas Diferenças (Diff-in-Diff ou DiD) foi popularizado por

David Card na segunda metade do século XX, quando o pesquisador se propôs a estudar o efeito da imigração e de leis trabalhistas no mercado de trabalho americano (Card, 1992). É uma técnica que permite trabalhar com dados observacionais em políticas públicas já aplicadas.

A engenharia por trás do método DiD consiste em comparar a mudança no resultado de interesse para o grupo de tratamento antes e depois da intervenção, com a mudança no mesmo resultado para o grupo de controle no mesmo período. Card (1992), por exemplo, analisou uma mudança no salário mínimo em Nova Jersey para avaliar o impacto dessa política no desemprego em relação a Pensilvânia (onde o salário permaneceu fixo) e concluiu que não houve diferença nas diferenças entre os estados.

A principal vantagem do método é que ele não exige aleatorização, além de que permite aplicar técnicas de efeitos fixos, controlando até mesmo variáveis não observáveis, como especificidades de cada estado. Além de ser um modelo de fácil aplicação e explicabilidade. A principal dificuldade no uso do modelo está na sua suposição de trajetórias paralelas, onde o modelo supõe que os grupos, na ausência de qualquer intervenção, manteriam diferenças constantes ao longo do tempo.

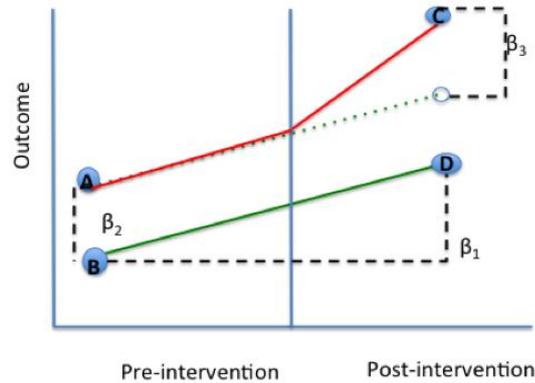
Apesar ser um modelo causal, o DiD numericamente é tão somente uma Regressão Linear. O modelo matemático pode ser dado por

$$y = \beta_0 + \beta_1 tempo + \beta_2 intervencao + \beta_3(tempo)(intervencao) + \epsilon \quad (2.4.1)$$

Onde a variável *tempo* é binária, onde o tempo pré-intervenção possui valor 0 e o pós-intervenção possui valor 1. A intervenção também é binário, sendo 1 para o grupo tratamento. O fator $(tempo)(intervencao)$ é a interação entre o grupo tratamento e o tempo, sendo, portanto, o impacto da intervenção em quem a recebeu.

O gráfico da Figura 4 facilita o entendimento da equação.

Figura 4 - Método Diferenças nas Diferenças



Fonte: Columbia University. 2024.

Neste caso, o valor esperado do grupo controle e do grupo tratado é dado pelas equações

$$E(y \mid \text{tempo} = 0 \text{ e } \text{intervencao} = 0) = \beta_0 \quad (2.4.2)$$

$$E(y \mid \text{tempo} = 0 \text{ e } \text{intervencao} = 1) = \beta_0 + \beta_2 \quad (2.4.3)$$

Para o $t = 1$ o valor esperado do grupo controle e do grupo tratado é dado por

$$E(y \mid \text{tempo} = 1 \text{ e } \text{intervencao} = 0) = \beta_0 + \beta_1 \quad (2.4.4)$$

$$E(y \mid \text{tempo} = 1 \text{ e } \text{intervencao} = 1) = \beta_0 + \beta_1 + \beta_2 + \beta_3 \quad (2.4.5)$$

O efeito causal da intervenção do tratamento, ATE, é o parâmetro β_3 , ele é a diferença das diferenças. Tal resultado é claro, bastando subtrair a diferença das equações 2.4.5 e 2.4.3 pela diferença das equações 2.4.4 e 2.4.2.

Nesse caso, trata-se a regressão da equação 2.4.1 como descrito na seção 2.1. A grande diferença é que para que se possa inferir causalidade (e não apenas correlação), é necessária toda a argumentação descrita na seção 2.2.

Zeldow, B., & Hatfield, LA (2021) sugerem, sem perda nenhuma de generalização, que é possível adicionar variáveis exógenas a esse modelo, sendo, então, o valor de y condicionado ao tempo, intervenção e, agora, as covariáveis adicionadas.

O modelo da equação 2.4.1 é também chamado de modelo 2x2, isso porque

trata-se de um modelo com dois grupos (controle e tratamento) em dois tempos (antes e depois da intervenção). Há extensões desse modelo, como o modelo diferenças nas diferenças 2xT onde dois grupos são observados em T tempos distintos. Essa extensão permite uma modelagem mais robusta da tendência e da suposição de trajetórias paralelas.

Baker (2025) sugere a utilização de efeitos fixos no tempo, onde para cada período T observado é estimado um efeito fixo comum a todos os grupos, sendo assim o modelo 2xT é dado por

$$y = \beta_0 + \beta_1 \text{intervencao} + \sum_{i \in T} \delta_i \text{tempo}_i + \sum_{i \in T^t} \beta_2 \text{tempo}_i \cdot \text{intervencao} + \epsilon \quad (2.4.2)$$

Onde tempo_i é uma variável dummie para cada tempo e δ_i é a média de y comum a todos os grupos, já $\text{tempo}_i \cdot \text{intervencao}$ é uma interação entre o tempo e a intervenção, sendo 1 somente nos tempos T^t onde o grupo tratamento já recebeu a intervenção. Trata-se tão somente de uma generalização do modelo 2x2.

2.5 Demanda pelo Público nos Estádios de Futebol

A gestão esportiva tem se consolidado como uma área de pesquisa estratégica, com foco em maximizar a performance financeira e administrativa dos clubes. Estudos aprofundados sobre salários de atletas, demanda por ingressos e outros fatores externos demonstram a importância de uma gestão eficiente para o sucesso das equipes.

A literatura sobre o público nos estádios pode ser considerada extensa. Veiga (2019) fez uma análise exploratória sobre o público são paulino nos mais diversos campeonatos e discorreu sobre hipóteses que poderiam explicar a presença do público no estádio.

A probabilidade de vitória é também uma variável significativa a ser analisada quando se trata de público nos estádios. Giovannetti, Rocha, Sanches & Silva (2006) analisaram a fidelidade do público dos estádios no campeonato brasileiro de 2004 de acordo com a probabilidade de vitória. Schaffer (2024) analisa um lado mais obscuro da probabilidade de vitória, onde seus resultados indicam que resultados de jogos diferentes dos previstos por casas de apostas (uma proxy pra probabilidade de vitória)

estava associada a um aumento de casos de violência doméstica.

Bortoluzzo (2017) analisou, sob um viés de marketing, de maneira preditiva os públicos no estádio, usando covariáveis como ambiente econômico, qualidade do produto, além de incentivos monetários e não-monetários para o torcedor.

O regulamento também é um fator decisor do público nos estádios. Simões (2023) realizou uma análise da CNN em conjunto com a QUAEST, conceituada empresa de pesquisas, onde torcedores argumentam preferir jogos eliminatórios. A pesquisa teria contado com 6.507 torcedores.

A estreia de jogadores e técnicos prestigiados também é um grande atrativo para ida do torcedor ao estádio. Lozetti e Hazan (2017) chamam de “Efeito Ceni” o público observado na reestrela do ídolo são paulino como técnico em 2017. Esse efeito pode ser observado em outros casos, como em 2012, quando o São Paulo também bateu seu recorde de público com a estreia de Paulo Henrique Ganso (Prates, 2012).

2.6 Estação Morumbi e Conflitos Sociais

O bairro Morumbi é uma região tradicional da cidade de São Paulo, tendo, segundo o IBGE, um dos maiores IDH (Índice de Desenvolvimento Humano) da cidade (0,938). O bairro passou por enormes transformações sociais nas últimas décadas e, como aponta Caldeira (2000), o bairro deixou de se ser um reduto de mansões e terrenos vazios na década de 70 para ter uma composição mista entre estes casarões e condomínios verticais de alto padrão, além, é claro, de favelas.

Para se ter uma ideia, em 2000 estima-se que em torno de 47,7% da população que morava na região do Morumbi era composta por favelas. A mais famosa delas, Paraisópolis, já chegou a estar entre as 5 maiores favelas do Brasil. Estima-se que 80% da população de Paraisópolis tenha origem nordestina e que tenham chegado à região durante a década de 80, auge do êxodo rural brasileiro, para trabalhar no boom da construção civil (Gohn, 2010). Hoje, grande parte desses trabalhadores são porteiros e faxineiros nos condomínios que construíram e suas esposas e filhas trabalham como empregadas domésticas nos apartamentos (Gohn, 2010).

Neste misto entre favelas, mansões e condomínios de alto padrão, os

moradores do bairro Morumbi se organizaram nas últimas décadas em associações de moradores. Evidentemente que, em meio a tamanha heterogeneidade social, os interesses são quase sempre distintos. Enquanto os representantes das áreas mais nobres defendem uma manutenção das origens do bairro, um lugar “fechado” em termos urbanísticos, as entidades representantes das favelas defendem um maior acesso a saneamento básico e espaços de lazer (Gohn, 2010).

A Estação dos Três Poderes é um bom caso que evidencia esse conflito. Planejada para ter o início de suas obras em 2005 no bairro Morumbi, foi retirada dos planos da prefeitura após reivindicação de moradores que argumentavam que a estação iria gerar uma mudança na característica do bairro (Mora, 2011). Fioravanti (2013) aborda esse conflito no bairro Morumbi, trazendo o conflito social entre moradores de alta renda, contra a construção de uma estação de metrô próxima às suas residências, e o morador “periférico” da região.

Nesse contexto, uma expansão da mobilidade urbana, apesar de estar associada a uma democratização do direito ao acesso a cidade (Boareto, 2008) torna-se um desafio para as prefeituras, especialmente quando se fala de bairros nobres.

3 MATERIAL E MÉTODOS

3.1 Material

O estudo foi conduzido utilizando a linguagem R e o software RStudio como editor de código e ambiente de desenvolvimento. Para a modelagem, foi utilizado o pacote `plm`, clássico na literatura de dados em painel.

Como base para o estudo foi utilizado as informações de bilheteria do campeonato paulista de 2015 a 2024 extraídos da FPF (Federação Paulista de Futebol). O escopo de jogos foi limitado a jogos nos Estádios Morumbi, Allianz Arena e Neo Química Arena, mandados respectivamente pelos clubes São Paulo Futebol Clube, Sociedade Esportiva Palmeiras e Sport Club Corinthians.

A limitação do escopo aos jogos do campeonato paulista se deve a dificuldade de se modelar o público em campeonatos nacionais e internacionais, uma vez que o risco de rebaixamento, a disputa de título ou a disputa pela eventual classificação para outros campeonatos está fortemente associado à presença do público, aumentando

consideravelmente a dificuldade da modelagem.

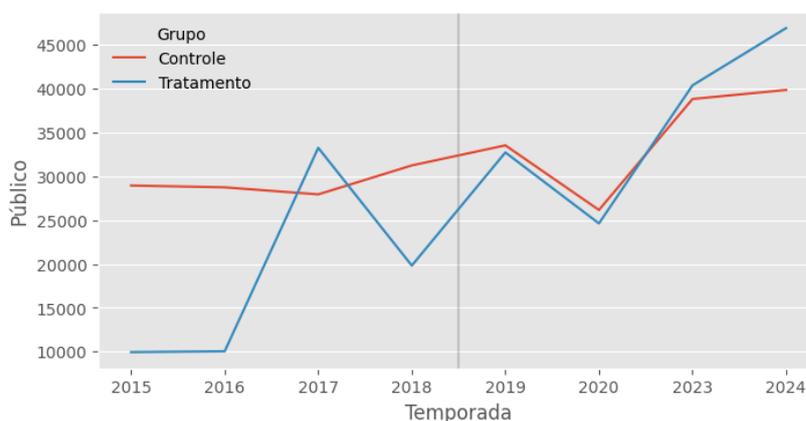
Além disso, foi retirado da análise os anos de 2021 e 2022, porque ambos foram afetados pelas políticas de lockdown durante o covid-19, o que dificultaria a robustez das suposições, uma vez que haveria o risco de estar sendo modelado outro tratamento (retorno da torcida aos estádios), o que fere as suposições do Modelo de Rubin.

3.2 Método

3.2.1 ANÁLISE EXPLORATÓRIA

O primeiro ponto a ser analisado é a trajetória do público ao longo das temporadas analisadas nos grupos controle e tratamento. A Figura 5 indica haver uma tendência geral no crescimento do público, com destaque para uma observação atípica em 2017 para o grupo tratamento.

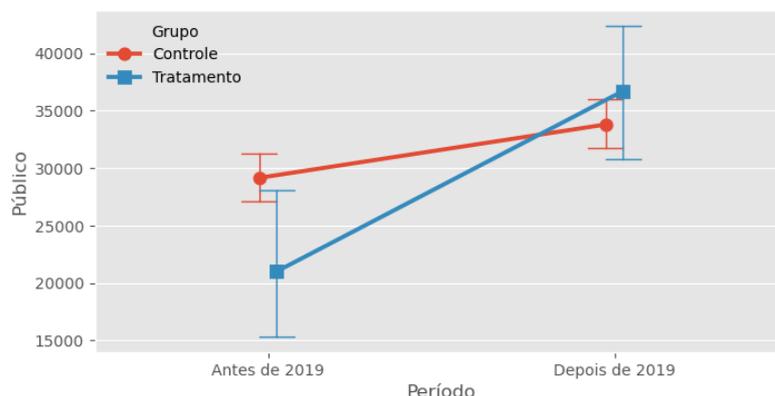
Figura 5 - Análise de Bilheteria Antes e Após a Construção da Estação São Paulo-Morumbi – Análise de Tendência



Fonte: Federação Paulista de Futebol.

A Figura 6 apresenta distribuição do público antes e depois da construção da Estação São Paulo – Morumbi, confirmando uma tendência geral do aumento da torcida, mas um aumento maior ainda para o público são paulino.

Figura 6 - Bilheteria Antes e Após a Construção da Estação São Paulo-Morumbi

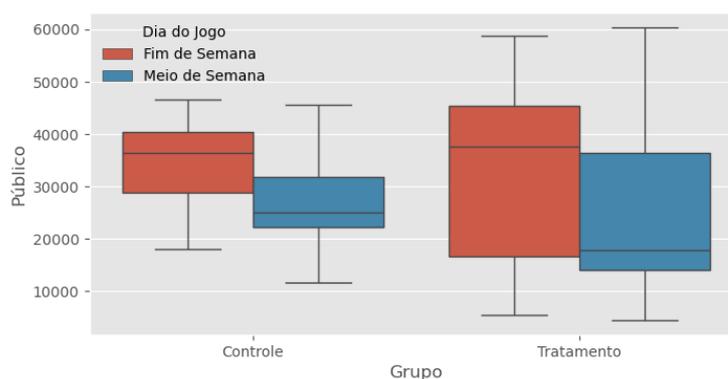


Fonte: Federação Paulista de Futebol.

É importante ressaltar que uma tendência no público para grupo controle (Corinthians e Palmeiras mandando nos seus respectivos estádios) não fere qualquer suposição do Método DiD. O método supõe uma trajetória paralela entre o grupo controle e a saída potencial do grupo de tratamento.

Algumas covariáveis de data podem ser analisadas também, como o dia da semana em que o jogo ocorreu. A Figura 7 evidencia que jogos no fim de semana estão associados a um público tanto para o grupo controle quanto para o grupo tratamento.

Figura 7 - Bilheteria em Jogos em Dia de Semana e Fim de Semana

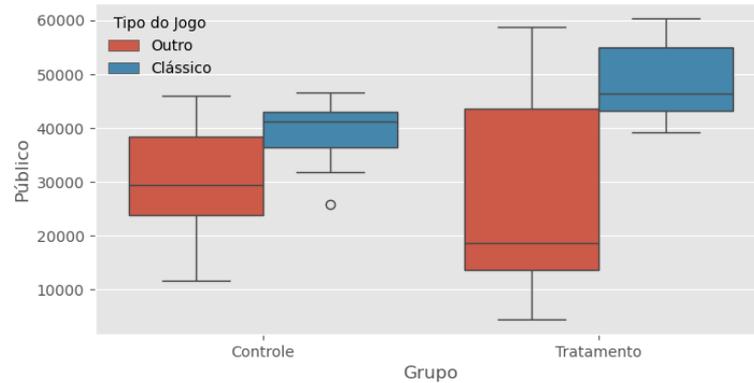


Fonte: Federação Paulista de Futebol.

A importância do jogo é outra covariável que explica a variação no público dos estádios. A Figura 8 traz o público observado no grupo de controle e tratamento quando a partida se tratava ou não de um clássico. Aqui clássico foi definido como

qualquer jogo entre os 4 principais clubes de São Paulo: São Paulo Futebol Clube, Palmeiras, Santos e Corinthians. É válido ressaltar que dos 26 registros de clássicos na base, 18 foram em um fim de semana, afinal, os clubes sabendo que jogos em fins de semana tem maior público, tendem a preferir clássicos nestas datas.

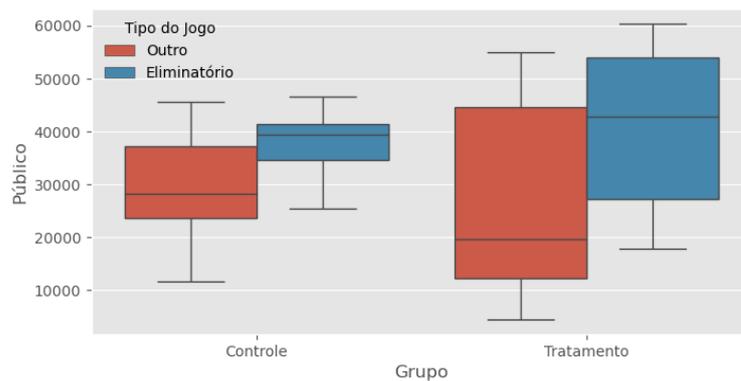
Figura 8 - Bilheteria em Clássicos



Fonte: Federação Paulista de Futebol.

Por fim, o fato de o jogo ser eliminatório ou não também está fortemente associado com a presença do público, como mostra a Figura 9, o que está em linha com a preferência do público brasileiro, levantado na seção 2.5.

Figura 9 - Bilheteria em Jogos Eliminatórios



Fonte: Federação Paulista de Futebol.

3.2.2 MÉTODO DIFERENÇAS NAS DIFERENÇAS

O primeiro modelo a ser testado foi o modelo Diferenças nas Diferenças na sua forma mais simples, sem variáveis exógenas, é dado por

$$Y_i = \beta_0 + \beta_1 \text{Morumbi} + \beta_2 \text{Pos}_{2019} + \beta_3 \text{Construcao_Estacao} + \epsilon_i \quad (3.1)$$

Onde Y_i é o público do i -ésimo jogo da base e todas as covariáveis são dummies. A covariável *Morumbi* recebe o valor 1 para os jogos mandados pelo São Paulo e 0 para os demais, ela é responsável por indicar a diferença sem intervenções do grupo controle e tratamento. A covariável *Pos_2019* recebe valor 1 para todos os jogos a partir de 2019, o ano da construção da estação, e 0 para os anteriores, captando qualquer tendência comum aos grupos controle e tratamento. Já a covariável *Construcao_Estacao* recebe valor 1 somente para jogos do São Paulo no Morumbi após 2019, sendo ela o efeito causal a ser estimado.

A Tabela 1 apresenta a saída desse modelo, onde foi estimado um efeito significativo de 10.876 torcedores no público do Estádio Morumbi a partir da construção da estação São Paulo – Morumbi. Além disso, foi estimado uma tendência de 6305 torcedores comum aos grupos controle e tratamento.

Tabela 1 – Saída do Modelo Diff-in-Diff

	Estimate	Std. Error	t value	Pr(> t)
β_0	29.1551	1.3377	21.80	0.0000
$\beta_1 \cdot \text{Morumbi}$	-8.1444	2.5468	-3.20	0.0017
$\beta_2 \cdot \text{Pos}_{2019}$	6.3054	2.0464	3.08	0.0024
$\beta_3 \cdot \text{Construcao_Estacao}$	10.8766	3.7428	2.91	0.0042

3.2.3 MÉTODO DIFERENÇAS NAS DIFERENÇAS COM COVARIÁVEIS CONFUNDIDAS

O modelo Diff-in-Diff tem uma forte suposição de que não há nenhum outro tratamento ocorrendo na análise. Esses tratamentos podem ser entendidos como

covariáveis confundidoras, que sua omissão pode estar enviesando a estimativa do efeito causal.

A seção 3.2.1 demonstrou que há várias covariáveis que podem explicar a variação do público entre os jogos. Nesse sentido, foi testado um modelo com covariáveis relacionadas ao dia da semana do jogo, o caráter eliminatório ou não, o jogo ser um clássico e, também, foi adicionado um fator específico em 2017 para controlar o efeito da volta do ídolo São Paulino, Rogerio Ceni, como treinador.

$$\begin{aligned}
 Y_i = & \beta_0 + \beta_1 \text{Morumbi} + \beta_2 \text{Pos}_{2019} + \beta_3 \text{Construcao_Estacao} & (3.9) \\
 & + \beta_4 \text{Classico} + \beta_5 \text{Fim_de_Semana} + \beta_6 \text{Eliminatorio} \\
 & + \beta_7 \text{Efeito_Ceni} + \epsilon_i
 \end{aligned}$$

Como pode ser visto na equação 3.9, não há suposição de interação entre as covariáveis. Isso é: a partida ser um clássico ou não tem o mesmo efeito quando se trata de um jogo eliminatório. Não há nenhum efeito adicional ao fato de o jogo ser um clássico numa semi-final. É uma suposição de modelagem, poder-se-ia utilizar um termo como $\beta_8 \text{Classico} \cdot \text{Eliminatório}$, por exemplo.

A Tabela 2 indica que agora há uma estimativa de um impacto maior da construção da estação São Paulo - Morumbi. Isso é, se antes havia uma estimativa de que o efeito da construção tinha uma distribuição $N(10.877, 3.743^2)$, agora a estimativa é que o efeito causal tem distribuição $N(15.556, 3.048^2)$.

Tabela 2 – Saída do Modelo Diff-in-Diff com Covariáveis

	Estimate	Std. Error	t value	Pr(> t)
β_0	21.2352	1.3500	15.73	0.0000
$\beta_1 \cdot \text{Morumbi}$	-13.8311	2.2400	-6.17	0.0000
$\beta_2 \cdot \text{Pos}_{2019}$	6.8025	1.5353	4.43	0.0000
$\beta_3 \cdot \text{Construcao_Estacao}$	15.5559	3.0476	5.10	0.0000
$\beta_4 \cdot \text{Fim_de_Semana}$	7.4436	1.3751	5.41	0.0000
$\beta_5 \cdot \text{Classico}$	9.8322	1.9464	5.05	0.0000
$\beta_6 \cdot \text{Eliminatorio}$	6.8960	1.5397	4.48	0.0000
$\beta_7 \cdot \text{Efeito_Ceni}$	17.6024	3.4446	5.11	0.0000

A mudança na estimativa do efeito causal pode ser explicada pela inclusão dessas covariáveis confundidoras. Por exemplo, em 2017 o São Paulo teve uma

média de público 33.243 pagantes, enquanto nos anos 2015, 2016 e 2018 teve um público médio de 14.894, menos da metade. Quando é informado para o modelo que em 2017 teve um “Efeito Ceni” único aos demais dados, o modelo reajusta de -8.144 para -13.883 a estimativa da diferença entre o público médio para o São Paulo pré 2019 e o grupo controle (β_1 *Morumbi*). Dessa forma, o impacto da intervenção torna-se maior, uma vez que a diferença das diferenças torna-se maior.

O modelo agora ganha também uma interpretação ligeiramente mais complexa. Agora pode se pensar que um jogo que não é no Morumbi (*Morumbi* = 0), antes de 2019 (*Pos_2019* = 0 e *Construcao_Estacao* = 0), no meio da semana (*Fim_de_Semana* = 0) e que não se trata de um clássico ou um jogo eliminatório (*Classico* = 0 e *Eliminatorio* = 0), tem um público esperado de 21.235 torcedores.

3.2.4 DIFERENÇAS NAS DIFERENÇAS COM MODELO DE EFEITOS FIXOS NO TEMPO

A forma com que foi conjecturado o modelo 3.10 nos diz que basicamente só há dois efeitos do tempo: antes de 2019 e depois de 2019. Essa é uma hipótese tradicional em modelos Diff-in-Diff, que geralmente trabalham só com duas observações no tempo (o antes e o depois da intervenção). Todavia, no nosso conjunto de dados as unidades amostrais foram observadas em 8 tempos distintos.

Uma possibilidade de enfrentar esse problema é estimando um efeito fixo temporal para o modelo. Isso é, vai existir um efeito comum a todas as unidades amostrais em cada tempo, conforme a equação 3.11. Nesse caso foi removida a covariável *Pos_2019*, pois ela seria uma combinação linear perfeita dos efeitos temporais de 2019 em diante.

$$\begin{aligned}
 Y_{ti} = & \delta_t \text{Temporada}_t + \beta_0 + \beta_1 \text{Morumbi} + \beta_2 \text{Construcao_Estacao} & (3.11) \\
 & + \beta_3 \text{Classico} + \beta_4 \text{Fim_de_Semana} + \beta_5 \text{Eliminatorio} \\
 & + \beta_6 \text{Efeito_Ceni} + \epsilon_i
 \end{aligned}$$

Tabela 3 – Saída do Diff-in-Diff com Efeitos Fixos no Tempo

	Estimate	Std. Error	t value	Pr(> t)
β_0	20.4769	1.7002	12.04	0.0000
$\beta_1 \cdot Morumbi$	-15.3080	2.1610	-7.08	0.0000
$\beta_2 \cdot Construcao_Estacao$	16.0288	2.8594	5.61	0.0000
$\beta_3 \cdot Fim_de_Semana$	7.9780	1.2735	6.26	0.0000
$\beta_4 \cdot Classico$	9.0691	1.7721	5.12	0.0000
$\beta_5 \cdot Eliminatório$	6.5443	1.3968	4.69	0.0000
$\beta_6 \cdot Efeito_Ceni$	19.7511	3.7124	5.32	0.0000

A Tabela 5 apresenta a saída dos parâmetros estimados pelo modelo. Nesse caso, a covariável presente nos demais modelos *Pos_2019* foi absorvida por efeitos fixos que podem ser vistos na Tabela 6.

Tabela 4 – Efeitos Fixos Estimados do Diff-in-Diff com Efeitos Fixos no Tempo

Temporada	Estimated_Effects
2015	-3.848
2016	-4.637
2017	-4.142
2018	0.146
2019	-0.841
2020	-4.160
2023	6.952
2024	10.272

Os efeitos fixos nesse caso trabalham como uma espécie de “baseline”. Isso é, pode ser que o campeonato paulista tenha uma heterogeneidade ao longo dos anos no público.

4 RESULTADOS E DISCUSSÃO

Foram treinados 3 modelos distintos na análise. O primeiro e o segundo modelo podem se dizer encaixados. Nesse caso, foi feito um teste de razão de verossimilhança para testar se o modelo com mais covariáveis é realmente melhor do que o mais enxuto, que obteve estatística chi-quadrado 94.969 ($p \text{ valor} < 2.2 \cdot 10^{-16}$) para a hipótese nula de que não há um ganho de explicabilidade da variância na adição de mais covariáveis, o que sustentou a decisão de que o modelo com covariáveis deve ser utilizado em detrimento do modelo sem.

Para o com feitos fixos no tempo fazer essa análise é um tanto mais delicado.

O modelo Diff-in-Diff já faz um controle temporal com a covariável Pos_2019. Desta forma, o uso de covariáveis com efeitos fixos no tempo se justificaria se, e somente se, houvesse uma heterogeneidade entre os tempos agrupados por esta covariável. Isso é, o efeito médio do ano 2019 deve ser igual ao efeito do ano 2020, 2023 e 2024. O que não é o que se observa na saída Tabela 8, onde 2020 não teve significância e os anos 2023 e 2024 tiveram. Desta forma o modelo de efeitos fixos no tempo não é descartado.

Tabela 5 – Saída do Modelo Diff-in-Diff com Efeitos Fixos no Tempo com p-valores

	Estimate	Std. Error	t value	Pr(> t)
β_0	20.4769	1.7002	12.04	0.0000
$\beta_1 \cdot Morumbi$	-15.3080	2.1610	-7.08	0.0000
$\beta_2 \cdot Construc\tilde{a}o_Estacao$	16.0288	2.8594	5.61	0.0000
$\beta_3 \cdot Fim_de_Semana$	7.9780	1.2735	6.26	0.0000
$\beta_4 \cdot Classico$	9.0691	1.7721	5.12	0.0000
$\beta_5 \cdot Eliminatorio$	6.5443	1.3968	4.69	0.0000
$\beta_6 \cdot Efeito_Ceni$	19.7511	3.7124	5.32	0.0000
$\delta_{2016} \cdot temporada_{2016}$	-0.7889	2.3437	-0.34	0.7369
$\delta_{2017} \cdot temporada_{2017}$	-0.2932	2.3099	-0.13	0.8992
$\delta_{2018} \cdot temporada_{2018}$	3.9947	2.1262	1.88	0.0623
$\delta_{2019} \cdot temporada_{2019}$	3.0070	2.2453	1.34	0.1826
$\delta_{2020} \cdot temporada_{2020}$	-0.3120	2.7031	-0.12	0.9083
$\delta_{2023} \cdot temporada_{2023}$	10.8000	2.1983	4.91	0.0000
$\delta_{2024} \cdot temporada_{2024}$	14.1204	2.3976	5.89	0.0000

Como os com covariáveis e efeitos fixos não são exatamente encaixados, a decisão do modelo a ser usado foi embasada ao atendimento as suposições da regressão linear e a critérios de ajuste, como AIC (critério de informação de Akaike). Isso é, o modelo que melhor nos garante ter estimativas não enviesadas para os parâmetros estimados. Todavia, é válido pontuar que o efeito causal dos dois modelos conversa entre si.

A Figura 10 indica que não há evidências de correlação entre os valores previstos, por outro lado a Figura 11 indica haver uma estrutura temporal no modelo diff-in-diff com covariáveis, o que não acontece no modelo com efeitos fixos no tempo.

Figura 10 – Resíduo dos modelos Diff-in-Diff com Covariáveis e Diff-in-Diff com Efeitos Fixos no Tempo

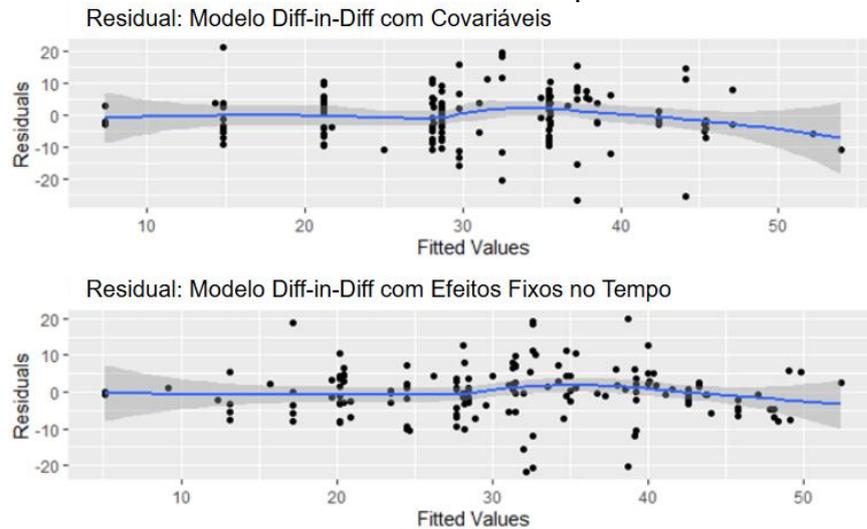
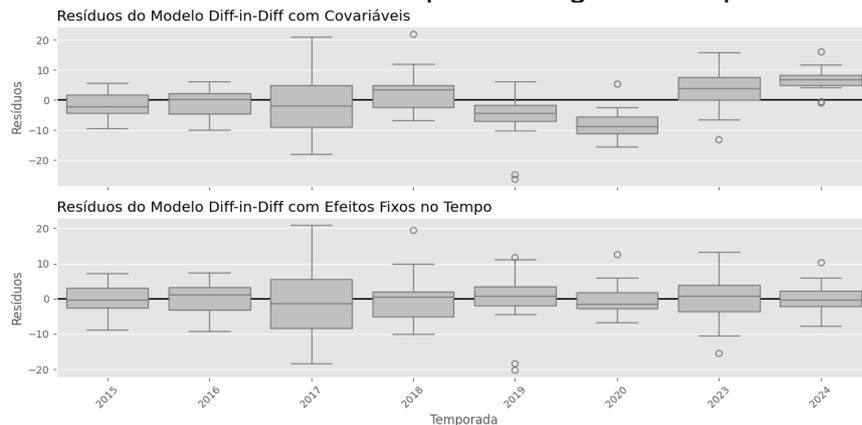


Figura 11 – Resíduo dos modelos Diff-in-Diff com Covariáveis e Diff-in-Diff com Efeitos Fixos no Tempo ao Longo do Tempo



Foi realizado um teste de Shapiro-wilk, apresentado na tabela 9, que indicou que somente o modelo com covariáveis obteve distribuição normal nos resíduos. Apesar disso, este modelo obteve um AIC de 1401.63 contra um AIC de 1345.73 do modelo diff-in-diff com efeitos fixos, indicando não ser o modelo o modelo que melhor equilibra o ajuste aos dados com o número de covariáveis. Por outro lado, o modelo de diff-in-diff com efeitos fixos obteve um *R-quadrado overall* de 0.401 contra um *R-quadrado overall* de 0.592 do modelo diff-in-diff com covariáveis, indicando um maior ajuste para o segundo.

Tabela 6 – Resultados do Teste Shapiro-Wilk

Model	Pr(> t)
DiD cov	0.05
DiD Efeitos Fixos	0.00

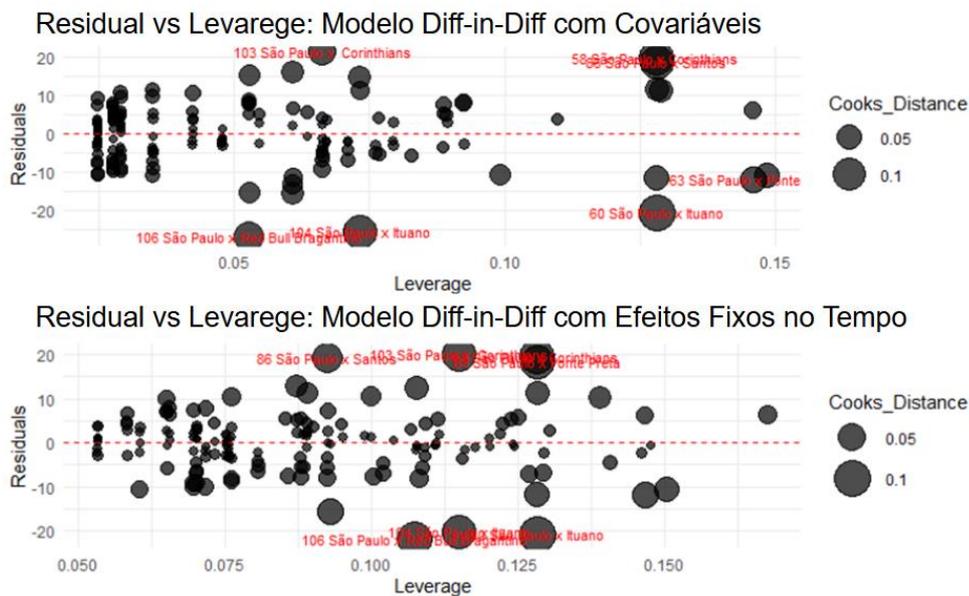
A Figura 13 indica que ambos os modelos possuem observações influentes que, por acaso, se trata das mesmas. No caso, em ambos os modelos tiveram as mesmas observações como influentes. As observações subestimadas pelos modelos foram: São Paulo x Corinthians na primeira fase do Paulistão 2017 e São Paulo x Ponte Preta, também em 2017, São Paulo x Santos em 2018 e São Paulo x Corinthians na final do Paulistão em 2019.

As duas podem ser atribuídas, ainda, ao Efeito Ceni (Lozetti e Hazan, 2017), posto que foram, respectivamente, no primeiro jogo de Ceni no Morumbi, bem como no seu primeiro clássico em casa. Como houve um declínio no público dos jogos conforme o Paulistão avançava, a covariável Efeito_Ceni captou apenas o “efeito médio” de 2017. As demais se trata de clássicos que podem ter efeitos não observáveis dentro do próprio jogo, como uma rivalidade mais acalorada nas semanas que antecedem a partida.

Os jogos onde os modelos superestimaram o público foram São Paulo x Bragantino e São Paulo x Ituano em 2019, bem como São Paulo x Ituano em 2017. Os dois primeiros jogos são observações atípicas justificadas pela eliminação na pré-libertadores do São Paulo alguns dias antes dos jogos (Canônico e Razan, 2019).

Apesar de que, sim, poderia ser adicionada covariáveis para explicar essas observações atípicas, fazer isto poderia desencadear um looping onde cada vez mais adiciona-se covariáveis para explicar novas observações atípicas e produz-se um modelo com um excesso de ajuste (*overfitting*). Desta forma, aceita-se essas covariáveis com mais peso, devido à complexidade do fenômeno que está sendo analisado.

Figura 12 - Resíduos x Peso com Distância de Cook



Devida a simplicidade maior do modelo e a melhor atendimento a suposição de distribuição normal, define-se o modelo Diff-in-Diff com covariáveis confundidoras, como o melhor modelo não enviesado para estimar o efeito causal da Construção da Estação São Paulo-Morumbi no público da torcida são paulina.

Um ponto importante agora é discutir se as suposições de causalidade do Modelo de Rubin foram atendidas. É um ponto interessante discutir vazamento de tratamentos quando se trata de mobilidade urbana. Há sempre uma limitação em afirmar que o grupo controle não se beneficia do tratamento. O que não é uma verdade, afinal, torcedores do Corinthians e do Palmeiras da região Morumbi se beneficiaram da Construção de Estação para acessar os estádios dos seus clubes. Todavia, o que se pode assumir é que esse efeito do vazamento é desprezível, afinal, sem essa suposição não tem como fazer inferência causal.

A estimativa final para o efeito causal no grupo tratado da construção da Estação São Paulo-Morumbi pode ser vista na equação 4.1.

$$ATT \sim N(15.556, 3.048^2) \quad (4.1)$$

Como ao todo desde 2019 o São Paulo Futebol Clube mandou 22 jogos pelo Paulistão no Morumbi, pode-se estimar o efeito acumulado dessa política apenas

somando 22 vezes o efeito ATT, sob a evidente suposição de que os jogos são independentes. A estimativa, portanto, para o efeito acumulado nos últimos 5 anos (ignorando os jogos de 2022) está na equação 4.2.

$$ATT_{acumulado\ 2019,2020,2023,2024} \sim N(342.232, 14.294^2) \quad (4.2)$$

Para os próximos 10 anos, supondo que o clube paulista mantenha sua média de 6 jogos no Morumbi e o efeito estimado se mantenha constante, espera-se que mais de 933.360 torcedores frequentem o estádio graças a essa política pública. Além do mais, pode se estimar que entre 2015 e 2018 373.344 torcedores deixaram de ir ao estádio devido a falta da Estação São Paulo-Morumbi.

5 CONCLUSÕES E COMENTÁRIOS FINAIS

O presente estudo buscou estimar um efeito da causal da construção da Estação São Paulo-Morumbi no público do Estádio Morumbi da torcida são-paulina. Para isso se valeu do modelo Diff-in-Diff usando covariáveis como possíveis confundidoras. A estimativa foi de que a construção da estação garantiu um aumento médio de 15.556 torcedores, tendo um erro padrão de 3.048 associado a essa estimativa. Além disso, estima-se que cerca de 342.232 torcedores que, num cenário contrafactual onde não foi construída a estação, não iriam, frequentaram o estádio do Morumbi pelo Paulistão nos últimos 5 anos.

O estudo possui seu escopo reduzido ao paulistão, bem como não destaca se esse aumento veio em arquibancadas populares, devido à escassez e dificuldade de se obter mais dados sobre o clube. A análise do impacto desse aumento de torcedores pode, em trabalhos futuros, ser expandida para outros campeonatos, bem como expandida para uma análise financeira que essa intervenção possibilitou para o clube, que, com mais torcedores, terá uma possibilidade de maior arrecadação com bilheteria e vendas de produtos oficiais no estádio. Outra análise futura a ser feita é o impacto desse aumento de torcedores na economia em volta do estádio.

REFERÊNCIAS

- CARD, D.; KRUEGER, A. B. **Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania**. *American Economic Review*, v. 84, n. 5, p. 772-793, 1994
- SCHÄFFER, Yuri. **Dia de Jogo: Uma análise da relação entre o futebol e a violência doméstica no Brasil entre 2015 e 2019**. In: Anais do XXIII Encontro Nacional de Estudos Populacionais, 2024, Brasília. Anais eletrônicos..., Campinas, Galoá, 2024.
- BORTOLUZZO, Adriana Bruscato et al. Previsão de consumo de ingressos para jogos do campeonato brasileiro. **Revista de Administração**, v. 52, n. 1, p. 70-80, 2017.
- VEIGA, R. P. **ESTUDO DA DEMANDA E PREÇO POR INGRESSOS DE FUTEBOL: O CASO DO SÃO PAULO FUTEBOL CLUBE**. Trabalho de Conclusão de Curso (Graduação em Ciências Econômicas) - Insper, São Paulo, 2019. Disponível em: <https://repositorio.insper.edu.br/handle/11224/3949>. Acesso em: 01/05/2024.
- GIOVANNETTI, B.; ROCHA, B. P.; SANCHES, F. M.; SILVA, J. C. D. da. **Medindo a fidelidade das torcidas brasileiras: uma análise econômica no futebol**. *Revista Brasileira de Economia*, ¹ Rio de Janeiro, v. 60, n. 4, p. 389-406, ² 2006.
- Gohn, M. da G. (2010). MORUMBI: o contraditório bairro-região de São Paulo. *Caderno CRH*, 23(59). <https://doi.org/10.9771/ccrh.v23i59.19103>
- BOARETO, R. A política de mobilidade urbana e a construção de cidades sustentáveis. **Revista dos Transportes Públicos - ANTP**, Rio de Janeiro, v. 30/31, n. 3/4, p. 143 – 160, 2008.
- OSTRENSKY, V. P.; PORSSE, A. A.; FROTA, L. M. da. Public transport and gentrification. Evidence from São Paulo metro new stations. **Volume 14, Issue 6**, Special Issue: Accessibility and Inequality in Latin-America, p. 254-269, dez. 2022.
- CALDEIRA, Tereza Pires. *Cidades de muros. Crime, segregação e cidadania em São Paulo*. São Paulo: Ed. 34/EDUSP, 2000.

MORA, Marcelo. Morador pode decidir melhor para bairro, diz líder sobre estação de Metrô. **G1**, São Paulo, 2011. Disponível em: <https://g1.globo.com/sao-paulo/noticia/2011/05/morador-pode-decidir-melhor-para-bairro-diz-lider-sobre-estacao-de-metro.html>. Acesso em: 01/12/2024.

SIMÕES, A. Vinte anos de pontos corridos: torcedor ainda prefere Brasileirão por mata-mata. **CNN Brasil**, 2024. Disponível em: <https://www.cnnbrasil.com.br/esportes/outros-esportes/brasileirao-em-20-anos-de-pontos-corridos-mata-mata-ainda-tem-a-preferencia-do-torcedor/>. Acesso em: 05/12/2024.

LOZETTI, A.; HAZAN, M. **Efeito Ceni: São Paulo consegue 3º maior público do século no Paulistão**. Globo Esporte, São Paulo, 2017. Disponível em: <https://ge.globo.com/futebol/times/sao-paulo/noticia/2017/02/efeito-ceni-sao-paulo-consegue-3-maior-publico-do-seculo-no-paulistao.html>. Acesso em: 01/12/2024.

PRATES, R. **Ganso vive tarde de ídolo logo na estreia pelo São Paulo e lida com pressão por status**. UOL Esporte, São Paulo, 2012. Disponível em: <https://www.uol.com.br/esporte/futebol/campeonatos/brasileiro/serie-a/ultimas-noticias/2012/11/19/ganso-vive-tarde-de-idolo-logo-na-estrela-pelo-sao-paulo-e-lida-com-pressao-por-status.htm>. Acesso em: 01/12/2024.

GLOBO ESPORTE. Maiores torcidas do Brasil: pesquisa mostra Flamengo na ponta, e Palmeiras mais perto do São Paulo. **ge.globo**, Rio de Janeiro, 21 ago. 2024. Disponível em: <https://ge.globo.com/futebol/noticia/2024/08/21/maiores-torcidas-do-brasil-pesquisa-mostra-flamengo-na-ponta-e-palmeiras-mais-perto-do-sao-paulo.ghtml>. Acesso em: 01/12/2024.

BAKER, Andrew; CALLAWAY, Brantly; CUNNINGHAM, Scott; GOODMAN-BACON, Andrew; SANT'ANNA, Pedro H. C. Difference-in-Differences Designs: A Practitioner's Guide. [S. l.]: [s. n.], 18 jun. 2025. Working Paper. Disponível em:

<https://arxiv.org/pdf/2503.13323>. Acesso em: 07 jul. 2025.