

UNIVERSIDADE NOVE DE JULHO - UNINOVE
PROGRAMA DE PÓS GRADUAÇÃO EM INFORMÁTICA E
GESTÃO DO CONHECIMENTO - PPGI

Henrique Pougy

O Plano e o Mercado

*Uma análise bayesiana da efetividade do planejamento urbano em direcionar o
adensamento construtivo em São Paulo*

São Paulo
2024

Henrique Pougy

O Plano e o Mercado

Uma análise bayesiana da efetividade do planejamento urbano em direcionar o adensamento construtivo em São Paulo

Dissertação apresentada ao Programa de Pós-Graduação em Informática e Gestão do Conhecimento (PPGIGC) da Universidade Nove de Julho - UNINOVE, como parte dos requisitos para a obtenção do título de Mestre em Informática e Gestão do Conhecimento.

Prof. Orientador: Dr. Wonder Alexandre Luz

São Paulo

2024

DEDICATÓRIA

Para Eliana.

AGRADECIMENTOS

Agradeço a meus pais, Eliana e Rodrigo, e à minha companheira Luísa, pelo apoio durante a escrita dessa dissertação.

Agradeço a meu orientador, prof. Wonder Alexandre Luz, pela orientação atenta e cuidadosa.

Agradeço ao prof. José Eduardo Storópoli por me ensinar estatística bayesiana e como a ciência deve ser feita: de forma aberta e reprodutível.

À Universidade Nove de Julho (UNINOVE), pela oportunidade e pela concessão da bolsa de estudos.

Agradeço, por fim, à CAPES, por ter me apoiado durante parte da elaboração deste trabalho.

EPÍGRAFE

“A última e grande ilusão é o próprio plano”.

Flavio Villaça (2005)
Arquiteto e Urbanista.

RESUMO

Promover o aumento do direito à cidade por meio do adensamento construtivo próximo aos eixos de transporte é um dos principais objetivos do Plano Diretor Estratégico do Município de São Paulo. No entanto, o sucesso desta política não depende apenas da decisão do Poder Público: requer a adesão do capital imobiliário, a quem compete produzir os empreendimentos que gerarão o crescimento objetivado. Este estudo tem por objetivo avaliar com base em evidências a efetividade do planejamento urbano em direcionar a produção imobiliária em São Paulo, utilizando os dados do Imposto Territorial e Predial Urbano e métodos bayesianos. Verifica-se que o adensamento construtivo ocorrido na cidade no período 2014-2023 localizou-se predominantemente no quadrante sudoeste da cidade, onde se concentra a população branca, com maior renda e que possui o menor tempo médio de deslocamento em viagens a trabalho no município. O estudo conclui que tal concentração na “área nobre” da cidade, onde o valor agregado do metro quadrado é maior, associada à inefetividade dos incentivos construtivos na periferia, resulta na perpetuação das estruturas de segregação espacial que o Plano Diretor objetivava transformar.

Palavras-chave: planejamento urbano, mercado imobiliário, estatística bayesiana.

ABSTRACT

Contextualization: Promoting the expansion of the right to the city through increased construction density near transportation corridors is one of the main objectives of São Paulo's Strategic Master Plan. However, the success of this policy is not solely dependent on the decisions made by the public authorities; it requires the engagement of the real estate sector, which is responsible for producing the developments that will drive the desired growth. This study aims to evaluate, based on evidence, the effectiveness of urban planning in directing real estate development in São Paulo, using data from the Urban Property Tax (IPTU) and Bayesian methods. The findings indicate that, between 2014 and 2023, the densification occurred predominantly in the southwestern quadrant of the city, where there is a concentration of a predominantly white population, with higher incomes and shorter average commute times to work. The study concludes that this concentration in the 'noble area' of the city, where the value per square meter is higher, combined with the ineffectiveness of construction incentives in the periphery, perpetuates the spatial segregation structures that the Master Plan aimed to transform.

Keywords: urban planning, real state market, bayesian statistics.

SUMÁRIO

Lista de Ilustrações	11
Lista de Tabelas	14
Lista de Quadros	16
Lista de Abreviaturas	18
Lista de Símbolos	22
1 Introdução	23
1.1 Organização.....	27
2 Planejamento urbano e densidade	29
2.1 Introdução.....	29
2.2 Acessibilidade e segregação	31
2.3 O Modelo Monocêntrico Padrão.....	35
3 Efetividade do Plano Diretor	44
3.1 Introdução.....	44
3.1.1 Objetivo.....	46
3.1.2 Organização	46
3.2 Materiais e Métodos.....	48
3.2.1 Unidades verticais produzidas entre 2014 e 2023.....	48
3.2.2 Quadras em Zona de Estruturação Urbana.....	50
3.2.3 Fatores associados à verticalização	54
3.2.4 Restrições construtivas	56
3.2.5 Identificação das áreas nobre e periféricas da cidade .	58
3.2.5.1 Análise de Componentes Principais	64
3.2.5.2 Indicadores Locais de Associação Espacial .	65
3.2.6 Resumo da base de dados	71
3.2.7 Modelos estatísticos.....	72
3.2.7.1 Tipificação da variável dependente	73
3.2.7.2 Poisson.....	76
3.2.7.3 Binomial Negativa	80

3.2.7.4	Zeros Inflados	82
3.2.7.5	Modelo Multinível	90
3.3	Resultados	102
3.3.1	Segregação social e agrupamentos espaciais.....	103
3.3.2	Produção imobiliária e modelos de contagem	115
3.3.3	Empreendimentos de grande porte e superdispersão .	122
3.3.4	Restrições construtivas e zeros inflados	129
3.3.5	Desigualdade e modelos multinível	146
3.4	Considerações Finais	162
4	Retorno ao Modelo Monocêntrico	165
4.1	Introdução.....	165
4.1.1	Objetivos	166
4.1.2	Organização	167
4.2	Materiais e Métodos.....	167
4.2.1	Materiais	167
4.2.2	Métodos	169
4.3	Resultados	170
4.4	Considerações Finais	176
5	Conclusão	180
	Referências Bibliográficas	183
	Anexos	195
	Anexo Único - Origem e Tratamento dos Dados	196
A	Introdução.....	196
B	Tecnologias utilizadas.....	197
C	Fontes de dados	199
C.1	Dados produzidos pela Prefeitura de São Paulo.....	199
C.1.1	Cadastro Imobiliário Fiscal.....	200
C.1.2	GeoSampa	204
C.2	Dados produzidos por outros órgãos públicos.....	212
C.2.1	Censo IBGE.....	212
C.2.2	Pesquisa Origem e Destino.....	213

	C.2.3	RAIS	214
D		Engenharia de atributos.....	215
	D.1	Densidade construtiva	215
		D.1.1 Identificação da produção imobiliária	216
		D.1.2 Quantidade de unidades produzidas.....	227
		D.1.3 Quantidade de metros quadrados produzidos	229
	D.2	Incentivos Construtivos	231
	D.3	Fatores associados à verticalização.....	235
	D.4	Restrições à edificação.....	239
	D.5	Áreas edificáveis	243
	D.6	Privilégio social.....	248
	D.7	Acessibilidade ao Mercado de Trabalho	252
E		Descrição das bases de dados.....	259
	E.1	Base de dados - adensamento medido por unidades produzidas	260
	E.2	Base de dados - adensamento medido por metros qua- drados de área construída	268

LISTA DE ILUSTRAÇÕES

1.1 Eixos de Estruturação da Transformação Urbana.	24
2.1 Ilustração para o modelo SMM.	36
2.2 Densidade construtiva de tipo residencial vertical em São Paulo (2023).	41
2.3 Simulação dos efeitos do aumento dos custos de transporte com base nos parâmetros do SMM.	42
3.1 Identificação das unidades verticais produzidas (2014-2023).	49
3.2 Distribuição dos EETU no território municipal.	52
3.3 Identificação das quadras em ZEU.	53
3.4 Imputação de dados faltantes para quadras criadas no período 2014-2023.	56
3.5 Anúncio de empreendimento <i>exclusivo</i>	59
3.6 Propaganda com a localização do empreendimento.	60
3.7 Setor sudoeste - área de residência da elite.	62
3.8 Gráfico de dispersão de Moran – exemplo.	68
3.9 Identificação dos <i>clusters</i> espaciais da área nobre e das áreas periféricas.	70
3.10 Exemplo: lojas de sorvete.	78
3.11 Paradoxo de Simpson.	93
3.12 Densidade de probabilidade conjunta em formato de funil.	100
3.13 Distribuição espacial das dimensões da segregação.	104
3.14 Matriz de correlação – segregação social.	106
3.15 Gráfico de declividade – PCA.	108
3.16 Distribuição Espacial da Variável Latente.	109
3.17 Correlação espacial global – I de Moran.	111
3.18 Gráfico de Dispersão de Moran.	112
3.19 Agrupamentos espaciais.	113
3.20 Distribuição <i>a priori</i> – Normal(0, 2.5 ²)	116
3.21 Traceplot - Poisson.	119
3.22 Empreendimento de grande porte.	123
3.23 Distribuição <i>a priori</i> – Gamma(0.01, 0.01).	124

3.24 Traceplot – Binomial Negativa.....	127
3.25 Localização dos Eixos de Estruturação da Transformação Urbana	129
3.26 Distribuição da área das quadras em ZEU	130
3.27 Distribuição espacial da área verticalizável.....	131
3.28 Valor nominal do metro quadrado de terreno	133
3.29 Histograma – Quantidade de unidades verticais produzidas por quadra 2014-2023	136
3.30 Média da proporção da área construída vertical por região da cidade	138
3.31 Distribuição dos dados simulados - Zeros Inflados	140
3.32 Traceplot ZIP - Contagem	142
3.33 Traceplot ZIP - Estrutural	143
3.34 Quadras mais próximas às quadras em ZEU	148
3.35 Traceplot - Modelo Multinível.....	151
4.1 Área verticalizável em São Paulo.	171
4.2 Distribuição da acessibilidade ao mercado de trabalho em 2007, sua variação no período 2007-2017 e do adensamento construtivo no mesmo período em São Paulo, por distrito	172
4.3 Clusterização Espacial da acessibilidade ao mercado de trabalho e adensamento construtivo em São Paulo	173
1 Representação da Cartela do IPTU.....	202
2 Plataforma GeoSampa.....	205
3 Lotes cadastrados na favela de Paraisópolis.....	219
4 Lotes cadastrados no Pacaembu.....	220
5 Correspondência entre quadras fiscais e quarteirões.	222
6 Quadra com dois polígonos - COPAN.	223
7 Unificação e desdobro de lotes em empreendimento vertical.	224
8 Estacionamento transformado em condomínio vertical - r. Frei Caneca (2004-2020).	226
9 Analogia – moldes de biscoito.....	246
10 Matriz de Correlação de Pearson - Restrições Construtivas	268
11 Matriz de Correlação de Pearson - Potencial de Verticalização	269
12 Matriz de Correlação de Pearson - Tratamentos e índices espaciais	270
13 Matriz de Correlação de Pearson - Variáveis Primárias	274

14	Matriz de Correlação de Pearson - Variáveis Construídas	275
----	---	-----

LISTA DE TABELAS

3.1	Resultados da simulação - Poisson	118
3.2	Descrição da tabela.....	119
3.3	Resultados da simulação - Binomial Negativa.....	125
3.4	Comparação com Dados Simulados – Poisson e Binomial Negativa.....	126
3.5	Comparação com Dados Reais – Poisson e Binomial Negativa.....	126
3.6	Descrição da tabela.....	127
3.7	Descrição da tabela.....	141
3.8	Descrição da tabela.....	142
3.9	Descrição da tabela.....	143
3.10	Comparação – Binomial Negativa e Zeros Inflados	146
3.11	Processo de contagem - Área Classe Média	152
3.12	Processo de zeros estruturais - Área Classe Média.....	152
3.13	Processo de contagem - Área Periférica	154
3.14	Processo de zeros estruturais - Área Periférica	154
3.15	Processo de contagem - Área Nobre.....	155
3.16	Processo de zeros estruturais - Área Nobre.....	156
3.17	Processo de contagem - Área Classe Média	158
3.18	Processo de zeros estruturais - Área Classe Média.....	158
3.19	Processo de contagem - Área Periférica	159
3.20	Processo de zeros estruturais - Área Periférica	160
3.21	Processo de contagem - Área Nobre.....	161
3.22	Processo de zeros estruturais - Área Nobre.....	161
4.1	Coeficientes – Modelo 1	174
4.2	Coeficientes – Modelo 2	175
4.3	Coeficientes – Modelo 3	176
1	Descrição da Base de Dados - Restrições à edificação por quadra da Cidade de São Paulo.	265
2	Descrição da Base de Dados - Potencial de verticalização por quadra da Cidade de São Paulo.....	266
3	Descrição da Base de Dados - Potencial de verticalização por quadra da Cidade de São Paulo.....	267

4	Descrição da Base de Dados - Variáveis Primárias e Dependente.	272
5	Descrição da Base de Dados - Variáveis Primárias e Dependente.	273

LISTA DE QUADROS

3.1	Variáveis utilizadas na modelagem.....	71
3.2	Exemplo: base de dados de lojas de sorvete.....	79
3.3	Exemplo: base de dados de lojas de sorvete.....	80
3.4	Exemplo: base de dados de lojas de sorvete.....	82
3.5	Exemplo: paradoxo de Simpson.....	92
3.6	Exemplo: paradoxo de Simpson.....	92
3.7	Análise de Componentes Principais.....	107
3.8	Pesos do Componente Principal.....	108
3.9	Variável Latente – maiores valores.....	110
3.10	Variável Latente – menores valores.....	110
4.1	Estatísticas – Modelo 1	174
4.2	Estatísticas – Modelo 2	175
4.3	Estatísticas – Modelo 3	176
1	<i>Stack</i> de tecnologias.....	198
2	Georreferenciamento – Prefeitura de São Paulo.....	206
3	Georreferenciamento – Outros Órgãos Públicos.....	206
4	Legislação Urbana – Variáveis Tratamento.....	207
5	Restrições relacionadas a Tombamentos.....	209
6	Restrições Ambientais	210
7	Restrições Construtivas Diversas.....	211
8	Unidades por Quadra	228
9	Área construída por Quadra.....	231
10	Área edificável por distrito.....	245
11	Área saturada por distrito	248
12	Média do tempo de desloc. a trabalho por distrito (minutos).....	252
13	Viagens pendulares - distritos de origem e destino.....	254
14	Média tempo desloc. - dist. origem e destino	254
15	Quantidade de vínculos empregatícios por distrito.....	255
16	Percentual de Postos de Trabalho Acessíveis em até 60 minutos .	257
17	BD Unidades – índices e <i>dummies</i>	261
18	BD Unidades – restrições	262

19	BD Unidades – potencial construtivo	263
20	BD Área Construída – Variáveis primárias	270
21	BD Área Construída – Variáveis construídas	271

LISTA DE ABREVIATURAS

Organizações

CASE Coordenadoria de Cadastro do Sistema Eletrônico de Licenciamento - SMUL

CONDEPHAAT Conselho de Defesa do Patrimônio Histórico, Arqueológico, Artístico e Turístico do Estado de São Paulo

CONPRESP Conselho Municipal de Preservação do Patrimônio Histórico, Cultural e Ambiental da Cidade de São Paulo

CPA Coordenadoria de Planejamento Ambiental - SVMA

DPH Departamento do Patrimônio Histórico

DEUSO Coordenadoria de Legislação, Uso e Ocupação do Solo - SMUL

DESAP Departamento de Desapropriações - SMJ

GEOINFO Coordenadoria de Análise e Produção da Informação - SMUL

IBGE Instituto Brasileiro de Geografia e Estatística

IPHAN Instituto do Patrimônio Histórico e Artístico Nacional

Metro-SP Companhia do Metropolitano do Estado de São Paulo

MTE Ministério do Trabalho e Emprego - Governo Federal

PMSP Prefeitura Municipal de São Paulo

SEHAB Secretaria de Habitação

SF Secretaria Municipal da Fazenda

SMC Secretaria Municipal de Cultura

SMJ Secretaria Municipal de Justiça

SMSU Secretaria Municipal de Segurança Urbana

SMUL Secretaria Municipal de Urbanismo e Licenciamento

SVMA Secretaria Municipal do Verde e Meio Ambiente

Conceitos e legislação associados ao urbanismo

APA Área de Proteção Ambiental

APP Área de Preservação Permanente

BDT Boletim de Dados Técnicos

CA Coeficiente de Aproveitamento

CBD Distrito Comercial Central (do inglês, *Central Business District*)

DIS Decreto de Interesse Social

DTCO Declaração Tributária de Conclusão de Obra

DUP Decreto de Utilidade Pública

FAR Proporção de Área Útil (do inglês, *Floor Area Ratio*)

FUNDURB Fundo Municipal de Desenvolvimento Urbano

ISS Imposto Sobre Serviços

IPTU Imposto Territorial e Predial Urbano

LPUOS Lei de Parcelamento, Uso e Ocupação do Solo

EETU Eixo de Transformação da Estruturação Urbana

OODC Outorga Onerosa do Direito de Construir

PDE Plano Diretor Estratégico

RPPN Reserva Particular do Patrimônio Natural

S-Q-L Setor-Quadra-Lote

TOD Desenvolvimento Orientado ao Transporte Público (do inglês, *Transit Oriented Development*)

UC Unidade de Conservação

ZC Zona de Centralidade

ZCOR Zona de Corredor

ZER Zona Exclusivamente Residencial

ZEIS Zona Especial de Interesse Social

ZEU Zona de Estruturação Urbana

ZEPAM Zona Especial de Preservação Ambiental

ZM Zona Mista

Programação e estatística

GLM Modelos Lineares Generalizados (do inglês, *Generalized Linear Models*)

AGPL Licença livre e aberta (do inglês, *Affero General Public Licensing*)

ETL Extração, Transformação e Carregamento (do inglês *Extract, Transform and Load*)

FTP Protocolo de transferência de arquivos (do inglês, *File Transfer Protocol*)

LGPD Lei Geral de Proteção de Dados Pessoais

LISA Indicadores Locais de Associação Espacial (do inglês, *Local Indicators of Spatial Association*)

NUTS Amostrador sem curvas U (do inglês, *No U-turn Sampler*)

SMM Modelo Monocêntrico Padrão (do inglês, *Standard Monocentric Model*)

PCA Análise de Componentes Principais (do inglês, *Principal Component Analysis*)

IID Independentes e Igualmente Distribuídos

RDD Regression Discontinuity Design

MSE Erro Quadrático Médio (do inglês, *Mean Squared Error*)

ZIP Modelo de Poisson para Zeros-Inflados (do inglês, *Zero-Inflated Poisson*)

Fontes de dados

OD Pesquisa Origem e Destino

RAIS Relação Anual de Informações Sociais

Outros

CEU Centro Educacional Unificado

LIDAR Detecção e Alcance de Luz (do inglês, *Light Detection and Ranging*)

LISTA DE SÍMBOLOS

CONCEITOS BÁSICOS

\mathbb{Z} Conjunto dos números inteiros

\mathbb{N} Conjunto dos números naturais

\mathbb{R} Conjunto dos números reais

IMAGENS

f e g Variáveis que representam imagens

$p, q \in \mathbb{D}$ Variáveis que representam pares (x, y) aqui chamados de *pixels*

1 INTRODUÇÃO

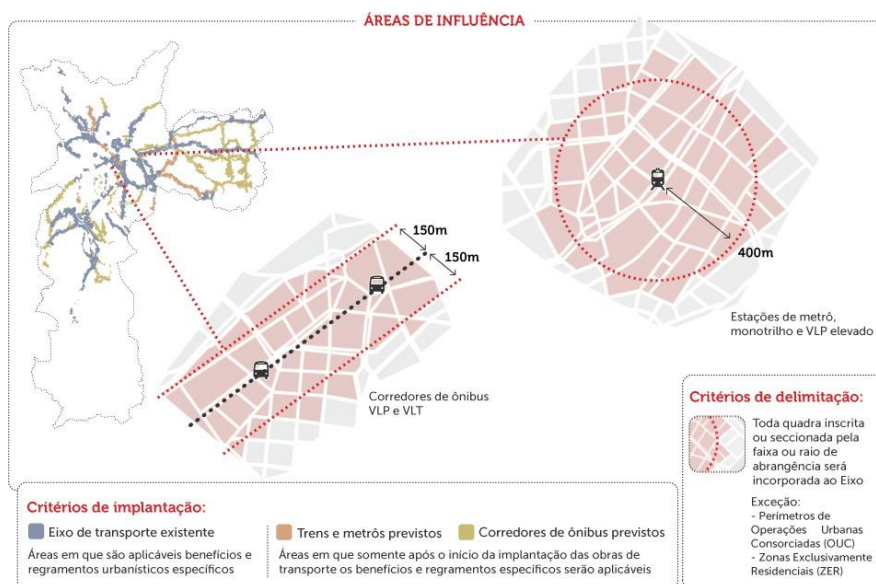
Nas últimas décadas, diversos movimentos do urbanismo se voltaram aos temas da densidade e da proximidade entre os locais de trabalho e moradia como formas de combater os efeitos adversos do processo de espraiamento da mancha urbana. Entretanto, muitos desses movimentos não esclarecem como o adensamento proposto será efetivamente alcançado, frequentemente recorrendo ao mercado para produzir os empreendimentos imobiliários que promoverão a transformação urbana pretendida (KHAVARIAN-GARMSIR et al., 2023).

Inspirado por um desses movimentos, o Plano Diretor Estratégico (PDE) da Cidade de São Paulo estabeleceu o objetivo de promover o adensamento nas áreas próximas aos terminais de transporte público da cidade, denominadas Eixo de Transformação da Estruturação Urbana (EETU) (FUNARI, 2016). Para isso, o PDE instituiu uma série de incentivos à produção imobiliária nesses locais.

Implementada em 2014, essa política se apresenta como um objeto de pesquisa ideal para estudos que visam avaliar a efetividade do planejamento urbano em orientar a produção imobiliária. Isso se deve ao fato de que os eixos estão distribuídos pelo território da cidade, abrangendo regiões com diferentes características socioeconômicas e distintos padrões de uso e ocupação do solo, como ilustra a Figura 1.1.

Este estudo tem por objetivo avaliar, com base em evidências, a efetividade do Plano Diretor em direcionar a produção imobiliária em São Paulo para os Eixos de Transformação da Estruturação Urbana no período compreendido entre os anos de 2014 e 2023.

Utilizando técnicas de Ciência de Dados Espaciais, foram analisadas diversas bases de dados públicas e abertas para identificar, de forma territorializada, o adensamento construtivo ocorrido na cidade durante o período em questão. A esses dados, foram adicionadas informações referentes às várias restrições construtivas atualmente vigentes, estabelecidas pela legislação urbanística e por políticas de preservação cultural e ambiental. Além de informações associadas ao potencial construtivo, como o valor da terra, a obsolescência do estoque construído, a disponibilidade de terrenos e a saturação por edificações pré-existentes.

Figura 1.1 – Eixos de Estruturação da Transformação Urbana.

Fonte: Prefeitura Municipal de São Paulo -

<https://gestaourbana.prefeitura.sp.gov.br/novo-pde-eixos-de-estruturacao-da-transformacao-urbana/>

A base de dados resultante registra a quantidade de unidades produzidas por empreendimentos verticais para cada uma das mais de 46 mil quadras do município de São Paulo, bem como a área construída total dessas unidades. Além disso, contém informações associadas à presença de restrições à edificação e ao potencial construtivo em cada uma dessas quadras. Esses dados foram combinados a uma análise das características socioeconômicas e de acessibilidade ao mercado de trabalho dos 96 distritos do município que permitiu identificar padrões espaciais de segregação urbana na cidade.

Os modelos estatísticos desenvolvidos têm por objetivo analisar se a localização de uma quadra da cidade de São Paulo em Zona de Eixo de Estruturação Urbana (ZEU) acarretou efeito positivo significativo sobre a quantidade esperada de unidades de tipo vertical produzidas no período de vigência do Plano Diretor (2014-2023).

O trabalho de análise seguiu o fluxo de trabalho *bayesiano*, segundo o qual são ajustados e comparados modelos estatísticos progressivamente

mais complexos com o objetivo de aproximá-los sucessivamente de uma modelagem mais adequada ao problema de pesquisa (GELMAN et al., 2020). O modelo final, um Modelo Multinível com Regressão Binomial Negativa para Zeros Inflados, fornece evidências de que os incentivos construtivos estabelecidos pelo Plano Diretor nas ZEU tiveram efeitos significativos apenas nas áreas de classe média e, especialmente, na área nobre da cidade.

Essa última é a área de residência da elite, onde se concentram os distritos com a maior proporção de pessoas brancas, ricas e que tiveram acesso à educação básica. As pessoas que nela residem têm os menores tempos de deslocamento a trabalho. Os resultados dos modelos indicam que essa região possui uma maior propensão a receber empreendimentos imobiliários, independentemente dos incentivos construtivos fornecidos.

Nas áreas periféricas, onde há maior vulnerabilidade social e os residentes enfrentam os maiores tempos de deslocamento para o trabalho, os incentivos construtivos estabelecidos pelos Eixos de Estruturação da Transformação Urbana não tiveram efeito significativo. Dessa forma, questiona-se se o objetivo de transformação urbana estabelecido pelo Plano Diretor foi realmente alcançado, ou se, na prática, houve um reforço das estruturas de segregação espacial já existentes.

Com o objetivo de compreender melhor os mecanismos mais amplos que estão associados ao adensamento urbano, foi realizado um estudo final com base no Modelo Monocêntrico. Trata-se de um modelo de equilíbrio teórico formulado pela economia urbana na década de 1960 que argumenta que há uma lógica de mercado que molda a estrutura urbana. Segundo esse modelo, há maior demanda pelos terrenos mais próximos ao centro de negócios da cidade devido aos menores tempos de deslocamento pendular de quem reside neles, o que resulta no adensamento dessas áreas.

O modelo estatístico desenvolvido neste estudo – uma Regressão Linear Múltipla – teve por objetivo analisar se um aumento na acessibilidade ao mercado de trabalho em um distrito da cidade de São Paulo esteve associado a um aumento na área construída de tipo residencial vertical produzida nesse distrito no período entre os anos de 2007 e 2017¹.

¹Como será detalhado no Anexo Único a escolha desse período, que difere do período de análise do primeiro estudo, se deve à disponibilidade dos dados da Pesquisa Origem e Destino que foram utilizados para identificar alterações nos tempos de deslocamento a trabalho na cidade.

Os resultados do modelo assinalam um efeito positivo substancial e altamente significativo ($R^2 = 0,57$ e $p < 0,001$) fornecendo evidências para corroborar a relação entre acessibilidade e adensamento que está no cerne da contribuição teórica do Modelo Monocêntrico (DURANTON; PUGA, 2015). Esses resultados sugerem que políticas públicas voltadas para a redução dos tempos de deslocamento, ou para a descentralização dos postos de trabalho, podem ser efetivas na promoção do adensamento urbano.

Em conjunto, os resultados dessa dissertação demonstram a importância de que as políticas públicas urbanas não estejam alicerçadas exclusivamente no conhecimento e experiências pessoais dos planejadores urbanos sobre a cidade, mas sim sejam baseadas em evidências sobre as dinâmicas socioeconômicas mais amplas que afetam a estrutura urbana e que só podem ser verificadas em grande escala.

Por fim, este trabalho evidencia a necessidade de que se realizem estudos posteriores sobre o impacto das restrições construtivas atualmente impostas pela legislação urbanística para impedir a verticalização das áreas de residência da elite da cidade. Tais restrições produzem artificialmente um adensamento construtivo baixo em áreas bem localizadas, com elevado acesso à infraestrutura urbana e alta acessibilidade ao mercado de trabalho. Essas áreas possuem forte propensão à verticalização, que é no entanto impedida por essas restrições.

Esse cenário pode acarretar aumentos de custos de moradia e/ou maiores tempos de deslocamento àqueles que não têm o privilégio de residir nesses locais, seja devido à escassez de imóveis gerada pelos impedimentos à verticalização, seja devido aos altos custos por metro quadrado que resultam dessa escassez. Haveria, assim, uma terceira via de intervenção em políticas públicas urbanísticas com alto impacto na redução de desigualdades no acesso à cidade e baixo custo de investimento público, que no entanto não é considerada pelo atual Plano Diretor do município de São Paulo. Trata-se da substituição dos zoneamentos de tipo residencial horizontal (conhecidos como *single family detached housing*) nas áreas centrais da cidade por zonas mistas de média a alta densidade.

1.1 ORGANIZAÇÃO

Esta dissertação está organizada da seguinte forma:

- O Capítulo 2 apresenta as bases teóricas do estudo por meio de uma revisão bibliográfica, abordando como o tema da densidade permeia as principais correntes do planejamento urbano. É identificada uma lacuna comum a essas correntes: a ausência de um modelo causal claro para avaliar a efetividade das intervenções planejadas na promoção do adensamento, o que aumenta o risco de efeitos adversos, como a gentrificação. O capítulo conclui com a retomada do Modelo Monocêntrico Padrão (SMM) como alternativa para entender as dinâmicas urbanas e avaliar a capacidade do planejamento em direcionar a produção imobiliária.
- O Capítulo 3 analisa a capacidade do Plano Diretor Estratégico de São Paulo em direcionar a produção imobiliária, focando no impacto dos incentivos construtivos em áreas próximas a terminais de transporte público (EETU) entre 2014 e 2023. Utilizando um modelo estatístico bayesiano multinível para zeros inflados, o estudo considera fatores que promovem e restringem a verticalização, como disponibilidade de terrenos e áreas de proteção ambiental. Os resultados indicam que os incentivos foram eficazes nas áreas nobres e de classe média, mas não tiveram impacto nas periferias, perpetuando as estruturas de segregação existentes.
- O Capítulo 4 examina a relação entre o adensamento construtivo residencial vertical e a acessibilidade ao mercado de trabalho em São Paulo (2007-2017). Utilizando um modelo estatístico com alta capacidade explicativa ($R^2 = 0.57$), revela uma forte associação positiva entre acessibilidade e adensamento. O capítulo contribui para a literatura empírica sobre a relação entre acessibilidade e estrutura urbana ao fornecer evidências que sustentam o Modelo Monocêntrico (SMM). Seus resultados permitem avaliar criticamente propostas recentes do urbanismo que retomam essa relação, mas negligenciam dinâmicas socioespaciais mais amplas, aumentando o risco de segregação e gentrificação.

- O Capítulo 5 resume e discute os resultados alcançados nos capítulos anteriores, concluindo sobre a necessidade de que o planejamento urbano seja baseado em evidências sobre as dinâmicas socioespaciais mais amplas que regem a estrutura urbana. Assinala também a necessidade de que se realizem estudos futuros sobre os impactos negativos das restrições à verticalização atualmente impostas sobre regiões da cidade com elevado acesso à infraestrutura urbana e alta acessibilidade ao mercado de trabalho.
- O Anexo Único é voltado ao leitor que tem por objetivo replicar os resultados deste estudo ou que busca utilizar as bases de dados aqui produzidas em estudos futuros. O anexo inicia-se com uma visão geral das tecnologias utilizadas, seguida pela listagem e descrição das fontes de dados. Em seguida, são detalhados os pipelines de extração e transformação de dados desenvolvidos, assim como os processos de *feature engineering* aplicados a cada variável das bases de dados finais. O anexo é concluído com uma descrição das bases resultantes.

2 PLANEJAMENTO URBANO E DENSIDADE

Resumo do capítulo

Este capítulo apresenta as bases teóricas do estudo. A partir de uma revisão bibliográfica, ele mostra como o tema da densidade está presente nas principais correntes de planejamento urbano, incluindo o movimento pela Cidade de 15 Minutos, que ganhou visibilidade nos últimos anos. No entanto, é identificada uma lacuna comum: a falta de um modelo causal claro que permita avaliar a efetividade das intervenções planejadas, especialmente quanto à capacidade de promover o adensamento. Isso resulta em um aumento do risco de efeitos adversos dessas intervenções, como a gentrificação. O capítulo conclui com a retomada do Modelo Monocêntrico Padrão (SMM), que associa a densidade urbana à acessibilidade ao mercado de trabalho. Esse modelo é proposto como alternativa para compreender as dinâmicas urbanas e avaliar a capacidade do planejamento em direcionar a produção imobiliária.

Objetivo: Apresentar a fundamentação teórica do estudo.

2.1 INTRODUÇÃO

A baixa densidade e a distância elevada entre os espaços de moradia e trabalho, possibilitadas pela popularização do automóvel no século XX, são problemas que afetam muitas cidades no mundo. Em um estudo longitudinal que envolveu 1100 áreas urbanas em 29 países, a OCDE constata que na maioria delas houve um aumento do espraiamento, agravando a dependência do automóvel (OECD, 2018). Segundo a organização, esse padrão de desenvolvimento urbano leva a maiores tempos de deslocamento, congestionamentos, poluição do ar e aumento das emissões de gases de efeito estufa. Além de elevar os custos de provisão de serviços públicos e tornar as cidades mais segregadas (OECD, 2018, p. 11).

Diversos movimentos do urbanismo nas últimas décadas se voltaram a esse problema, tais como o design orientado ao transporte (CALTHORPE, 1993) e o novo urbanismo (ANDERSON, 2017). Mais recentemente, du-

rante a pandemia de Covid-19, com as restrições de circulação impostas pelas medidas sanitárias para conter o vírus, os impactos negativos da separação entre os espaços de moradia e trabalho se acentuaram

Nesse contexto, a prefeitura de Paris decidiu colocar em prática a proposta da cidade de 15 minutos, que defende uma cidade na qual as principais necessidades humanas – como a moradia, o trabalho, o entretenimento, a alimentação, a educação e a saúde – estão acessíveis a um deslocamento de menos de 15 minutos a pé ou de bicicleta do local de residência (MORENO et al., 2021).

Esse modelo tem sido proposto para a cidade de São Paulo pela candidatura de Guilherme Boulos, da coligação PT-PSOL, que visitou Paris durante a pré-campanha e identificou a cidade como um exemplo de boa prática. Segundo o candidato em um discurso à câmara dos deputados, o modelo de cidade de 15 minutos em Paris aponta para a necessidade de aproximar os polos de residência, trabalho, cultura e lazer em São Paulo (BOULOS, Guilherme, 2023).

Esse ponto de vista, no entanto, já havia sido incorporado no último governo do PT ao principal instrumento de planejamento urbano da cidade, o Plano Diretor (PDE). Com efeito, os Eixos de Estruturação da Transformação Urbana (EETU), propostos no PDE, consistem em áreas da cidade próximas à rede estrutural de transportes coletivos (metrô, trens e terminais de ônibus) nas quais o PDE planeja promover o adensamento construtivo e habitacional. Assim como incentivar os usos mistos e a fruição pública dos térreos dos edifícios, aproximando os locais de moradia e de trabalho e, com isso, desincentivando os deslocamentos por automóvel¹.

Balbin e Krause (2016, 12), em uma avaliação *ex-ante* dessa política, afirmam que ela foi inspirada no modelo de Desenvolvimento Orientado ao Transporte (TOD), inicialmente proposto por Calthorpe (1993). Esse modelo possui diversos pontos em comum com o modelo de Cidade de 15 minutos, especialmente no que diz respeito à defesa do adensamento e da proximidade entre residência e emprego como forma de combater o espraiamento (KHAVARIAN-GARMSIR et al., 2023).

Contudo, na medida em que essas políticas não se baseiam no Estado

¹Isso se manifesta inclusive na redução da quantidade mínima de vagas por unidade.

como produtor de unidades habitacionais², escapando portanto da ação direta do poder público, sua eficácia passa a ser determinada por fatores externos, sobretudo pela adesão do mercado imobiliário. Pois caberá ao mercado produzir o adensamento construtivo almejado pelo planejamento urbano, gerando a transformação urbana nas áreas por ele designadas.

Nesse cenário, em que a efetividade do planejamento urbano é mediada por sua capacidade de direcionar a produção imobiliária, faz-se necessário mobilizar um arcabouço teórico definido que possa ser operacionalizado para mensurar se o adensamento eventualmente ocorrido (ou não) foi de fato causado pela política urbana, como argumentam DeSalvo e Su (2018)³. O presente capítulo tem por objetivo apresentar esse arcabouço teórico, partindo de uma avaliação crítica dos movimentos mais recentes do planejamento urbano.

2.2 ACESSIBILIDADE E SEGREGAÇÃO

A proposta da Cidade de 15 Minutos tem muitos pontos em comum com as correntes do planejamento urbano que a antecederam (KHAVARIAN-GARMSIR et al., 2023). Especificamente, a densidade, os usos mistos e a proximidade entre moradia e trabalho são princípios compartilhados por diversos movimentos do urbanismo a partir da década de 1980, dentre os quais o design orientado ao transporte e o novo urbanismo (KHAVARIAN-GARMSIR et al., 2023). A Cidade de 15 Minutos aborda esses temas ao propor cidades hiper-policêntricas, nas quais cada vizinhança deverá se adensar e se tornar uma centralidade em si mesma, promovendo a proximidade entre moradia e trabalho a partir dos usos mistos e eliminando a necessidade de deslocamentos que atravessem a cidade.

Tal foco nos deslocamentos de até 15 minutos causou controvérsia ao ser alvo de protestos de parte do público leigo, que considerou a medida autoritária por entender tratar-se de um impedimento à livre circulação (HSU, 2023). Na academia, autores questionaram a proposta por subesti-

²Há inclusive uma crítica aos conjuntos habitacionais produzidos pelo Estado sob a perspectiva do urbanismo moderno, em geral exclusivamente residenciais e situados em áreas periféricas da cidade, como é o caso, em São Paulo, da Cidade Tiradentes.

³Funari (2016) apresenta questionamento similar para o caso específico dos EETU definidos no Plano Diretor de São Paulo

mar o esforço necessário para alterar os padrões socioterritoriais vigentes e por conter riscos de efeitos adversos.

Casarin, MacLeavy e Manley (2023) argumentam que, se por um lado a promoção de usos mistos defendida pelo movimento da Cidade de 15 Minutos pode ser um benefício às vizinhanças de baixa renda, historicamente desprovidas de serviços, comércio e equipamentos públicos, por outro lado a proposta tem o risco de resultar em processos de gentrificação. O sucesso das vizinhanças de 15 minutos pode acarretar um aumento no valor dos aluguéis e, com isso, a expulsão de seus residentes originais para localidades mais distantes, agravando a segregação e aumentando seu tempo de deslocamento. Segundo os autores, para que isso não ocorra é necessária uma abordagem que não se concentre apenas na escala local e que defenda o direito à cidade como um todo (CASARIN; MACLEAVY; MANLEY, 2023, 3180).

Um estudo realizado por Freemark (2020) com base em dados do mercado imobiliário na cidade de Chicago mostra que o efeito direto de políticas de adensamento por meio da alteração dos parâmetros construtivos do zoneamento pode acarretar um resultado similar. Os autores argumentam que as políticas de *upzoning*⁴ na cidade tiveram como efeito um rápido aumento nos valores das transações de propriedade ocorridas, inclusive para as unidades já existentes anteriormente ao momento de publicização da alteração legislativa que estabeleceu o *upzoning*.

Rodríguez-Pose e Storper (2020) argumentam que, quando realizado sem o acompanhamento de outras políticas urbanas, o *upzoning* reforça os efeitos da desigualdade de renda não apenas na medida em que o novo potencial construtivo passa a ser contabilizado nos valores dos imóveis, tornando-os mais valorizados. Pois, quando esse potencial de fato se efetiva por meio da construção de empreendimentos verticais, isso ocorre em vizinhanças que já eram alvo do interesse do mercado imobiliário. As novas unidades produzidas, que passam a substituir as residências antigas, em geral são voltadas a pessoas de maior renda, possuindo um valor por metro quadrado com os quais os antigos moradores não são capazes de arcar. Segundo os autores, esse processo resulta em uma redução da oferta da

⁴Termo utilizado para denotar o aumento do potencial construtivo permitido em uma determinada zona da cidade.

moradia para pessoas de baixa renda nesses vizinhanças, o que configura gentrificação (RODRÍGUEZ-POSE; STORPER, 2020, 18).

Com base em dados massivos, obtidos a partir dos GPS dos aparelhos móveis de cidadãos de mais de 400 aglomerações urbanas nos Estados Unidos em 2019, Abbiasov et al. (2023) estimam que o norte-americano mediano realiza apenas 14% de suas viagens orientadas para consumo em até 15 minutos. Dado que demonstra o quão distante o comportamento atual está do ideal proposto por (MORENO et al., 2021), o que pode indicar não apenas uma preferência individual mas uma questão estrutural da forma como as pessoas se relacionam com o espaço urbano.

Os autores identificam ainda que um maior percentual de viagens de 15 minutos implica uma redução significativa da quantidade esperada de interações sociais. Segundo eles, as pessoas de baixa-renda são especialmente suscetíveis a esse efeito, pois possuem uma tendência maior a realizar viagens para consumo apenas em sua vizinhança em comparação a pessoas das classes média e alta. Os pontos comerciais das vizinhanças de baixa renda, geralmente localizadas em áreas periféricas, tendem a ser visitados apenas por pessoas dessa classe social. Esse padrão contrasta com os pontos comerciais localizados em regiões centrais, cuja frequência é mais socialmente diversa.

A realização de viagens em apenas 15 minutos acarreta, assim, segundo os autores, em um risco de “intensificação da segregação socioeconômica” (ABBIASOV et al., 2023, 15). Como argumentam (MANI; RILEY, 2021) a diversidade e extensão das redes sociais de que as pessoas participam está relacionada à ascensão social, pois é por meio destas redes que oportunidades de emprego e outros recursos são obtidos e compartilhados. Redes sociais reduzidas ou pouco-diversas, especialmente se restritas apenas a pessoas da mesma classe social, tendem a diminuir as chances de ascensão.

Assim, devido ao risco de reduzir as interações sociais de pessoas de baixa renda com pessoas de outras classes sociais, em especial pessoas de maior renda, a proposta da Cidade de 15 Minutos pode limitar a mobilidade social positiva nas cidades, resultando em um aumento da segregação socioeconômica. Além de ir de encontro a um dos principais motivos que faz as pessoas migrarem para cidades: a busca por oportunidades.

Bertaud (2022a) analisa a localização de estabelecimentos em Paris

e conclui que a cidade já pode ser considerada uma cidade de 15 minutos antes mesmo de qualquer intervenção do poder público. No entanto, dados de viagens pendulares mostram que apenas 12% dos parisienses levam menos de 15 minutos para ir de casa ao trabalho. Isso não é causado por uma falta de empregos, pois há 1.6 postos de trabalho para cada morador (BERTAUD, 2022a, 12).

Apesar da abundância de empregos, 14% dos parisienses trabalham fora da cidade. Segundo o autor, isto se deve ao fato de que os trabalhadores otimizam a localização de suas moradias procurando equilibrar diversos fatores, dentre os quais a remuneração e as oportunidades de crescimento dos postos de trabalho acessíveis, mas também o tempo de deslocamento e o tamanho da residência. Para ele, a proposta da cidade de 15 minutos deve não lograr sucesso por desconsiderar essa dinâmica urbana.

Ferrari, Monteiro e Amaral (2020) encontram evidências sobre essa estratégia de alocação residencial com base nos dados do censo demográfico de 2010 para a cidade de São Paulo. Os autores analisam a relação entre estrutura familiar e local de moradia e identificam um padrão espacial que concentra famílias unipessoais e casais sem filhos nas regiões centrais da cidade, enquanto famílias maiores tendem a situar-se em regiões mais afastadas. Segundo eles, a renda é um fator importante nesta dinâmica. Pois famílias com mais dependentes possuem menos renda disponível para gastos com moradia e maior necessidade de espaço. Como o valor da terra é maior em áreas centrais, isso faz com que essas famílias se situem em regiões mais afastadas (FERRARI; MONTEIRO; AMARAL, 2020, 419-420).

Em conjunto, os estudos aqui apresentados contêm evidências de que a proposta de uma cidade hiper-policêntrica, que está no cerne do movimento pela cidade de 15 minutos, não considera padrões espaciais mais amplos de deslocamento e moradia nas cidades, acarretando riscos de agravar a segregação. Como afirma o economista E. Glaeser em uma avaliação crítica sobre o movimento da Cidade de 15 Minutos:

“[o] foco em permitir que pessoas de renda média-alta possam passear em suas agradáveis pequenas vizinhanças de 15-minutos obscurece o problema muito mais amplo de como podemos fazer com que nossas cidades mais uma vez se tornem lugares de oportunidade para todos” (GLAESER, 2021).

Bertaud (2022b) argumenta que o urbanismo contemporâneo deve incorporar os achados da economia urbana para que se torne mais efetivo. Segundo o autor, propostas como a do movimento pela Cidade de 15 Minutos tendem não apenas a falhar, mas sobretudo a gerar efeitos negativos inesperados, por não considerar os mecanismos socioeconômicos que produzem o espaço urbano. O autor cita o Modelo Monocêntrico (SMM) como um dos principais referenciais teóricos a serem considerados para compreender as dinâmicas que regem a estrutura urbana.

2.3 O MODELO MONOCÊNTRICO PADRÃO

Inspirado no modelo de uso da terra agrária desenvolvido por Von Thunen em 1826 (THÜNEN, 1826), aquele que viria a ser conhecido como o Modelo Monocêntrico Padrão (SMM) foi inicialmente proposto por Alonso (1964), com contribuições significativas de Mills (1967) e Muth (1969) nos anos seguintes⁵.

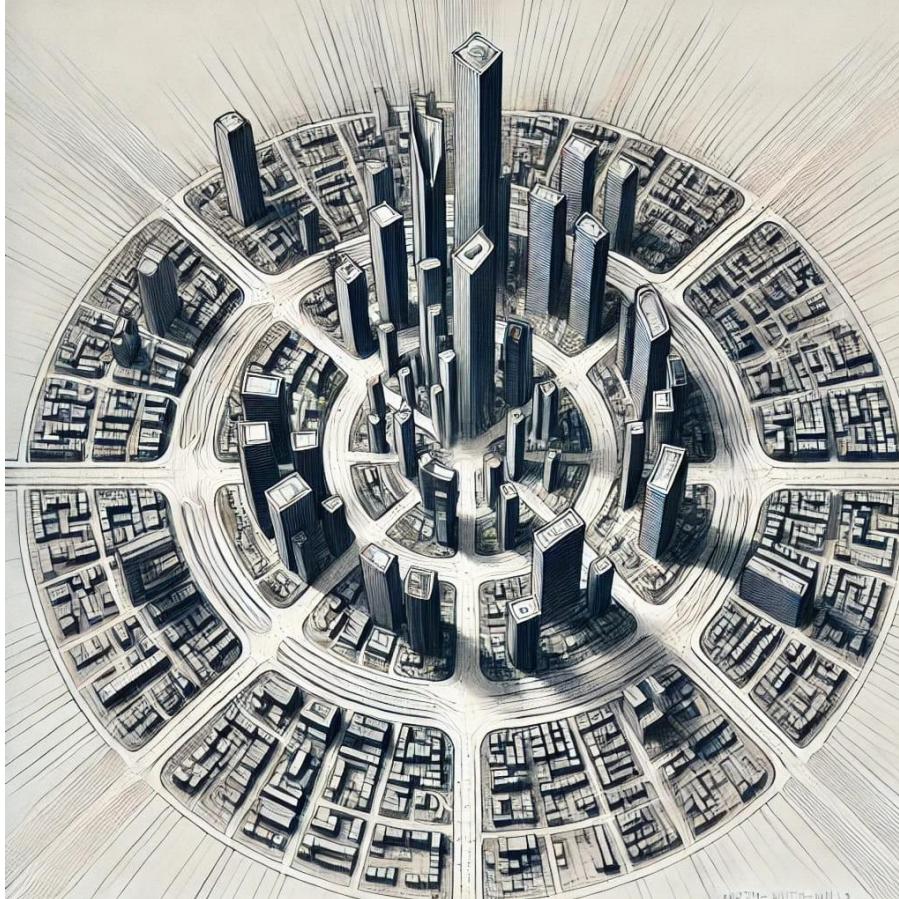
Trata-se de um modelo de equilíbrio fundamentado em fortes simplificações do espaço urbano real. Essa abordagem, que é caracterizada por Alonso (1964, 1) como um “pálido esqueleto”, não obstante teve grande influência na literatura sobre estrutura urbana. Bruckner (1987, 822-828) apresenta um tratamento unificado das formulações clássicas do SMM. Os parágrafos que seguem referenciam seu trabalho.

O modelo considera que a cidade possui um único distrito comercial central (CBD – do inglês, *Central Business District*) para o qual os cidadãos – que possuem todos a mesma renda e as mesmas preferências de consumo – se deslocam diariamente para trabalhar. Nas palavras de Alonso (1964, 15), a cidade é concebida como se estivesse localizada em uma planície uniforme. Nela, o deslocamento é realizado sem obstáculos, assim como toda a terra é de igual qualidade, com o mesmo potencial de edificação. As moradias são produzidas por um único dono de terras, que não encontra restrições legais ou urbanísticas para a construção, assim como não tem custos de demolição. Mas os custos de edificação aumentam de acordo com a quantidade de unidades construídas, o que faz com que os empreendimentos

⁵De onde advém a outra sigla pela qual o modelo é conhecido: modelo AMM (Alonso, Muth e Mills)

imobiliários tenham rendimentos decrescentes. A Figura 2.1 ilustra essa cidade idealizada.

Figura 2.1 – Ilustração para o modelo SMM.



Fonte: Elaboração própria, com o auxílio do ChatGPT

Nessa cidade, o cidadão ($c \in C$) é concebido como consumidor de apenas dois bens: a terra urbana para uso residencial e um único outro bem composto, que representa suas demais necessidades e pode ser concebido, em uma analogia, como *alimentação*. Os cidadãos arcam também com os custos de transporte, que são determinados pela distância dos deslocamentos pendulares ao CBD. Há, portanto, uma mesma restrição orçamentária para todos, com a seguinte forma:

$$r^c = tx^c + a + z^c s^c \quad (2.1)$$

Na qual c representa o cidadão e r^c representa sua renda (o c superscrito identifica a indexação para cada cidadão). O custo de deslocamento por metro linear – que, em uma analogia, pode ser pensado como o preço do combustível – é representado por t . Os custos com *alimentação*⁶ são representados por a . Como no modelo de Alonso (1964) todos os cidadãos possuem as mesmas preferências de consumo, esse valor é uma constante.

A distância em metros lineares da moradia do cidadão em relação ao CBD é representada por x^c ⁷. Na cidade idealizada por Alonso (1964) há apenas um único CBD que se encontra no centro geométrico da cidade, em $x = 0$, portanto. Dessa forma, os custos totais de transporte do cidadão são definidos por tx^c : o valor de deslocamento por metro quadrado multiplicado pela distância da moradia em relação ao CBD.

Assim como os custos totais de transporte, também o valor por metro quadrado do aluguel varia em relação à distância x do CBD, sendo representado por z^x . A quantidade de metros quadrados da residência do cidadão, por sua vez, é representada por s^c . Os custos totais de moradia são então definidos por $z^x s^c$.

Em outras palavras, a equação estabelece que os gastos com transporte (tx^c), com alimentação (a) e com moradia ($z^x s^c$) devem ser iguais à renda do cidadão (r^c).

Nesse cenário, o tamanho do imóvel alugado pelo cidadão, s^c , será dado pela inequação:

$$s^c \leq \frac{r^c - (tx^c + a)}{z^x} \quad (2.2)$$

Ou seja, ele é limitado pela renda disponível do cidadão (r^c), descontados seus gastos com transporte (tx^c) e com alimentação (a). E é inversamente proporcional ao valor do aluguel naquela região (z^x). Assim, cidadãos que optam por residir em localizações com menores custos por metro quadrado de aluguel podem arcar com imóveis maiores. Trata-se da estratégia de alocação residencial descrita por Ferrari, Monteiro e Amaral (2020), que foi originalmente conceitualizada por Beckmann (1973) com base no SMM.

De acordo com a mesma lógica, o valor por metro quadrado de aluguel

⁶Mais especificamente, o bem composto que representa as demais necessidades de consumo no dia a dia do cidadão.

⁷Aqui, a indexação em c identifica tratar-se da moradia do cidadão

que pode ser arcado pelo cidadão c é dado pela inequação:

$$z^x c \leq \frac{r^c - (tx^c + a)}{s^c} \quad (2.3)$$

Ou seja, ele é inversamente proporcional ao tamanho do imóvel alugado (s^c) e tende a aumentar conforme a distância ao CBD (x) diminui, pois os cidadãos que residem mais próximos ao local de trabalho têm menos gastos com transporte.

Como no modelo proposto por Alonso (1964) todos os cidadãos possuem a mesma renda (r), os mesmos gastos com *alimentação* (a) e os mesmos custos por metro quadrado de deslocamento (t)⁸, o valor por metro quadrado de aluguel tende a ser constante para cada distância x . De forma que a inequação pode ser reescrita sem a indexação em c e de forma independente do tamanho do imóvel(s):

$$z^{-x} \leq r^{-} - (tx + a) \quad (2.4)$$

Em outras palavras, o valor médio por metro quadrado de aluguel a uma dada distância x do CBD será limitado pela renda média dos cidadãos (r^{-}), subtraída dos gastos por transporte àquela distância (tx) somados aos gastos com *alimentação* (a).

A área construída residencial total localizada à distância x do CBD da cidade é definida por $S^x = \sum_{i=1}^n s_i$, onde n é a quantidade de unidades localizada nessa distância e s_i a área construída de cada uma delas. Há, evidentemente, um máximo natural para essa área construída se considerarmos apenas construções horizontais, pois esse tipo de construção poderá ocupar no máximo a área total do terreno. Assim, para uma determinada região da cidade definida por $x^{i+1} - x^i$, onde x representa um ponto distante x metros em relação ao CBD e $x^i + 1$ o ponto seguinte, a área construída máxima nessa porção $\max S^x$ será igual à área entre $x^{i+1} - x^i$. Mas esse máximo pode ser ultrapassado por meio da verticalização, que permite edificar, para uma dada porção da cidade $x^{i+1} - x^i$, uma quantidade de área construída S que seja superior à área dessa porção da cidade.

A verticalização permite, assim, o adensamento acima do limite da área de terreno. No entanto, esse adensamento será limitado pela viabi-

⁸Que podem ser pensados, por analogia, como o preço de combustível.

lidade financeira dos empreendimentos. Como afirmado anteriormente, o modelo **SMM** considera que o custo de construção, aqui definido por k , aumenta conforme a quantidade de metros quadrados construída, ou seja, $\frac{dk}{dS} > 0$. O custo total de construção de um empreendimento situado em x será assim definido por $K^x = k^S S^x$, onde S^x é a área construída total do empreendimento conforme definido anteriormente e k^S é o custo por metro quadrado de área construída para essa área construída total. Isso reflete a intuição de que edifícios altos requerem mais investimentos em fundação, materiais e técnicas construtivas, de forma que um prédio maior terá, em relação a um prédio menor, não apenas um custo total maior, como também um maior custo por metro quadrado de área construída.

O valor total retornado por aluguéis por um empreendimento imobiliário vertical localizado em x será definido por $Z^x = z^x S^x$. Ou seja, pela área construída total do empreendimento (S^x) multiplicada pelo valor do metro quadrado de aluguel naquela região (z^x). O empreendimento se tornará financeiramente inviável no ponto em que o custo total de construção, K , ultrapassar o retorno total com aluguéis Z . Pois não há incentivo para construir um edifício cujo custo de construção for maior que o rendimento que ele pode fornecer. Além disso, como o custo de construção por metro quadrado de área construída, representado por k , aumenta conforme a área construída total do edifício, S^x , há um valor ótimo de S^x que maximiza o lucro do empreendimento π^S , definido pelo retorno total esperado com aluguéis subtraído pelo custo de construção, $\pi^S = Z^x - K^x$, para um determinado valor de metro quadrado de aluguel k^x .

Como mencionado, o valor do metro quadrado de aluguel k^x reduz conforme aumenta a distância em relação ao **CBD** devido aos maiores custos de transporte das viagens pendulares (tx), que resultam em menor disponibilidade financeira dos cidadãos para arcar com os aluguéis. Dessa forma tem-se que $\frac{dk}{dx} < 0$. Disso decorre que tanto a área construída total máxima do empreendimento⁹ quanto sua área construída ótima¹⁰ diminuirão caso o empreendimento seja construído em uma região mais distante do centro

⁹Aquela na qual o custo total de construção é igual ao retorno total esperado com aluguéis a uma dada distância do **CBD** para uma determinada área construída.

¹⁰Aquela que maximiza o lucro, dado pelo retorno total esperado com aluguéis subtraído pelo custo total de construção a uma dada distância do **CBD** para uma determinada área construída.

comercial. Assim, de $\frac{dk}{dx} < 0$ decorre também que $\frac{ds}{tx} < 0$. Ou seja, como os aluguéis diminuem conforme aumenta a distância em relação ao CBD, também a área construída tende a diminuir.

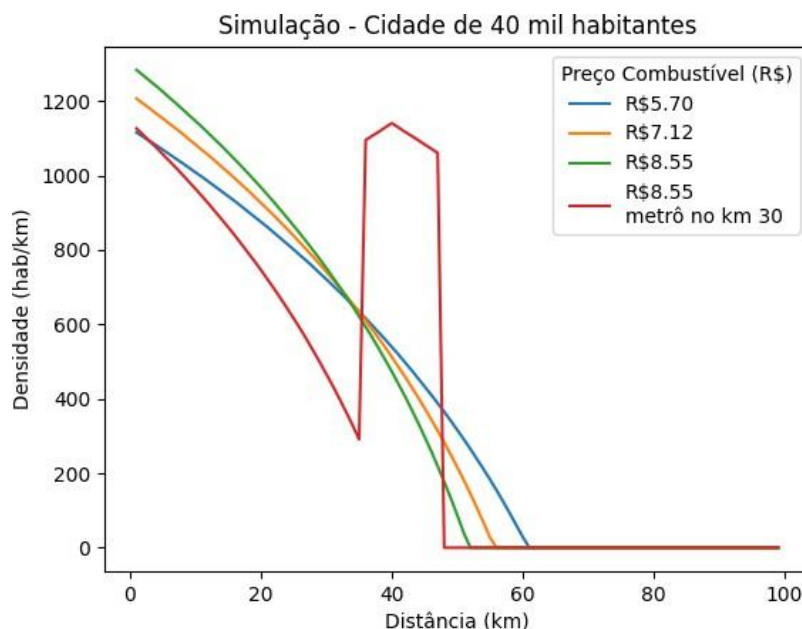
Por fim, o limite da área urbana edificada será dado pelo ponto em que o retorno esperado com aluguéis for menor que a renda da terra rural. Pois não há sentido em realizar um empreendimento imobiliário – um loteamento por exemplo – em um terreno que pode render mais caso seja utilizado para produção agrícola. Isso pode ser verificado em cidades com alto retorno da produção agrícola, como as cidades francesas na região de Champagne, nas quais a mancha urbana tende a não se expandir para a área agrícola (BERTAUD, 2022b). Nesse cenário, não apenas um aumento nos retornos da atividade agrícola, mas também um aumento dos custos de transporte t acarretará em uma redução da área urbanizada.

Isto pois, ao reduzir a renda disponível dos cidadãos, esse aumento de custo reduz também o valor esperado de retorno com aluguéis para todas as distâncias x do CBD. Dessa forma, o limite da área urbanizada é “movido para a esquerda”, ou seja, ele passa a se localizar em uma nova posição x' , dado que $x' < x$. Tal redução da área urbanizada implicará também em um aumento da densidade, pois será necessário que o tamanho médio das unidades habitacionais \bar{s} seja reduzido para abrigar a mesma quantidade de pessoas no mesmo espaço. Pela mesma lógica, uma redução drástica nos custos de transporte – por exemplo, pela popularização do automóvel e preços baixos de gasolina nos anos 1950-1960 – resulta em um aumento da área urbanizada e em uma redução da densidade. Dessa forma o modelo SMM explica o espraiamento (BERTAUD, 2022b).

O padrão monocêntrico que caracteriza as principais representações do modelo SMM decorre dessa dinâmica, na qual a densidade tende a ser maior nas áreas mais próximas ao CBD. Evidentemente, trata-se de uma simplificação excessiva dos fatores complexos que de fato influenciam a morfologia urbana. No entanto, apesar da simplicidade do SMM e dos fortes pressupostos nos quais se baseia, esse padrão é surpreendentemente aderente aos dados. A Figura 2.2 representa o mapeamento da densidade construtiva de tipo residencial vertical em São Paulo para o ano de 2023, obtida a partir da metodologia descrita na seção D.1 do Anexo Único. Como se pode notar, há uma concentração da área construída nas áreas centrais da

40 mil habitantes e diferentes custos de transporte. Como se pode conferir, o aumento dos custos de transporte resulta em um aumento da densidade e diminuição da área urbanizada, de acordo com o padrão monocêntrico previsto pelo [SMM](#). No entanto, a construção de uma estação de metrô, reduzindo o custo de transporte dos cidadãos que residem em até 5km de distância dela, acarreta um aumento substancial da densidade nessa área, como se pode conferir pela linha vermelha na Figura 2.3.

Figura 2.3 – *Simulação dos efeitos do aumento dos custos de transporte com base nos parâmetros do SMM.*



Fonte: Elaboração própria.

A simulação assinala assim a possibilidade de surgimento de outras centralidades a depender de alterações dos parâmetros do modelo [SMM](#), como demonstraram Fujita e Ogawa (1982). Esses dados mostram que a discussão sobre se as cidades são policêntricas ou monocêntricas – que dominou a literatura empírica em torno do [SMM](#) – passa ao largo da principal contribuição teórica do modelo, qual seja, a relação positiva entre acessibilidade, valor da terra e densidade (DURANTON; PUGA, 2015, 527). Além disso, esses resultados o [SMM](#) permitem realizar uma avaliação crítica de

políticas públicas de planejamento urbano que buscam direcionar a produção imobiliária por vias regulatórias apenas, sem considerar as dinâmicas mais amplas que regem a estrutura urbana, em especial os efeitos da acessibilidade sobre a densidade urbana.

O capítulo 3, que segue, consiste em uma avaliação, com base em evidências, da capacidade do Plano Diretor Estratégico do município de São Paulo em direcionar o adensamento construtivo na cidade no período entre os anos de 2014 e 2023. Como será demonstrado, os incentivos construtivos estabelecidos pelo Plano Diretor tiveram sucesso apenas na região nobre da cidade, que já era buscada pelo mercado imobiliário. Não surpreendentemente, trata-se da região com os menores tempos de deslocamento a trabalho, como previsto pelo SMM.

Esse tema será retornado no capítulo 4, que tem por objetivo verificar empiricamente a relação entre acessibilidade e densidade, tomando como estudo de caso a cidade de São Paulo no período 2007-2017. A hipótese sustentada no capítulo é que um aumento na acessibilidade ao mercado de trabalho em uma dada região da cidade (ou seja, uma redução de t) está positivamente associada a um aumento da densidade construtiva de tipo residencial vertical nessa mesma região, como prevem os resultados da simulação apresentada na Figura 2.3. Como será demonstrado, os dados apresentam evidências que apoiam essa hipótese, contribuindo para a validação empírica do SMM.

3 EFETIVIDADE DO PLANO DIRETOR

Resumo do capítulo

Este capítulo analisa a capacidade do planejamento urbano de direcionar a produção imobiliária, com foco no Plano Diretor Estratégico de São Paulo. O estudo investiga o impacto dos incentivos construtivos concedidos pelo Plano Diretor em áreas próximas a terminais de transporte público (EETU) sobre a construção de unidades residenciais verticais entre 2014 e 2023. Utilizou-se um modelo estatístico bayesiano multinível para zeros inflados, que considera tanto fatores que promovem quanto restrições à verticalização, como a disponibilidade de terrenos livres e áreas de proteção ambiental. Os resultados mostram que os incentivos tiveram efeito positivo na área nobre da cidade, que já era alvo do interesse do mercado imobiliário, e nas áreas de classe média, não havendo impacto nas periferias. Dessa forma, o planejamento urbano não foi capaz de reverter as estruturas de segregação vigentes na cidade.

Objetivo: Analisar a efetividade do planejamento urbano em direcionar a produção imobiliária.

3.1 INTRODUÇÃO

O PDE de São Paulo, aprovado em 2014, estabelece diretrizes para induzir o desenvolvimento ordenado e sustentável da cidade até 2029. Dentre seus objetivos, destaca-se a meta de aumentar o adensamento urbano nas proximidades dos eixos de transporte público do município por meio do incentivo à produção imobiliária nestes locais. Com isso, o Poder Público busca reverter os padrões de segregação vigentes na cidade, reduzindo a desigualdade no tempo de deslocamento da residência ao trabalho e ampliando o acesso da população em geral às regiões mais bem providas de infraestrutura, serviços e equipamentos públicos.

As regiões onde este adensamento construtivo é esperado foram identificadas no Plano Diretor sob o termo EETU. Nome que indica a esperança que os planejadores urbanos possuíam, no momento da elaboração do PDE, de que este dispositivo promovesse uma alteração no *status quo* do muni-

cípio no que diz respeito à ampliação do direito à cidade. No entanto, a despeito das expectativas elevadas, cumpre reconhecer que o sucesso desta política urbanística não depende apenas da ação do Poder Público, uma vez que sua efetividade está diretamente relacionada à adesão do setor de produção imobiliária formal. Pois cabe ao mercado imobiliário incorporar, projetar e construir a esmagadora maioria dos empreendimentos que gerarão o crescimento almejado pelo planejamento municipal nas áreas por ele designadas.

O presente estudo foi desenvolvido no contexto dos debates que acompanharam o processo de revisão do [PDE](#) entre os anos de 2022 e 2023. Ele tem por objetivo contribuir com uma análise baseada em evidências sobre a efetividade do [PDE](#), especificamente em relação à sua capacidade de promover o adensamento construtivo nos Eixos de Estruturação da Transformação Urbana na cidade como um todo – e não apenas de forma local, em áreas já valorizadas previamente pelo mercado imobiliário.

A base de dados do Imposto Territorial e Predial Urbano ([IPTU](#)) foi utilizada para identificar o adensamento construtivo de tipo vertical na cidade de forma territorializada, medido pela quantidade de unidades produzidas no período 2014-2023 em cada quadra da cidade. Além disso, para cada quadra foram construídas variáveis que representam fatores associados ao seu potencial construtivo – por exemplo, a área de terrenos e galpões –, assim como restrições à edificação, como a presença de remanescentes de Mata Atlântica. A esses dados foi adicionada a identificação de se a quadra está situada na área nobre ou nas áreas periféricas da cidade.

A caracterização da área nobre e das áreas periféricas da cidade foi realizada por meio de uma *clusterização* espacial (ANSELIN, 1995) a nível de distrito do componente principal (GREENACRE et al., 2022) que representa as correlações entre diferentes medidas de privilégio socioeconômico, segregação racial e a duração das viagens pendulares. Como será demonstrado, em São Paulo os distritos que possuem maior proporção de pessoas brancas, ricas e alfabetizadas são também aqueles cujos moradores possuem os menores tempos de deslocamento a trabalho.

Esses distritos de alta concentração de privilégio estão localizados de forma avizinhada em uma área que corresponde ao Setor Sudoeste da cidade, descrito por Villaça (2001) como a região de residência da elite. Essa

área é referida neste estudo como “área nobre”, de forma a referenciar o jargão utilizado pelo mercado imobiliário que tem nesta região o destino prioritário de seus investimentos. Nos extremos Leste, Sul e Noroeste, por sua vez, encontram-se agrupamentos de distritos com os maiores tempos de deslocamentos e uma maior proporção de pessoas pretas, pardas, pobres e analfabetas. Esses *clusters* foram denominados como “áreas periféricas” neste estudo, por representarem as regiões de maior concentração de vulnerabilidade socioeconômica na cidade.

Os dados descritos nos parágrafos anteriores foram analisados por meio de uma série de modelos *bayesianos* progressivamente mais complexos, seguindo a metodologia estabelecida por Gelman et al. (2020). O resultado da análise, em especial o último modelo, uma regressão multinível com função de ligação binomial negativa para zeros inflados (MCELREATH, 2020) apresenta evidências de que os incentivos construtivos fornecidos pelos Eixos de Estruturação da Transformação Urbana tiveram efeitos apenas na área nobre e de classe média da cidade, não havendo efeito mensurável nas áreas periféricas. Além disso, a comparação dos interceptos para as áreas nobre e periférica mostra que a simples localização em área nobre, tudo o mais mantido igual, acarreta um aumento considerável da quantidade esperada de unidades de tipo vertical produzida no período.

3.1.1 Objetivo

O objetivo deste capítulo é analisar a capacidade dos Eixos de Estruturação da Transformação Urbana definidos pelo Plano Diretor do município de São Paulo de direcionar a verticalização da cidade entre os anos de 2014 e 2023, medida em termos da quantidade de unidades de tipo vertical (comerciais e/ou residenciais) produzidas no período.

3.1.2 Organização

O capítulo está organizado da seguinte forma. A seção 3.2, [Materiais e Métodos](#) é iniciada com a descrição de como a base de dados utilizada no presente estudo foi construída e descreve as variáveis utilizadas nos modelos estatísticos. Esses dados são detalhados posteriormente no Anexo [Único](#),

que tem por objetivo permitir a replicabilidade do presente estudo. Essa seção é finalizada com a descrição metodológica de cada um dos modelos desenvolvidos.

Em seguida, a seção 3.3 apresenta os resultados dos modelos estatísticos ajustados para o estudo. O estudo foi realizado de acordo com o método do fluxo de trabalho *bayesiano* (GELMAN et al., 2020), de modo que foram especificados modelos progressivamente mais complexos de forma a se aproximar sucessivamente de uma modelagem mais adequada aos dados e ao problema de pesquisa. Os resultados são apresentados nas seguintes seções:

- A seção 3.2.5 apresenta os resultados da análise de agrupamentos espaciais que foi realizada para caracterizar a estrutura de segregação social existente na cidade de São Paulo;
- A seção 3.3.2 apresenta os resultados do primeiro modelo ajustado, uma Regressão de Poisson para dados de contagem;
- A seção 3.3.3 apresenta os resultados do modelo de Regressão Binomial Negativa, que possibilita capturar a superdispersão presente nos dados;
- A seção 3.3.4 apresenta os resultados do Modelo de Regressão Binomial Negativa para Zeros Inflados, um modelo de mistura que permite modelar simultaneamente o processo de contagem, associado ao potencial construtivo, e o processo de zeros estruturais, associado às restrições à edificação vigentes;
- A seção 3.3.5 apresenta os resultados do Modelo Multinível com Regressão Binomial Negativa para Zeros Inflados, que estende o modelo da seção 3.3.4 em uma modelagem multinível na qual os dados são analisados no contexto dos agrupamentos espaciais identificados na seção 3.2.5.

A seção 3.4, *Considerações Finais*, resume e discute os resultados encontrados nos modelos ajustados.

3.2 MATERIAIS E MÉTODOS

Nesta seção descrevemos como foi produzida a base de dados que foi analisada neste estudo, assim como os métodos estatísticos empregados. Mais detalhes sobre as fontes de dados utilizadas e os procedimentos de engenharia de atributos são apresentadas no Anexo Único.

3.2.1 Unidades verticais produzidas entre 2014 e 2023

A resposta ao problema de pesquisa apresentado na seção 3.1.1 requer a identificação do adensamento construtivo de tipo vertical ocorrido em São Paulo durante o período de vigência do PDE, de forma territorializada. Para isso, nos valem da base de dados do IPTU, que contém informações censitárias sobre mais de 3 milhões e meio de imóveis no município. Essa base foi tornada pública pela Prefeitura de São Paulo, e teve os polígonos georreferenciados disponibilizados pela Secretaria da Fazenda no Portal GeoSampa em 2019. Ela é descrita em maiores detalhes na seção C.1.1 do Anexo Único.

Por meio da comparação entre os imóveis cadastrados em diferentes anos da base de dados, é possível identificar as unidades produzidas no período compreendido entre eles. Para o estudo apresentado no presente capítulo, foram comparados os dados de todos os imóveis de tipo de construção vertical (residências ou horizontais) cadastrados nas bases do IPTU para os anos de 2014 e 2023. Como a base do IPTU representa a situação do cadastro imobiliário fiscal em 1º de janeiro do ano a que ela corresponde, a base de 2014 representa, assim, o registro mais próximo da situação base da cidade quando da promulgação do Plano Diretor, que ocorreu em julho desse ano.

Para contabilizar as unidades produzidas, foram identificados todos os imóveis verticais presentes em 2023 que não estavam previamente cadastrados em 2014, assim como aqueles já cadastrados em 2014 mas que tiveram mudança de uso – o que permite identificar *retrofits*. Em seguida, foram selecionados apenas os imóveis cujo ano de construção é igual ou superior a 2014, de modo a garantir que eventuais regularizações fundiárias não se-

jam contabilizadas como produção imobiliária¹. Por fim, foram removidos todos os cadastros referentes a vagas de garagem e a usos que não sejam residenciais ou comerciais (por exemplo, depósitos e antenas de celular), com o objetivo de certificar tratar-se de fato de unidades em condomínios verticais.

Foi então realizada a contagem da quantidade de registros por quadra fiscal e, por meio do cruzamento com a camada de quadras disponibilizada no GeoSampa, os dados foram georreferenciados. Foi obtida assim a quantidade de novas unidades verticais produzidas em cada quadra da cidade de São Paulo entre os anos de 2014 e 2023. Essa informação corresponde à variável *y* do presente estudo. O Diagrama 3.1 representa os passos descritos nesta seção. A metodologia aqui empregada é detalhada no Anexo Único, na seção D.1.2.

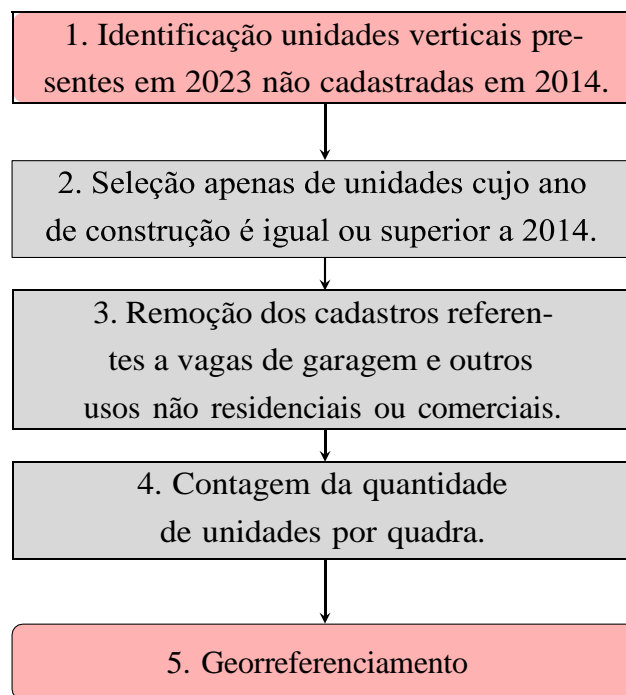


Figura 3.1 – *Identificação das unidades verticais produzidas (2014-2023).*

¹Para mais detalhes sobre essa questão, veja a seção D.1 do Anexo Único

3.2.2 Quadras em Zona de Estruturação Urbana

Os Eixos de Estruturação da Transformação Urbana ([EETU](#)) são porções do território do município definidas pela área de influência do transporte coletivo de média e alta capacidade (trens, metrô e corredores de ônibus) que o Plano Diretor estabeleceu como prioritárias para o adensamento urbano (Prefeitura Municipal de São Paulo, 2014b, art. 75). Isso pode ser conferido no artigo 23 da lei que promulgou o [PDE](#), que define como um dos objetivos estratégicos nos [EETU](#) a promoção do melhor aproveitamento do solo por meio do aumento na densidade construtiva, habitacional e populacional (Prefeitura Municipal de São Paulo, 2014b, art. 23).

Trata-se de uma abordagem inspirada no movimento do *transit-oriented development* ([TOD](#)) que busca conter o espraiamento e reduzir a dependência do automóvel por meio do incentivo ao adensamento próximo aos eixos de transporte coletivo (BALBIM; KRAUSE, 2016). Ela é uma das principais estratégias do [PDE](#), que é melhor compreendida por meio do artigo 7º, incisos I a III, citados a seguir:

“Art. 7º A Política de Desenvolvimento Urbano e o Plano Diretor Estratégico se orientam pelos seguintes objetivos estratégicos:

- I - conter o processo de expansão horizontal da aglomeração urbana, contribuindo para preservar o cinturão verde metropolitano;
- II - acomodar o crescimento urbano nas áreas subutilizadas dotadas de infraestrutura e no entorno da rede de transporte coletivo de alta e média capacidade;
- III - reduzir a necessidade de deslocamento, equilibrando a relação entre os locais de emprego e de moradia;”

(Prefeitura Municipal de São Paulo, 2014b, art. 7º)

Essa estratégia pode ser concebida por meio da analogia com uma barragem que busca direcionar o curso de um rio. De um lado, o [PDE](#)

estabelece uma série de restrições ao adensamento e à verticalização em diversas áreas da cidade, com destaque para as áreas de proteção ambiental, protegidas por Zona Especial de Preservação Ambiental (**ZEPAM**) e às áreas exclusivamente residenciais unifamiliares, protegidas pela Zona Exclusivamente Residencial (**ZER**). De outro lado, são conferidos uma série de incentivos construtivos nas áreas de **EETU**, acompanhados de regramentos que resultam na produção de maiores quantidades de unidades por empreendimento. Veja a seção **D.2** do Anexo **Único** para maiores detalhes sobre isso. Essas restrições representam a barragem, que busca conter o fluxo da produção imobiliária, enquanto os incentivos construtivos podem ser concebidos como o canal, para o qual o planejamento urbano busca direcionar essa produção.

As áreas de influência que definem os **EETU** são delimitadas pelos mapas 3 e 3A, anexos à lei que promulga o **PDE**. Elas foram construídas por meio da aplicação de *buffers* de 300 metros a partir da localização das estações de trem e metrô e 150 metros a partir dos corredores de ônibus². O artigo 76 da lei do **PDE** esclarece que, mesmo que apenas parte da quadra tenha sido intersectada pelo *buffer* da área de influência, toda a quadra deverá ser considerada como localizada em **EETU**, desde que essa quadra esteja contida em um *buffer* com o dobro do tamanho do *buffer* original, ou seja, de 600 metros para trens e metrô e 300 metros para corredores de ônibus (Prefeitura Municipal de São Paulo, 2014b, art. 76).

A Figura 3.2, extraída do GeoSampa, apresenta a localização dos **EETU** em cinza. Como se pode notar, os **EETU** estão localizados em todas as zonas da cidade³, incluindo os extremos Norte, Leste e mesmo Sul (há **EETU** em Parelheiros). Mas há uma maior concentração nas áreas central e sudoeste, que são mais providas de corredores de ônibus e estações de metrô.

O parágrafo único do art. 75 da lei que promulgou o **PDE** estabeleceu que as disposições relativas aos usos e atividades, índices e parâmetros de ocupação do solo definidas no **PDE** para as áreas de **EETU** entraram em vigor na data de publicação da lei, prevalecendo inclusive sobre a Lei de

²E também veículos leves sobre pneus (VLP), que não existem ainda em São Paulo mas foram contemplados pelo **PDE**

³Isto é, as zonas Leste, Oeste, Norte, Sul e o Centro

Figura 3.2 – Distribuição dos **EETU** no território municipal.

Fonte: GeoSampa/Prefeitura de São Paulo

Parcelamento, Uso e Ocupação do Solo (**LPUOS**) então vigente, isto é, a Lei 13.885, de 2004 (Prefeitura Municipal de São Paulo, 2014b, art. 75). No entanto, quando da revisão da **LPUOS**, foram formalizadas nas áreas de **EETU** as Zonas de Estruturação Urbana (**ZEU**). Elas implementam, por meio do zoneamento, os parâmetros construtivos definidos para os **EETU** no **PDE**, integrando a estratégia definida no Plano Diretor às atividades de controle edilício estabelecidas na **LPUOS** e conferindo maior segurança jurídica a elas.

Neste estudo, utilizamos a camada de zoneamento referente à Lei 16.402/2016 (a **LPUOS**) para identificar as quadras situadas em **ZEU**. Para realizar isso, em primeiro lugar os polígonos da camada de zoneamento foram dissolvidos por meio da sigla da zona que lhes foi atribuída na **LPUOS**, de modo a conter apenas um multipolígono para cada tipo de zona. Em seguida, foi realizado o *join* espacial entre a camada de quadras fiscais e o

zoneamento por meio da intersecção entre os polígonos das quadras fiscais e os multipolígonos que representam cada tipo de zona. Nos casos (raros) em que as quadras são intersectadas por mais de um tipo de zona, foi calculada a área da intersecção entre o polígono da quadra fiscal e os multipolígonos que a intersectam. Foi atribuída como zona da quadra a zona referente ao multipolígono com maior área de intersecção com o polígono da quadra. Em outras palavras, a zona “preponderante” na quadra.

Por fim, foi criada uma coluna *booleana* que identifica se a zona atribuída à quadra é **ZEU**. Essa informação corresponde à variável tratamento deste estudo, cujos efeitos sobre a quantidade esperada de unidades produzidas por quadra no período temos por objetivo medir. O mesmo método foi também utilizado para identificar as quadras situadas em **ZER**. O Diagrama 3.3 resume os procedimentos realizados.

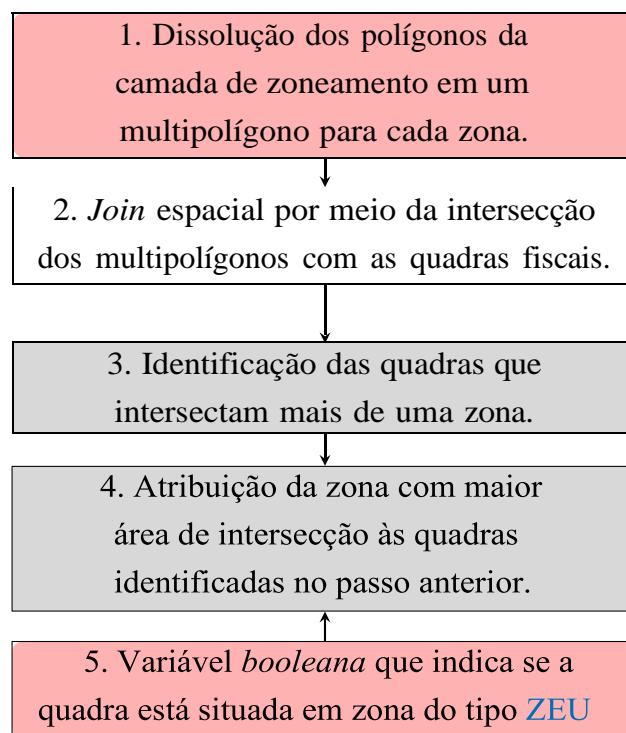


Figura 3.3 – Identificação das quadras em **ZEU**.

3.2.3 Fatores associados à verticalização

Por meio dos dados dos imóveis disponibilizados na base do IPTU, foram calculadas para cada quadra uma série de variáveis associadas ao potencial de edificação da mesma. Essas variáveis foram utilizadas como controles nos modelos estatísticos desenvolvidos para este estudo. Todas elas têm como referência o ano de 2014, representando assim a situação-base do estoque de imóveis na cidade anterior à aplicação do tratamento – isto é, à promulgação do PDE, que forneceu os incentivos construtivos às áreas de ZEU. Elas estão listadas a seguir:

- Valor médio do metro quadrado de terreno dos imóveis situados na quadra;
- Média da idade dos imóveis situados na quadra, calculada a partir do ano de construção;
- Média da idade dos proprietários dos imóveis situados na quadra, calculada a partir do ano de nascimento do contribuinte;
- Média da quantidade de pavimentos dos imóveis na quadra;
- Proporção da área de terreno do imóvel não-verticalizado com o maior terreno situado na quadra em relação à área total da quadra⁴;
- Proporção da área construída total de edificações de tipo vertical em relação à área da quadra;
- Proporção da área construída total de edificações de tipo horizontal em relação à área da quadra;
- Proporção da área total dos terrenos ocupados por galpões e usos similares em relação à área da quadra;
- A área da quadra;

⁴Essa variável busca identificar por exemplo a presença de casas com grandes terrenos ainda remanescentes na quadra, o que facilita a incorporação

Como detalhado na seção [D.1](#), há a possibilidade de que sejam criadas novas quadras na cidade, devido sobretudo à regularização de loteamentos e ao parcelamento do solo, ainda que nas últimas décadas isso seja um evento relativamente raro devido ao esgotamento das glebas loteáveis. No período de nosso estudo, apenas 860 quadras novas quadras foram criadas, em um universo de mais de 45 mil (correspondendo assim a menos de 0,2%). No entanto, por não existirem anteriormente, essas quadras não possuem informações cadastrais na base do [IPTU](#) para o ano de 2014, de modo que não é possível obter diretamente as variáveis controle descritas nesta seção para elas.

Para contornar esse problema, foi desenvolvida a metodologia descrita no Diagrama [3.4](#). Basicamente, ela consiste em imputar para as quadras criadas o valor médio das variáveis apresentadas nessa seção das quadras que se situam em seu entorno e que já existiam anteriormente. Essa metodologia se justifica pois as variáveis aqui consideradas possuem certa homogeneidade espacial. Por exemplo, quadras vizinhas em um mesmo loteamento tendem a ter imóveis com a mesma idade e com o mesmo valor de metro quadrado de terreno. O mesmo vale para os usos, havendo bairros com predominância residencial e bairros com predominância comercial, assim como antigos bairros industriais que ainda possuem muitos galpões. Dessa forma, a metodologia desenvolvida busca identificar como era a situação, em 2014, das quadras vizinhas da quadra que foi criada, pois é bastante provável que os lotes que deram origem à quadra fossem similares aos da sua vizinhança. Assim, na ausência de outras informações, é sensato pressupor que a situação original da quadra que foi criada tenha sido originalmente correspondente à situação média de seu entorno.

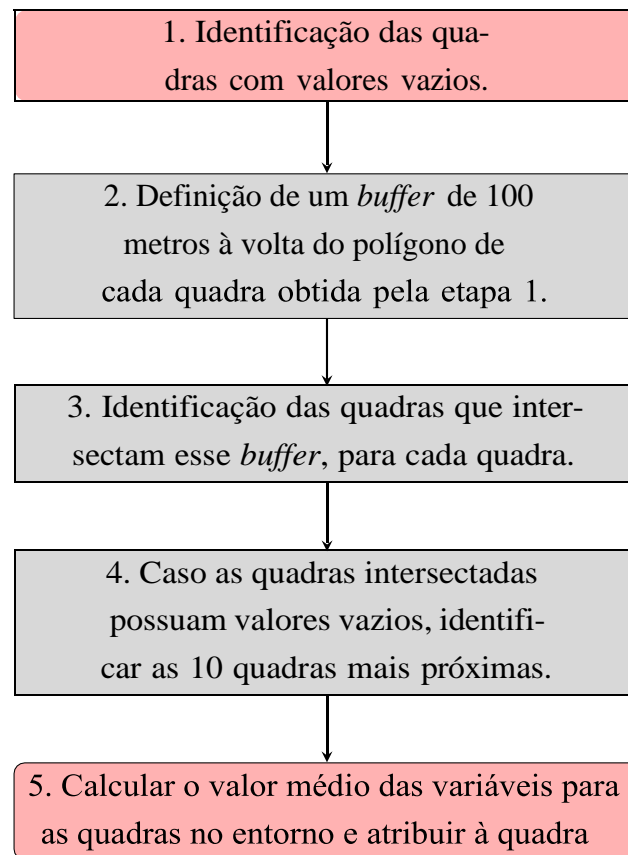


Figura 3.4 – *Imputação de dados faltantes para quadras criadas no período 2014-2023.*

Essas variáveis, assim como a razão para a escolha de cada uma delas, são detalhados na seção D.3 do Anexo Único. O Quadro 19 contém a descrição de cada uma delas. Para permitir a comparabilidade dos tamanhos de efeito, as variáveis descritas nesta seção foram padronizadas com o uso do método *Z-scores*. Nesse método, os valores padronizados são obtidos a partir da aplicação da equação $z_i = \frac{x_i - \mu}{s}$, onde x_i representa o valor da variável x para a i -ésima quadra, μ representa a média da variável x e s seu desvio-padrão.

3.2.4 Restrições construtivas

Contudo, além dos fatores associados ao potencial construtivo das quadras, é preciso considerar que há também uma série de restrições à edificação estabelecidas na legislação vigente. Trata-se sobretudo de restrições

associadas à proteção ambiental e do patrimônio arquitetônico da cidade, como é o caso das Áreas de Proteção Ambiental ([APA](#)) e dos tombamentos. Mas há também a possibilidade de que o próprio zoneamento implique em restrições à edificação, como é o caso das Zonas Exclusivamente Residenciais Unifamiliares ([ZER-1](#)), nas quais a [LPUOS](#) proíbe usos verticais.

Neste estudo as seguintes variáveis foram construídas para identificar restrições construtivas:

- Área envoltória de imóvel tombado;
- Tombamento;
- Bairro Ambiental;
- Área contaminada;
- Área de proteção ambiental ([APA](#));
- Área de proteção aos mananciais;
- Área com restrição geotécnica;
- Área de risco geológico;
- Reserva Particular do Patrimônio Natural ([RPPN](#));
- Unidade de conservação;
- Rampa de pouso e decolagem de heliponto;
- Zona Exclusivamente Residencial ([ZER](#));

Todas elas consistem em variáveis *booleanas* que indicam se o polígono da quadra foi intersectado por algum polígono contido na camada do GeoSampa que representa essa restrição. Por exemplo: caso a quadra esteja localizada dentro de uma área de proteção ambiental, ainda que apenas parte dela esteja dentro dos limites da [APA](#), a variável indicará que essa restrição incide sobre a quadra. A seção [D.4](#) do Anexo [Único](#) detalha a metodologia utilizada e a escolha das variáveis. O Quadro [18](#) contém a descrição de cada uma delas.

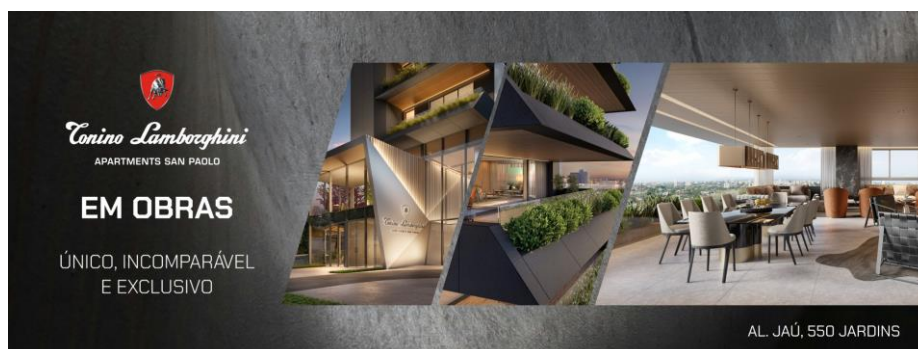
Diferentemente das variáveis descritas na seção 3.2.3, que foram padronizadas com o método *Z-scores*, as variáveis *booleanas* apresentadas nesta seção foram apenas convertidas em *dummies*. A padronização não foi realizada para permitir uma interpretação semanticamente mais significativa para o contexto do estudo. Por exemplo: faz muito mais sentido mensurar o efeito da quadra estar situada em área de proteção ambiental, sim ou não, do que o efeito do acréscimo de um desvio padrão dessa variável.

3.2.5 Identificação das áreas nobre e periféricas da cidade

Todas as variáveis descritas neste capítulo até então têm como nível de agregação as quadras da cidade. No entanto, há que se considerar que o fenômeno da verticalização não é determinado exclusivamente pela escala diretamente local, isto é, apenas pelas características dos imóveis situados no quarteirão no qual futuramente estará situado o edifício. Isso pode ser facilmente notado pela visibilidade que as empresas do ramo imobiliário dão à localização do empreendimento. A importância é tamanha que isso ficou gravado na cultura popular por meio da expressão “Location, location, location”, bordão utilizado pela apresentadora de um *reality show* britânico de mesmo nome (CONTRIBUTORS, 2024b).

Considere-se, por exemplo, o anúncio na Figura 3.5, em especial o destaque dado à *exclusividade* do empreendimento, assim como ao bairro em que está situado. Note-se que essas informações são apresentadas na ausência de qualquer dado sobre as unidades a serem vendidas ou as características do edifício em si, como a metragem dos apartamentos, quantidade de quartos, número de vagas na garagem ou a presença de estruturas lazer nas áreas comuns, como piscinas ou *rooftops*. Na página do anúncio no site da construtora, afirma-se que “o bairro Jardins é um bairro de alto padrão, sendo uma das regiões mais cobiçadas para se morar em São Paulo”, seguida de uma descrição da acessibilidade do bairro a comércio e serviços, assim como equipamentos públicos, como parques e museus. Isso é reforçado pela apresentação de um mapa com a localização do empreendimento e no qual esses pontos de interesse estão destacados (GAFISA, 2024).

Outros anúncios, como o apresentado na Figura 3.6 utilizam-se de imagens aéreas anotadas para indicar a proximidade do empreendimento

Figura 3.5 – Anúncio de empreendimento exclusivo.

Fonte: Gafisa - <https://www.gafisa.com.br/sp>

com regiões valorizadas – muitas vezes de forma enganosa, com o uso de técnicas de perspectiva que dão a entender que os elementos estão mais próximos do que realmente estão. A própria Figura 3.6 foi extraída de uma postagem de *blog* na qual uma empresa de marketing afirma que as empresas do ramo imobiliário devem utilizar a informação da localização de seus empreendimentos de “maneira inteligente, persuasiva e estratégica” (IMÓVEIS, 2020) para maximizar o valor de venda das unidades. Essa expectativa de valorização por meio da localização é tamanha que não é raro verificar um esforço discursivo de alargamento das fronteiras históricas de alguns bairros para que se possa afirmar que o empreendimento está localizado neles.

Isso é reforçado por portais de anúncios imobiliários, que produzem listagens dos chamados “bairros nobres”. Considere-se por exemplo postagens de blog da Zap Imóveis e QuintoAndar, duas das principais empresas de tecnologia atuantes no ramo de anúncios imobiliários, nas quais são listados os “bairros nobres” da cidade de São Paulo (ZAP, 2024; QUINTO-ANDAR, 2024). Esse fenômeno não se restringe a São Paulo ou mesmo ao Brasil, estando presente em diversos países, como pode ser constatado em um artigo na Wikipedia, que lista os bairros nobres de diversas cidades no mundo (CONTRIBUTORS, 2024a).

Trata-se das áreas de residência da elite. Villaça (2001) analisa esse tema, mostrando como nas cidades brasileiras as classes sociais mais altas ocupam porções privilegiadas do espaço urbano. Segundo o autor, no início

Figura 3.6 – *Propaganda com a localização do empreendimento.*



Fonte: Acqua -

<https://www.aqua.com.br/blog/mapa-localizacao-no-setor-imobiliario>

do processo de urbanização, essas áreas possuíam melhores condições naturais, como um clima mais ameno, menor propensão de serem atingidas por inundações, melhor circulação do ar, vistas a belas paisagens, entre outros fatores. Quando da escolha dos locais de moradia, essas características naturais e paisagísticas foram associadas à busca por uma localização de fácil acesso ao setor comercial central das cidades (CBD), de modo que as elites pudessem usufruir dessas qualidades sem ter que arcar com viagens pendulares de duração elevada.

Ao longo do tempo, o direcionamento dos investimentos públicos nessas áreas por meio da influência dos membros da elite que nela residem reforçou o caráter privilegiado das mesmas ao fazer com que nelas, ou próximo a elas, fossem estabelecidos os principais equipamentos públicos, como parques, praças e museus. Ademais, se formaram setores de comércio e serviços pujantes no entorno dessas vizinhanças, com o objetivo de atrair os clientes com maior poder aquisitivo que nelas residem. Políticas públicas higienistas realizadas no final do século XIX e início do século XX, concebidas e patrocinadas por membros dessas elites, resultaram na expulsão dos pobres e das classes trabalhadoras que porventura ainda residiam nessas áreas.

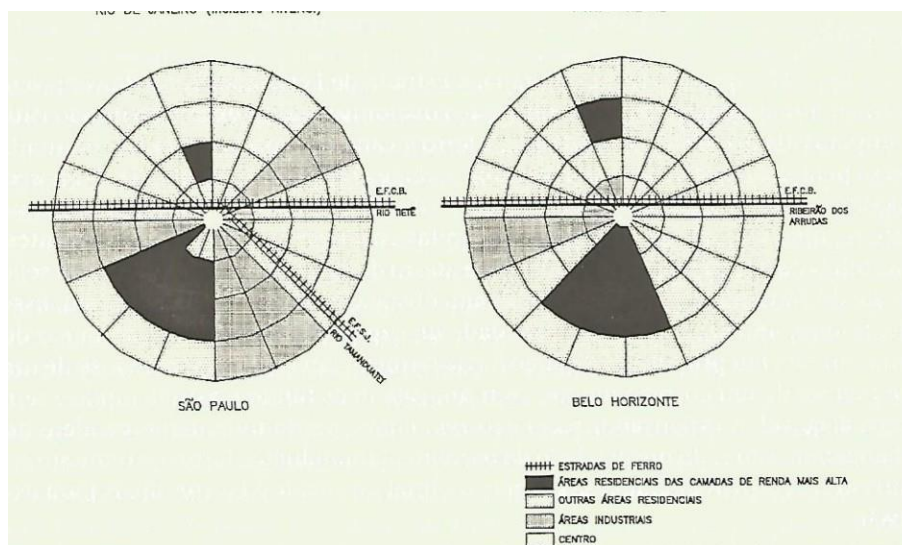
Essas medidas foram reforçadas ao longo do século XX pelo planejamento urbano, em especial o zoneamento, que direcionou os usos com impactos ambientais desagradáveis para regiões distantes desses bairros. Avanços recentes na literatura sobre o tema têm apresentado evidências de que o zoneamento foi mobilizado de forma a promover a segregação social, mesmo quando não ostensivamente concebido sob o viés racial, como no caso dos guetos em situações de *apartheids* (SHERTZER; TWINAM; WALSH, 2022). Além disso, políticas de habitação social no século XX reforçaram essa situação de segregação espacial ao construir conjuntos habitacionais em bairros afastados, como é o caso de Cidade Tiradentes, no extremo Leste de São Paulo.

Segundo Villaça, esse processo, que levou ao estabelecimento dos “bairros residenciais das camadas de alta renda”, foi posteriormente capturado pelo capital imobiliário. Grandes incorporadoras se valeram de forte *lobby* para direcionar os investimentos públicos em infraestrutura urbana, sobretudo de transportes, de modo a realizar elevados lucros a partir de movimentos especulativos, nos quais glebas previamente adquiridas e de difícil acesso aos centros de negócios das cidades repentinamente se valorizaram após a construção de uma ponte ou túnel (VILLAÇA, 2001, cap. 8).

No entanto, segundo o autor, cumpre ressaltar essa capacidade de direcionamento do capital imobiliário é limitada pela própria estrutura urbana, sobretudo a acessibilidade aos centros de negócios. Dessa forma, mesmo esses novos bairros de elite produzidos pelo capital imobiliário tendem a situar-se de forma contígua – ou com rápido acesso por automóvel – aos bairros tradicionais.

Em São Paulo, o autor identifica essa região de concentração dos locais de residência da elite como o setor Sudoeste da cidade (VILLAÇA, 2001, p. 116-118), que ele representa de forma esquematizada na Figura 3.7. Sob essa perspectiva, mesmo os novos bairros de elite da cidade, supostamente localizados em subúrbios situados em municípios distantes do centro urbano, seguem o mesmo padrão. Esse é o caso de Itapacerica da Serra, Cotia e Barueri, que são contíguos ao setor Sudoeste do município de São Paulo.

Entretanto, ao invés de apenas referenciar o trabalho de Villaça (2001), neste estudo, utilizamos método estatísticos para atualizar, com base em

Figura 3.7 – Setor sudoeste - área de residência da elite.

Fonte: (VILLAÇA, 2001, p. 115)

evidências, a localização desta área de concentração da elite da cidade de São Paulo a partir da década de 2010. Tal medida é necessária pois, como argumenta o próprio autor, ainda que haja forte continuidade histórica na localização dos bairros nobres, novos bairros se consolidam ao longo do tempo, com o aumento da população e crescimento da mancha urbana. Do mesmo modo, com a mudança da localização do **CBD** – que em São Paulo, por exemplo, migrou do Largo do Café, no centro histórico, para a região da Paulista, em seguida Avenida Berrine e nos últimos anos tem se direcionado para a Faria Lima – é possível que bairros originalmente de elite se desvalorizem, como aconteceu com os Campos Elísios. Para isso, nos valem os microdados do Censo Demográfico de 2010, assim como da Pesquisa Origem e Destino de 2017 realizada pelo Metrô de São Paulo. Essas fontes de dados foram detalhadas, respectivamente, nas seções **C.2.1** e **C.2.2** do Anexo **Único**.

Os dados do Censo Demográfico de 2010 foram agrupados por distrito de modo a gerar as seguintes variáveis:

- 1. A proporção de pessoas brancas residentes no distrito, em relação ao total da população residente no mesmo;

- 2. A proporção de responsáveis de domicílio alfabetizados no distrito, em relação ao total de domicílios particulares situados no distrito;
- 3. A proporção de domicílios com renda *per capita* superior a 10 salários mínimos, em relação ao total de domicílios particulares situados no distrito;
- 4. A média da renda nominal média dos responsáveis por domicílio dos setores censitários situados no distrito;

Em conjunto, essas variáveis têm por objetivo realizar uma caracterização da desigualdade sociodemográfica nos distritos da cidade, compreendendo as dimensões da segregação racial (1), acesso à educação básica (2) e concentração de renda (3 e 4).

Em relação à concentração de renda, foram utilizadas duas variáveis distintas pois há forte desigualdade interna em alguns distritos. Como é o caso da Vila Andrade, que contém a fronteira entre a favela de Paraisópolis e condomínios verticais de luxo⁵. Nesse cenário, apenas a renda média não seria suficiente para capturar a concentração de domicílios com alta renda, sobretudo pois as favelas possuem densidade demográfica muito superior aos bairros de alto padrão, o que resulta na redução dessa variável. Ao mesmo tempo, considerar apenas os domicílios com renda maior que 10 salários mínimos não permitiria diferenciar distritos que concentram bairros de residência de famílias de classe média-baixa daqueles nos quais se situam as pessoas de baixa renda mais vulneráveis. Pois em ambos os casos essa proporção tenderia a zero, apesar de, do ponto de vista socioeconômico, serem situações muito distintas.

Os dados provenientes da Pesquisa Origem e Destino, por sua vez, foram utilizados para calcular a média de duração das viagens pendulares⁶ que partem de cada distrito. Essa variável tem por objetivo medir a acessibilidade dos moradores do distrito ao mercado de trabalho da cidade, conforme previsto no modelo SMM da economia urbana, descrito no Capítulo 2. Conceitualização que também está presente em Villaça (2001, cap.

⁵É nele que se localiza a famosa foto do edifício com apartamentos um por andar e piscinas na varanda logo ao lado de uma área de ocupação irregular com domicílios autoconstruídos.

⁶Isto é, as viagens da residência ao local de trabalho ou do local de trabalho à residência.

5), quando o autor estabelece a relação entre os bairros de residência da elite e a proximidade ao por ele denominado “centro principal”.

A metodologia para cálculo dessas variáveis é detalhada na seção [D.6](#) do Anexo Único. A escolha das datas de referência das bases de dados (2010 e 2017 para o Censo e a pesquisa OD, respectivamente) deve-se ao fato de que ambas consistem nos dados disponíveis para cada uma dessas pesquisas mais próximos a 2014, ano base deste estudo. Há que se pontuar, no entanto, que há distância considerável entre a caracterização socioeconômica e a medição dos tempos de deslocamento. Trata-se de uma limitação do presente estudo, que é no entanto minimizada pela continuidade histórica do fenômeno aqui estudado, que, como mostra o trabalho de Villaça (2001), não se altera substancialmente em menos de uma década. Ademais, ressalta-se que os resultados alcançados neste estudo confirmam em larga medida as afirmações do autor em 2001, apresentando mais uma evidência da permanência do fenômeno estudado.

3.2.5.1 Análise de Componentes Principais

As variáveis descritas nesta seção⁷ foram então padronizadas por meio do método *Z-scores* referido anteriormente na seção [3.2.3](#), conforme a equação $z_i = \frac{x_i - \mu}{s}$. Uma vez padronizadas, foi ajustada uma Análise de Componentes Principais (do inglês, *Principal Component Analysis*) ([PCA](#)) para redução de dimensionalidade. Essa metodologia consiste em um mapeamento linear dos dados para um espaço com menor dimensionalidade de forma que a variância na representação de menor dimensionalidade é maximizada. O componente principal, primeira dimensão desse espaço, consiste na combinação linear das variáveis originais que é capaz de explicar a maior proporção da variância das mesmas. Os demais componentes são ortogonais ao primeiro (isto é, não há correlação linear entre eles) e explicam porções progressivamente menores da variância restante.

A [PCA](#) é obtida por meio do cálculo dos autovetores da matriz de correlação das variáveis originais. Esses autovetores são ordenados em uma matriz, aqui identificada por W , de acordo com seus autovalores, do maior

⁷Qual seja, a caracterização sociodemográfica dos distritos o e tempo médio de deslocamento a trabalho

ao menor. O autovetor de maior autovalor corresponde ao componente principal e é portanto W_1 . A redução de dimensionalidade para k dimensões é obtida a partir da multiplicação da matriz correspondente aos k primeiros autovetores ordenados por seus autovalores, W_k , pela matriz de dados originais X , conforme a equação $T_k = XW_k$. A matriz de dados transformada resultante T_k terá a quantidade original de linhas n mas apenas k colunas. Cada uma dessas colunas é uma combinação linear de todas as variáveis originais que corresponde a dimensões ortogonais entre si que representam porções progressivamente menores da covariância total dos dados. Neste estudo, utilizamos a primeira dessas dimensões, que corresponde assim à multiplicação da matriz de dados original pelo componente principal, na forma $T_1 = XW_1$.

Essa metodologia permitiu tornar a medida de privilégio social construída a partir das cinco dimensões distintas descritas nesta seção em uma métrica univariada que combina todas elas. Isso permite comparar diretamente os distritos da cidade de modo ranqueá-los, com os valores mais elevados dessa métrica correspondendo aos distritos mais privilegiados e os mais reduzidos aos mais vulneráveis (GREENACRE et al., 2022).

A variável construída segundo essa metodologia, que possui 96 registros, um para cada distrito da cidade de São Paulo, foi então georreferenciada por meio do *join* com a camada de distritos disponibilizada no GeoSampa. Os dados foram plotados em um mapa coroplético, que permitiu identificar visualmente a existência de áreas de alta concentração de privilégio na cidade, assim como áreas de alta vulnerabilidade. Com o objetivo de confirmar a hipótese de concentração espacial advinda da análise exploratória realizada por meio do mapa coroplético, foi então realizada a análise de Indicadores Locais de Associação Espacial (do inglês, *Local Indicators of Spatial Association*) (LISA) com o uso do método Moran Local.

3.2.5.2 Indicadores Locais de Associação Espacial

Proposta originalmente por Anselin (1995), esse método não supervisionado permite identificar *clusters* espaciais a partir do cálculo da covariação entre a variável de interesse para determinado registro e sua rede de vizinhança, de acordo com seguinte equação:

$$I_i = (X_i - \bar{X}) \sum_{j=1}^n w_{ij} (X_j - \bar{X}) \quad (3.1)$$

Nessa equação, X_i representa o valor de X para o i -ésimo registro e \bar{X} a média de X . A matriz W_{ij} , por sua vez, representa a matriz de vizinhança, indexada para o registro i e seus vizinhos, representados por j , de modo que X_j representa o valor de X para o registro j . A matriz de vizinhança w_{ij} consiste em uma matriz quadrada de dimensões $n \times n$, onde n é o número de registros da base de dados e cujos valores indicam a vizinhança entre os registros indexados por i e j . O valor resultante, denominado Moran Local, indica o tipo de associação espacial da rede de vizinhança de um determinado registro. Como o nome indica, essa estatística é uma especialização local da estatística I de Moran (ANSELIN, 1995).

A I de Moran busca mensurar, para a totalidade dos dados, se há autocorrelação espacial, isto é, se registros próximos possuem valores similares. Um valor positivo para a estatística I de Moran indica que valores elevados tendem a estar circundados de valores elevados, e valores reduzidos também tendem a estar circundados de valores reduzidos. Já um valor negativo indica que valores elevados tendem a estar circundados de valores reduzidos, e vice-versa. Um valor próximo de zero, por fim, indica que não há autocorrelação espacial significativa, de modo que a distribuição espacial dos dados se assemelha à permutação aleatória dos mesmos na rede de vizinhança (MORAN, 1948). A equação a seguir define o cálculo dessa estatística, que utiliza as mesmas variáveis descritas anteriormente para a Moran Local, com exceção da variável W , que representa a soma-tória de todos os pesos da matriz de vizinhança w , conforme a equação $W = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$.

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum_{i=1}^n (x_i - \bar{x})^2} \quad (3.2)$$

As matrizes de vizinhança binárias – isto é, que indicam se há ou não relação de vizinhança entre i e j – os valores podem ser 1 ou 0. Valores iguais a 1 indicam a vizinhança entre i e j , enquanto valores iguais a 0 indicam que ela não ocorre. Por padrão, a diagonal das matrizes de vizinhanças binárias é preenchida por zeros, pois um registro não é considerado vizinho

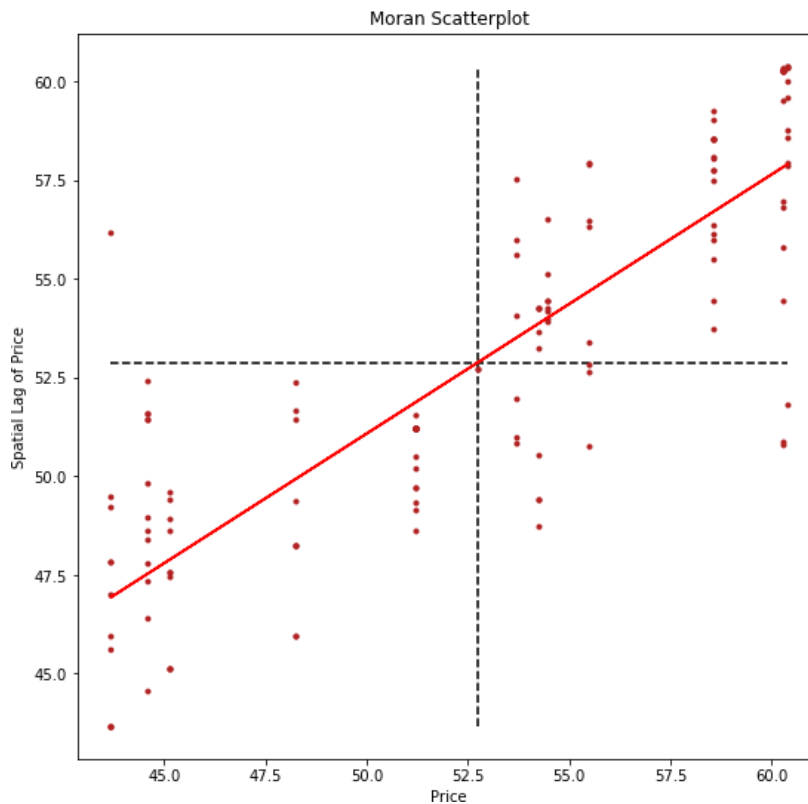
de si mesmo. Além disso, as matrizes de vizinhança podem ser padronizadas por linha, de modo que a soma de todos os pesos para todos os vizinhos de um determinado registro i seja igual a 1, isso é $\sum_{j=1}^n W_{ij} = 1$. No caso de matrizes de vizinhança binárias, essa padronização requer apenas a multiplicação de todos os valores por $\frac{1}{n}$.

A matriz de vizinhança utilizada neste estudo consiste em uma matriz binária do tipo rainha, isto é, que aceita vizinhanças em todos os sentidos, inclusive nas diagonais. Essa matriz foi construída a partir das relações de contiguidade espacial entre os polígonos que representam os limites dos distritos da cidade de São Paulo. Em outras palavras, distritos que compartilham fronteiras uns com os outros são considerados vizinhos, enquanto distritos que não são contíguos não são considerados vizinhos.

É possível realizar a defasagem espacial da variável de interesse, que consiste na média, ponderada pela matriz de vizinhança, de seu valor, conforme a equação $y_i = \frac{\sum_{j=1}^n W_{ij} \times y_j}{n}$. A plotagem de um mapa de dispersão com a variável de interesse padronizada no eixo x e seu valor padronizado defasado espacialmente no eixo y permite identificar diferentes tipos de agrupamentos espaciais. A Figura 3.8 apresenta um exemplo desse gráfico, denominado Gráfico de Dispersão de Moran (ANSELIN, 1996) ⁸. Pontos nesse gráfico de dispersão situados no quadrante que possui valores elevados para os eixos x e y correspondem a agrupamentos do tipo *Alto-Alto*. Ou seja, a registros que possuem valores acima da média para a variável de interesse e que são cercados de vizinhos que também possuem valores elevados.

Já os agrupamentos de tipo *Baixo-Baixo* correspondem ao quadrante oposto, que é definido por valores baixos para os eixos x e y . Trata-se, como é evidente, de registros que possuem valores abaixo da média para a variável de interesse e cujos vizinhos encontram-se na mesma situação. Há ainda dois outros tipos de *clusters*, *Alto-Baixo*, que correspondem a registros com valores acima da média cercados de valores abaixo da média, e *Baixo-Alto*, que é o oposto: registros com valores abaixo da média com vizinhos acima da média.

⁸Como demonstra Anselin (1996), o valor para do coeficiente angular da regressão linear de y (a variável defasada espacialmente) em x (a variável em seus valores normais) corresponde à estatística I de Moran (MORAN, 1948)

Figura 3.8 – Gráfico de dispersão de Moran – exemplo

Fonte: Documentação da biblioteca PySal - (REY; ANSELIN, 2007). Link: <https://pysal.org/esda/notebooks/spatialautocorrelation.html>

Por meio da realização de simulações (permutações aleatórias da variável de interesse condicionadas à matriz de vizinhança), a análise LISA permite identificar quais dos registros localizados nesses *clusters* espaciais possuem significância estatística (ANSELIN, 1995). Neste estudo, selecionamos os *clusters* que possuem mais de 95% de confiança, ou seja, nos quais o valor de I_i teria ocorrido aleatoriamente em apenas 5% dos casos, considerando 1000 permutações.

Como será demonstrado na seção 3.3, essa metodologia permitiu identificar um *cluster Alto-Alto* significativo, ou seja, uma região de alta concentração de privilégio na cidade de São Paulo, que foi denominada *área nobre*, com o objetivo de referenciar o termo utilizado pelo mercado imobiliário da cidade. Foram também identificados 3 outros *clusters Baixo-Baixo*, que cor-

respondem a regiões de concentração de vulnerabilidade social, que foram denominadas *áreas periféricas*.

Foi construída então uma variável binária do tipo *dummie* que identifica os distritos que estão situados no *cluster Alto-Alto* correspondente à *área nobre*. Outra variável *dummie* similar foi construída para identificar os distritos situados na *área periférica*. Os demais distritos foram denominados como situados na *área média*, referenciando o conceito socioeconômico de *classe média* como os indivíduos que se situam entre a elite e a população de baixa renda. Não foi construída uma variável *dummie* para esse caso, pois ele corresponderá ao caso de referência.

O Diagrama 3.9 abaixo resume a metodologia desenvolvida na identificação das áreas nobre e periféricas da cidade:

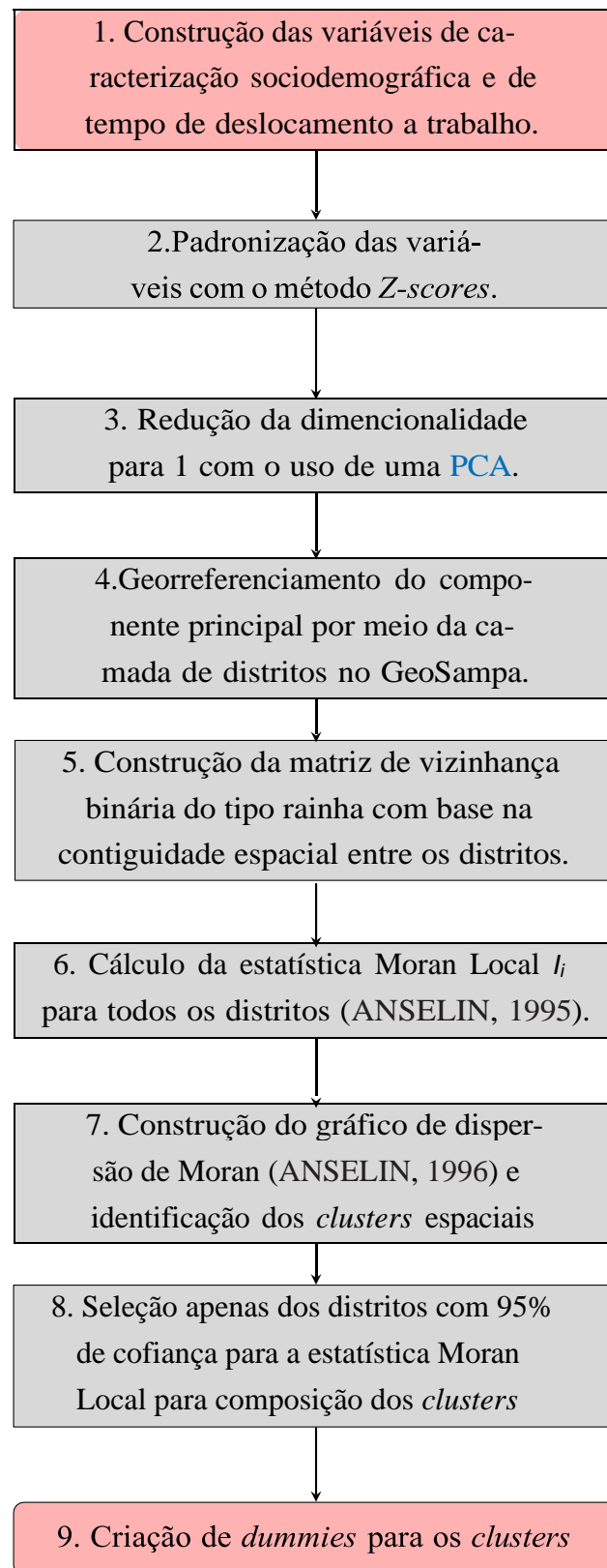


Figura 3.9 – Identificação dos clusters espaciais da área nobre e das áreas periféricas.

3.2.6 Resumo da base de dados

A base de dados final resultante de todos os processos descritos nas seções 3.2.1, 3.2.3, 3.2.4 e 3.2.5, acompanhada do dicionário de variáveis, foi descrita em detalhes na seção E.1 do Anexo Único. O Quadro 3.1 resume o papel de cada uma das variáveis utilizadas nos modelos estatísticos desenvolvidos neste estudo:

Quadro 3.1 – Variáveis utilizadas na modelagem

Categoria	Variáveis
Variável dependente	qt_novas_unidades
Tratamento	ZEU
Grupo	area_nobre, area_periferica
Controles potencial construtivo	valor_medio_m2_terreno, idade_media_imoveis_em_2014, idade_media_contribuintes_em_2014, media_qtd_pavimentos, prop_maior_terreno_nao_vertical, prop_area_total_edificada_horizontal, prop_area_total_edificada_vertical, prop_area_total_terrenos_e_galpoes_etc, qd_area
Controles restrições edificação	area_envoltoria_tombamento, tombo, bairro_ambiental, area_contaminada, area_protecao_ambiental, protecao_aos_mananciais, restricao_geotecnica, risco_geologico, rppn, unidade_de_conservacao, rampa_heliponto

Fonte: Elaboração Própria

A variável dependente dos modelos é a quantidade de unidades de tipo vertical produzida por quadra no período 2014-2023. A indicação de se a quadra encontra-se em ZEU representa o tratamento cujos efeitos se objetiva mensurar. E há dois conjuntos de controles: um primeiro conjunto, referente ao potencial construtivo da quadra, descrito na seção 3.2.3, e um segundo conjunto, referente à presença de restrições à edificação na quadra, descrito na seção 3.2.4. Além disso, há duas variáveis *dummie* que identificam o pertencimento da quadra a um agrupamento (*cluster*), qual seja a área nobre ou as áreas periféricas da cidade, como descrito na seção 3.2.5.

Como mencionado na seção 3.2.3 apenas as variáveis de controle referentes ao potencial construtivo da quadra, que são variáveis métricas, foram padronizadas com o método *Z-scores*. As variáveis de controle referentes à presença de restrições à edificação não foram padronizadas pois trata-se de variáveis binárias, que foram codificadas como *dummies* na seguinte forma:

$$X_i = \begin{cases} 1 & \text{se a condição é verdadeira,} \\ 0 & \text{caso contrário.} \end{cases} \quad (3.3)$$

De modo que $X_i \in \{1, 0\}$. O valor de referência para essas *dummies* representa o caso no qual não há nenhuma restrição à edificação. As variáveis que indicam o pertencimento a um *cluster*, nomeadamente a área nobre ou as áreas periféricas, também foram codificadas como *dummies*, de forma que o caso base representa o não-pertencimento a qualquer desses dois *clusters* – ou seja, que a quadra encontra-se na área de classe média. Por fim, a variável dependente do modelo não foi padronizada ou transformada, tendo sido mantida na escala original. Isso foi realizado com o objetivo de tornar a interpretação dos coeficientes dos modelos semanticamente mais próxima da realidade estudada, que será em termos da quantidade esperada de unidades de tipo vertical produzidas no período.

3.2.7 Modelos estatísticos

Os dados apresentados na seção 3.2.6 foram analisados neste estudo com o uso de modelos estatísticos *bayesianos*. Ainda que haja diversas razões epistemológicas para se utilizar métodos *bayesianos* (CLAYTON, 2022; MCELREATH, 2020), essa dissertação não tem por escopo adentrar essa seara. Neste capítulo, os métodos *bayesianos* são utilizados devido à melhor adequação dos mesmos ao problema de pesquisa. Pois, como será demonstrado, o problema de pesquisa requer o ajuste de um modelo multinível com poucos grupos (apenas 3), sendo que eles são desbalanceados, havendo disparidade na quantidade de amostras pertencentes a cada grupo. Esse cenário é bastante desafiador para modelos frequentistas, resultando em vieses nas estimativas dos coeficientes do modelo (MAAS; HOOX, 2005), o que não acontece para estimadores *bayesianos*.

Na subseção 3.2.7.1, que segue, a variável dependente do estudo é caracterizada à luz da metodologia de modelagem mais adequada a ela. Em seguida, cada um dos modelos desenvolvidos é detalhado em uma subseção própria. A metodologia aqui empregada segue o fluxo de trabalho *bayesiano* (*bayesian workflow*) (GELMAN et al., 2020). Esse método consiste em:

- Iniciar com um modelo matematicamente mais simples;
- Utilizar dados simulados para verificar se a especificação do modelo, principalmente as distribuições *prior*es estão capturando adequadamente o fenômeno⁹;
- Caso a especificação se mostre adequada, ajustar o modelo aos dados (se não, reespecificar);
- Verificar se o modelo convergiu e não houve problemas computacionais (se não, reespecificar o modelo, transformar os dados ou ajustar os hiperparâmetros do amostrador para alcançar a convergência);
- Utilizar validação cruzada para verificar os resultados do modelo;
- Aceitar o modelo provisoriamente e interpretar os resultados das distribuições posteriores;
- Especificar um modelo mais complexo e voltar ao passo 1. Em seguida, comparar esse modelo com o modelo desenvolvido anteriormente – manter provisoriamente o modelo com melhores resultados.

Ao longo dessa seção, um mesmo exemplo – a loja de sorvetes – é utilizado para ilustrar as razões pelas quais cada um dos modelos, progressivamente mais complexos, foram escolhidos para modelar o problema de pesquisa.

3.2.7.1 Tipificação da variável dependente

Como apresentado na seção 3.2.6, a variável dependente a ser modelada neste estudo representa a quantidade de unidades produzidas por

⁹Método que é denominado *prior predictive check* por Gelman et al. (2020) e McElreath (2020)

quadra em empreendimentos verticais no período 2014-2023. Uma vez que não estamos considerando os imóveis demolidos para dar lugar às novas unidades, a variável consiste portanto em números inteiros positivos iniciados em zero. Pois é possível que não tenha sido construído nenhum empreendimento vertical na quadra, de modo que o total de unidades produzida era igual a zero. Isto é, $Y = \{0, 1, 2, 3, \dots\}$.

Caso o estudo considerasse a diferença na quantidade de unidades habitacionais, levando em conta não apenas as unidades produzidas mas aquelas que foram demolidas para dar lugar aos empreendimentos (por exemplo, as casinhas que deram origem ao prédio), seria possível haver resultados negativos. Isso ocorre sobretudo nos movimentos especulativos descritos na seção D.1 do Anexo Único que têm como resultado a manutenção de terrenos vazios por vários anos, em geral sediando serviços de estacionamento de automóveis.

No entanto, como o estudo tem por objetivo avaliar a capacidade do planejamento urbano em direcionar a produção imobiliária, a variável dependente aqui considerada é mais adequada, pois a quantidade bruta de unidades produzidas é uma medida mais sensível ao fenômeno que se quer explicar. Considere-se, por exemplo, os seguintes casos: (a) vários imóveis residenciais com terrenos de baixa metragem são demolidos para dar lugar a um empreendimento e (b) apenas um grande galpão é demolido, com a mesma área total de terreno das casas e dando lugar ao mesmo empreendimento. Do ponto de vista da mensuração da produção da atividade imobiliária, (a) e (b) são iguais, uma vez que o que importa para nosso estudo é a quantidade de unidades que esse edifício lançou no mercado. Nesse cenário, descontar as casas demolidas resultaria em diferenciar dois empreendimentos que contém a mesma quantidade de unidades, mas que foram construídos em locais com diferentes padrões de uso do solo¹⁰.

Outra medida possível – que é utilizada no Capítulo 4 – é considerar a área construída total acrescida por esses empreendimentos, medida como a somatória das áreas construídas de cada uma de suas unidades. Essa variável é mais sensível ao investimento realizado pelo mercado imobiliário

¹⁰Note-se inclusive que, dependendo da região da cidade, o galpão do empreendimento (b) pode até mesmo ter um valor por metro quadrado de terreno superior ao das casas do empreendimento (a).

do que a quantidade de unidades, pois permite diferenciar empreendimentos com produtos imobiliários de padrões distintos. Considere-se, por exemplo, (a) um edifício com 4 andares e 6 apartamentos por andar e (b) um edifício de 24 andares, mas no padrão um por andar, ambos construídos em terrenos de mesmo tamanho. Evidentemente o edifício de alto padrão (b) implica em um maior investimento do mercado imobiliário. Mas ambos (a) e (b) possuem a mesma quantidade de unidades e seriam portanto idênticos caso seja essa a medida utilizada.

No entanto, este estudo tem por objetivo mensurar especificamente a capacidade do Plano Diretor Estratégico de São Paulo em direcionar a produção imobiliária para as Zonas de Estruturação Urbana. Como detalhado na seção D.2, o PDE estabelece uma série de dispositivos e incentivos que foram calibrados com a intencionalidade de gerar uma maior produção de unidades – em especial de habitação. De modo que a avaliação desta política deve considerar seus efeitos sobre a quantidade de unidades e não sobre a área construída total.

A quantidade de unidades produzida foi mensurada para unidades espaciais bem definidas – as quadras da cidade – e por um tempo determinado de exposição, que corresponde ao período entre 2014 e 2023. Como foi calculado o total de unidades construído em cada quadra ao longo desse período, o banco de dados do estudo não possui amostras repetidas (cada quadra está registrada uma única vez). Além disso, as quadras possuem fronteiras bem definidas, não se confundindo umas com as outras. Por essas razões, podemos afirmar que a variável dependente consiste em dados de contagem, sendo portanto definida por $y \in \mathbb{N}_0$.

A modelagem estatística deste estudo terá assim por objetivo avaliar o efeito da variável de tratamento, todos os demais controles listados acima mantidos iguais, sobre a quantidade esperada de unidades produzidas em cada uma das quadras da cidade no período 2014-2023. Nesse cenário, como a variável dependente consiste em dados de contagem e a amostragem é realizada sobre unidades espaciais bem definidas sob um mesmo tempo de exposição, os modelos da família de Modelos Lineares Generalizados (do inglês, *Generalized Linear Models*) (GLM) com função de ligação logarítmica apresentam-se como uma primeira aproximação adequada para modelagem do problema de pesquisa.

Para esta família de modelos, os valores esperados do vetor da variável dependente y , condicionados às variáveis de controle e tratamento representadas pela matriz X , são concebidos como a média μ de uma distribuição da família exponencial. Essa média é obtida por meio de uma função de ligação, em geral representada por sua função inversa, g^{-1} , aplicada sobre resultado da combinação linear de um vetor de parâmetros desconhecidos β com a matriz de dados X , conforme a seguinte equação:

$$E(y|X) = \mu = g^{-1}(\beta X) \quad (3.4)$$

Os parâmetros desconhecidos β são então estimados de acordo com diferentes técnicas. Neste estudo, é realizada a atualização *bayesiana* das distribuições *a priori* para os parâmetros β com o uso do Amostrador sem curvas U (do inglês, *No U-turn Sampler*) ([NUTS](#)). Esse amostrador constrói correntes de Markov utilizando uma técnica de amostragem baseada no método de Hamiltonian Monte Carlo, contudo sem requerer a definição dos hiper-parâmetros de tamanho da passada ϵ e quantidade de passos L (HOFFMAN; GELMAN, 2011), consistindo atualmente no estado da arte dos amostradores *bayesianos*.

A aplicação do *vorflow bayesiano* teve como resultado a progressiva complexificação dos modelos [GLM](#) para dados de contagem inicialmente ajustados, de modo a aproximá-los melhor dos dados e da realidade do problema de pesquisa. Dessa forma, foram desenvolvidos também modelos de mistura e modelos multinível, como será detalhado nas seções a seguir.

3.2.7.2 Poisson

No caso dos modelos para dados de contagem, a função de ligação g indicada é o logaritmo natural, que tem como função inversa g^{-1} a função exponencial. O modelo mais comumente utilizado, denominado Regressão de Poisson, modela o valor esperado de Y condicionado a X como o parâ-

metro λ da distribuição de Poisson, de modo que:

$$\begin{aligned} y_i &\sim \text{Poisson}(\lambda_i) \\ \log(\lambda_i) &= X_i\beta \\ \lambda_i &= e^{X_i\beta} \end{aligned} \quad (3.5)$$

Essa função de ligação possui dois principais benefícios. Em primeiro lugar, a aplicação da função exponencial com base e sobre o preditor linear ($e^{X\beta}$) permite que o domínio desse preditor, que pertence ao conjunto dos números reais $R = (-\infty, +\infty)$, seja mapeado para o contradomínio dos números reais positivos $R = (0, \infty)$. Dessa forma, o valor esperado λ previsto pelo modelo será sempre um número real positivo, o que permite que ele represente melhor os dados de contagem, que também são positivos¹¹. Isso evita previsões negativas, que são descabidas para esse tipo de dado: por exemplo, não faz sentido prever que haverá -3 gols em uma partida de futebol. Além disso, o fato de ser utilizado um número real e não um número natural ($N_0 = \{0, 1, 2, 3, \dots\}$) possibilita que o modelo realize previsões com maior acurácia ao permitir erros menores que 1¹². Por essas razões, é comum definir com distribuição *a priori* pouco informativa para os parâmetros β a distribuição Normal com média zero e desvio padrão um, sobretudo caso os dados tenham sido previamente padronizados. Mas, quando há conhecimento prévio que justifique, é possível também estabelecer outros valores de μ e σ^2 para as *priors*.

$$\forall \beta_i \in \beta, \beta_i \sim N(\mu_i, \sigma_i^2) \quad (3.6)$$

Em segundo lugar, essa função de ligação modela os coeficientes do preditor linear como efeitos multiplicativos. Realizada a exponenciação, isso permite interpretar os coeficientes como um fator de aumento que, aplicado sobre a taxa basal representada pelo intercepto (coeficiente β_0 ou α), resulta no valor esperado de Y . Considere-se o seguinte exemplo: o intercepto de um modelo ajustado corresponde a 2, de modo que o modelo prevê uma taxa basal de em média aproximadamente 7,4 ocorrências, pois $\exp(2) = 7,389$.

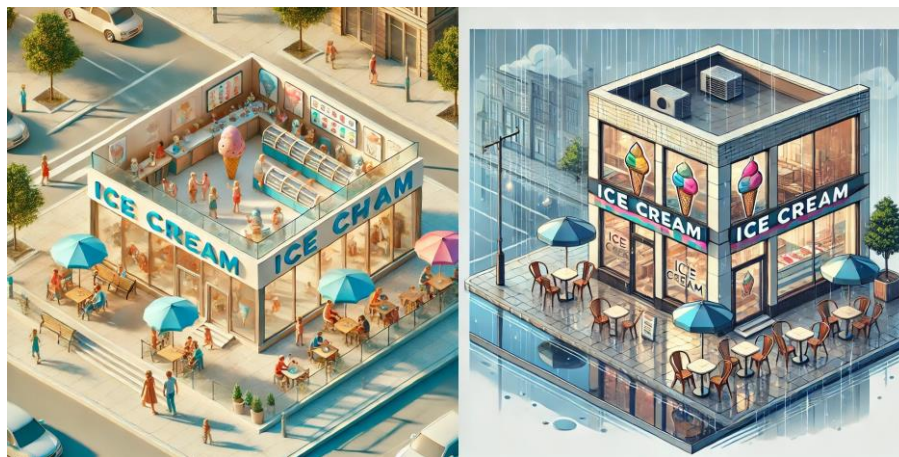
¹¹Além de ser um parâmetro válido para a distribuição de Poisson, que requer que λ seja real positivo.

¹²Por exemplo, é melhor prever 4,8 quando o valor é 5, do que 4

Nesse caso, um coeficiente angular β_1 equivalente a 0,2 corresponderia então a um aumento de aproximadamente 22% no valor esperado, pois $\exp(0.2) = 1,221$, resultando em um valor previsto de aproximadamente 9,2.

Essa interpretação em termos de um fator de aumento é bastante aderente à maior parte dos problemas de pesquisa associados a dados de contagem. Para ilustrar essa questão, considere-se uma base de dados que contém a informação da quantidade de clientes em um mesmo dia do ano para diferentes lojas de sorvete pertencentes a uma multinacional com presença em diversos países. Essas lojas estão representadas na Figura 3.10.

Figura 3.10 – Exemplo: lojas de sorvete.



Fonte: Elaboração própria com o auxílio do ChatGPT

Podemos imaginar uma base de dados que representa essas e outras lojas de sorte de uma mesma cadeia internacional conforme o Quadro 3.2 apresentado a seguir. Essa base contém a medição da quantidade de clientes em diferentes lojas dessa cadeia localizadas em diversas localidades no mundo em um determinado dia do ano. Ela contém, além do identificador da loja e a contagem da quantidade de clientes que visitou a loja nesse dia, uma variável que identifica se choveu naquela localidade naquele dia.

Quadro 3.2 – Exemplo: base de dados de lojas de sorvete

id_loja	qtd_clientes	chuva
1	23	0
2	3	1
...

Fonte: Elaboração Própria

Uma regressão de Poisson pode ser ajustada para essa base de dados de acordo com a seguinte equação:

$$\begin{aligned}
 y_i &\sim \text{Poisson}(\lambda_i) \\
 \log(\lambda_i) &= X_{i0}\beta_0 + X_{i1}\beta_1 \\
 \lambda_i &= e^{\beta}
 \end{aligned}
 \tag{3.7}$$

Na qual β_0 representa o intercepto, isto é, a taxa basal de clientes que cada loja tende a receber por dia. Enquanto β_1 representa o efeito da variável *booleana* que indica se houve chuva sobre essa taxa basal de clientes. Suponhamos que, após a realização da amostragem das correntes de Markov, os valores médios para as distribuições *posteriori* dos parâmetros β_0 e β_1 correspondam, respectivamente a 3 e $-1,8$. Isso significa que a taxa basal de clientes esperada em um dia é de aproximadamente 20 clientes, pois $e^3 = 20,085$. No entanto, caso esteja chovendo, há uma queda de aproximadamente 84% nesse valor, pois $e^{-1,8} = 0,16$. Assim, a equação completa para o valor previsto do modelo, $\lambda_i = e^{3-1,8}$, determina que a taxa basal deverá ser multiplicada por 16%, resultando em um valor esperado de 3 clientes apenas. O que é natural, pois as pessoas não tendem a desejar tomar sorvete em dias de chuva.

Esse efeito multiplicativo é preferível a um efeito aditivo, que resultaria do ajuste de uma regressão linear simples. Como se pode notar pela Figura 3.10 ambas as lojas possuem uma capacidade total de clientes distinta – a loja da direita possui um salão no segundo andar. Nesse cenário, um efeito aditivo, que resultaria em uma redução absoluta no número de clientes (por exemplo, -10 clientes) não seria representativo dessa diferença de capacidade entre as lojas, podendo subestimar a queda no número esperado

de clientes da loja maior ou então superestimá-la para a loja menor.

3.2.7.3 Binomial Negativa

A distribuição de Poisson acarreta no entanto uma pressuposição forte para o modelo: a de que a média e a variância da variável prevista \hat{y} terão o mesmo valor. Esse pressuposto não se sustenta para muitos casos reais. Retornando ao exemplo das lojas de sorvete, imagine-se que, no dia em que os dados foram coletados, era feriado em algumas das localidades em que as lojas estão situados. De modo que a base de dados passará a ter o formato apresentado no Quadro 3.3.

Quadro 3.3 – *Exemplo: base de dados de lojas de sorvete*

id_loja	qtd_clientes	chuva	feriado
1	23	0	0
2	3	1	0
3	46	0	1
...

Fonte: Elaboração Própria

Como se pode ver pela base de dados no Quadro 3.3, as lojas situadas em uma localidade com feriado no dia em que os dados foram coletados receberam uma quantidade elevada de clientes nesse dia. Mesmo que apenas uma pequena quantidade de lojas se enquadre nesse caso, ainda assim isso resulta em um aumento considerável da variância em relação à média. Considere-se a equação para obtenção da variância, $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$. Como um de seus componentes é o quadrado do desvio em relação à média, valores extremos tendem a elevar o valor da variância numa proporção maior do que eles elevam a própria média, $\mu = \frac{1}{n} \sum_{i=1}^n x_i$.

Uma vez que a distribuição de Poisson pressupõe que a média e a variância são iguais, esse modelo tenderá a subestimar os valores previstos para as lojas dos locais onde era feriado, aumentando o erro para esses casos. Uma solução para essa questão é modelar os dados utilizando outra distribuição adequada para dados de contagem, a Binomial Negativa. Essa distribuição foi originalmente concebida para modelar o número de tentati-

vas necessário para obter uma quantidade determinada de sucessos em um número definido de experimentos de Bernoulli – por exemplo, o número de tentativas até se obter 5 caras em 10 jogos de uma moeda fiel. Assim como a Poisson, ela é uma distribuição de probabilidade discreta, sendo portanto adequada para dados de contagem.

A distribuição Binomial Negativa pode ser especificada pela seguinte equação:

$$y_i \sim \text{NegBinom}(\mu_i, \phi) \quad (3.8)$$

Na qual μ_i representa a média esperada de contagens para a i -ésima observação e ϕ representa um parâmetro de dispersão. A relação entre a média μ_i e a variância é dada pela equação $\text{Var}(y_i) = \mu_i + \frac{\mu_i^2}{\phi}$. Dessa forma, um parâmetro ϕ elevado indica que há baixa dispersão nos dados, sendo que quando se aproxima do infinito, $\phi \rightarrow \infty$, a distribuição Binomial Negativa equivale à distribuição de Poisson. Uma regressão *bayesiana* para dados de contagem que utilize a Binomial Negativa para definição da função de verossimilhança da variável dependente pode ser especificada da seguinte forma:

$$\begin{aligned} y_i &\sim \text{NegBinom}(\mu_i, \phi) \\ \log \mu_i &= X_i \beta \\ \mu_i &= e^{X_i \beta} \end{aligned} \quad (3.9)$$

Como o parâmetro ϕ deve ser positivo, é comum utilizar a distribuição Gamma como distribuição *a priori* para ele. Já as distribuições *a priori* para os parâmetros β podem ser representadas pela distribuição Normal, pois, assim como na regressão de Poisson apresentada na seção 3.2.7.2, a função de ligação logarítmica implica em sua exponenciação com a base e , de modo que eles serão mapeados de números reais para números reais positivos.

$$\begin{aligned} \forall \beta \in \beta, \beta_i &\sim N(\mu_i, \sigma^2) \\ \phi &\sim \text{Gamma}(\alpha, \theta) \end{aligned} \quad (3.10)$$

Por meio da atualização *bayesiana*, o ajuste do modelo estatístico per-

mitirá identificar a distribuição *a posteriori* de ϕ mais provável, condicionada à variável dependente y e à matriz de dados X , $P(\phi \mid y, X)$. Caso o valor ajustado de ϕ seja elevado, a distribuição *a posteriori* de y se aproximará da distribuição de Poisson, com média e variância com valores similares. No entanto, caso o valor de ϕ seja reduzido, a variância de y poderá ser superior à sua média. Essas propriedades fazem com que a regressão Binomial Negativa seja mais robusta para dados de contagem, sobretudo quando há hiperdispersão da variável dependente y – como é o caso de nosso exemplo, no qual a ocorrência de feriado resulta em valores extremados para a quantidade de clientes nas lojas.

3.2.7.4 Zeros Inflados

Ainda considerando o exemplo da cadeia de lojas de sorvete, considere-se que, no dia da coleta dos dados, houve lojas que não foram abertas ao público, de modo que elas não receberam nenhum cliente. Uma das razões para isso ter ocorrido é que algumas dessas lojas estavam passando por reformas nesse período, sendo que algumas dessas reformas – por exemplo, porque obrigaram os estabelecimentos a desligarem a câmara fria – impediram a abertura. Essa informação foi capturada em nossa base de dados, conforme exemplifica o Quadro 3.4.

Quadro 3.4 – Exemplo: base de dados de lojas de sorvete

id_loja	qtd_clientes	chuva	feriado	reforma
1	23	0	0	0
2	3	1	0	0
3	46	0	1	0
4	0	1	0	0
5	0	0	0	1
...

Fonte: Elaboração Própria

A consideração dos dados das lojas com números identificadores 4 e 5 indica que há duas diferentes razões para que uma loja não tenha recebido clientes no dia da coleta de dados:

- 1. A loja estava aberta, mas nenhum cliente foi até ela porque estava chovendo;
- 2. A loja não abriu, porque estava em reformas.

Há, portanto, dois processos gerativos diferentes que resultam em um valor de $y_i = 0$. O primeiro deles é denominado *processo de contagem*. Como apresentado nas seções 3.2.7.2 e 3.2.7.3, esse processo pode ser modelado por meio de uma regressão da família GLM com função de ligação logarítmica e função de verossimilhança definida pela distribuição de Poisson ou Binomial Negativa – sendo que essa última é mais robusta a resultados extremos. Esses modelos permitirão identificar variáveis que representam ocorrências ou fatores que *podem* resultar em valores esperados iguais a zero, $\hat{y}_i = 0$, mas não necessariamente, como é o caso do efeito da chuva para o exemplo das lojas de sorvete. Esses casos são denominados *zeros de contagem*.

O segundo processo gerativo que resulta no registro de valores iguais a zero para a variável dependente é denominado *processo estrutural*. Como no exemplo das lojas que não foram abertas ao público, esse processo gerativo representa situações que *necessariamente* resultam em valores $y_i = 0$, denominados *zeros estruturais*. Essas situações, representadas por S , podem ocorrer ou não. Trata-se assim de uma variável binária, de modo que $S \in \{0, 1\}$. A probabilidade de ocorrência $P(S = 1)$ pode ser modelada por meio de um modelo da família GLM para dados binários denominado Regressão Logística. Trata-se de um modelo linear com função de ligação logística e função de verossimilhança definida pela distribuição de Bernoulli. Esse modelo pode ser especificado da seguinte forma:

$$\begin{aligned} P(S_i = 1 \mid X_i) &\sim \text{Bernoulli}(p_i) \\ \text{logit}(p_i) &= \log\left(\frac{p_i}{1-p_i}\right) = X_i\beta \\ p_i &= \frac{1}{1+e^{X_i\beta}} \end{aligned} \quad (3.11)$$

A aplicação da função de ligação logit sobre a probabilidade de ocorrência $S_i = 1$, indicada por p_i , tem como resultado transformar essa probabilidade no logarítmo da chance (*odds*) de ocorrência. A chance é definida pela razão entre a probabilidade de ocorrência e a probabilidade de não

ocorrência, ou seja, $\frac{P(S=1)}{P(S=1)+P(S=0)}$. Como a variável é binária, de modo que $P(S=0) = 1 - P(S=1)$, essa equação pode ser reescrita como $\frac{P(S=1)}{1-P(S=1)}$.

A chance, O , consiste portanto em um número real positivo, de modo que $O \in \mathbb{R}^+$. A aplicação da função logarítmica sobre esse número permite mapeá-lo para o conjunto dos números reais, $\mathbb{R} = (-\infty, +\infty)$. Isso possibilitará à Regressão Logística representar o preditor linear, $X_i\beta$, que também pertence ao conjunto dos números reais, como o logaritmo da chance de ocorrência.

A equação $\frac{1}{1+e^{X_i\beta}}$ que representa a aplicação da função inversa da função logit, denominada logística, permite mapear o preditor linear novamente para uma probabilidade, de modo que a previsão do modelo, \hat{p} , estará sempre entre 0 e 1: $\hat{p} \in [0, 1]$. Para se obter o valor previsto \hat{y}_i , considerando que $y \in \{0, 1\}$, é definido um limiar de \hat{p} a partir do qual considera-se que há probabilidade suficiente para ocorrência, de modo que caso se alcance esse limiar, $\hat{y} = 1$ e caso não se alcance, $\hat{y} = 0$. Esse limiar é em geral parametrizado como 0, 5. A equação abaixo operacionaliza essa transformação:

$$(\hat{y}_i | X_i) = \begin{cases} 1 & \text{se } \frac{1}{1+e^{X_i\beta}} \geq 0,5 \\ 0 & \text{se } \frac{1}{1+e^{X_i\beta}} < 0,5 \end{cases} \quad (3.12)$$

Como a função logística é não-linear, a interpretação de um único coeficiente β_i é dificultada, pois não é possível inferir sua contribuição individual para a probabilidade prevista, \hat{p}_i , uma vez que essa probabilidade será resultado da combinação entre os efeitos dos diferentes β . No entanto, é possível exponenciar o coeficiente, e^{β_i} , o que resulta na interpretação deste como a razão de chances (*odds-ratio*). A razão de chances, como o próprio termo indica, é definida pela razão entre as chances de ocorrência de um evento e as chances de ocorrência de outro evento. No caso de variáveis binárias, nas quais há apenas dois eventos possíveis (ocorrência, $S = 1$ e não-ocorrência, $S = 0$), a razão de chances é definida pela equação:

$$OR = \frac{\left(\frac{P(S=1)}{1-P(S=1)}\right)}{\left(\frac{P(S=0)}{1-P(S=0)}\right)} = \frac{P(S=1)(1-P(S=0))}{P(S=0)(1-P(S=1))} \quad (3.13)$$

Caso a razão de chances seja igual a 1, $OR = 1$, a chance de ambos os

eventos ocorrerem é idêntica. Ou seja, há igual probabilidade de ocorrência e não-ocorrência, de modo que $P(S = 1) = 0,5$. Caso a razão de chances seja maior que 1, $OR > 1$, há aumento na probabilidade de ocorrência, na proporção da razão de chances. Se $OR = 1,65$, por exemplo, a chance de ocorrência, $O(S = 1)$, é 65% maior que a chance de não ocorrência, $O(S = 0)$. Caso a razão de chances seja menor que 1, $OR < 1$, há diminuição na chance de ocorrência, também na proporção da razão de chances. Assim, se $OR = 0,65$, por exemplo, a chance de ocorrência, $O(S = 1)$, será 35% menor que a chance de não-ocorrência, $O(S = 0)$.

Ao exponenciar os coeficientes, $e^{\beta x_i}$, eles passam a ter uma relação multiplicativa entre si, e não mais aditiva, da mesma forma que ocorre na regressão de Poisson apresentada na seção 3.2.7.2. De modo que se $OR = e^{\beta x_i}$, caso $e^{\beta_i} > 1$, um aumento de uma unidade variável X_i contribui positivamente para a ocorrência do fenômeno modelado. Caso, no entanto, $e^{\beta_i} < 1$, o efeito é negativo. Em ambos os casos, a magnitude do efeito pode ser interpretada como a proporção da contribuição de um aumento de uma unidade da variável X_i sobre a chance de ocorrência, $O(S = 1)$. Um coeficiente angular de 1,33, $\beta_i = 1,33$, corresponde assim a um aumento de 33% na chance de ocorrência para cada unidade de X_i . Por fim, se $e^{\beta_i} = 1$, o efeito é nulo.

Os Modelos de Contagem para Zeros-Inflados (ZIP¹³) consistem em modelos de mistura que unificam a modelagem dos *zeros estruturais* com a modelagem dos *zeros de contagem*. A modelagem dos *zeros estruturais* é realizada por meio da Regressão Logística, apresentada nos parágrafos anteriores. Já a modelagem dos *zeros de contagem* é feita por meio de regressões GLM para dados de contagem, como as regressões de Poisson e Binomial Negativa, apresentadas nas seções 3.2.7.2 e 3.2.7.3, respectivamente.

A Regressão Logística modelará então a probabilidade de ocorrência da situação que resultará necessariamente em $y_i = 0$, ou seja, o *processo estrutural*, representado pela variável latente binária S descrita anteriormente. Já a Regressão Poisson ou Binomial Negativa modelará a proba-

¹³Modelo de Poisson para Zeros-Inflados (do inglês, *Zero-Inflated Poisson*) (ZIP). Ainda que a sigla refira-se especificamente aos modelos de Poisson, ela é utilizada neste texto para referenciar os modelos de contagem para zeros-inflados de forma mais ampla, que inclui também as regressões Binomial Negativa, pois é esta a sigla que se consolidou na literatura.

bilidade de que o *processo de contagem* resulte em zero. McElreath (2020, cap. 12) apresenta esse modelo em detalhes. A combinação de ambas as verossimilhanças ocorre da seguinte forma:

- 1. Se $y_i = 0$, a probabilidade de 0 será definida pela probabilidade π de ocorrência do *zero estrutural*, somada à probabilidade de não-ocorrência do *zero estrutural*, ou seja, $1 - \pi$, multiplicada pela probabilidade de ocorrência do *zero de contagem* de acordo com a distribuição escolhida (Poisson ou Binomial Negativa).
- 2. Se $y_i > 0$, a probabilidade de y_i será definida pela probabilidade de não ocorrência do *zero estrutural*, $1 - \pi$, multiplicada pela probabilidade de ocorrência da contagem y_i de acordo com a distribuição escolhida.

Retomando o exemplo das lojas de sorvete, isso pode ser traduzido da seguinte forma:

- 1. A probabilidade de que a loja não tenha recebido nenhum cliente no dia da coleta de dados, $P(y_i = 0)$, é definida pela probabilidade π de que a loja não tenha sido aberta ao público naquele dia, somada à probabilidade $1 - \pi$ de que a loja tenha sido aberta ao público. A probabilidade de que a loja tenha sido aberta é então multiplicada pela probabilidade de que a quantidade de clientes y_i seja zero (ou seja, que de que a loja tenha sido aberta, mas que ainda assim não tenha recebido nenhum cliente no dia);
- 2. A probabilidade de que a loja tenha recebido uma dada quantidade de clientes, isto é, $P(y_i | y_i > 0)$, é definida pela probabilidade de que a loja tenha sido aberta ao público, $1 - \pi$, multiplicada pela probabilidade de que a loja tenha recebido essa quantidade de clientes y_i naquele dia;

A probabilidade de que a loja não tenha sido aberta, isto é, a probabilidade π de ocorrência do *zero estrutural* é então modelada por meio de uma Regressão Logística, que usa a distribuição de Bernoulli como função de verossimilhança. Já a probabilidade de que a loja tenha sido aberta mas

não tenha recebido nenhum cliente é modelada por meio de uma Regressão **GLM** para Dados de Contagem, que utiliza a distribuição de Poisson ou a Binomial Negativa como função de verossimilhança. Essa Regressão para Dados de Contagem é a mesma que modela a quantidade de clientes que a loja recebeu em geral, ou seja, o *processo de contagem* como um todo. Considerando-se que foi utilizada a distribuição Binomial Negativa, o modelo **ZIP** assume então a seguinte forma:

$$\begin{aligned}
 & \text{se } y_i = 0, \\
 & P(y_i = 0 \mid X_i) \sim \text{Bernoulli}(1 \mid \pi_i) + \text{Bernoulli}(0 \mid \pi_i) \times \text{NegBinom}(\mu_i, \phi) \\
 & \text{se } y_i = 1, \\
 & P(y_i = y_i \mid y_i > 0, X_i) \sim \text{Bernoulli}(0 \mid \pi_i) \times \text{NegBinom}(\mu_i, \phi)
 \end{aligned} \tag{3.14}$$

Os parâmetros π da distribuição de Bernoulli e μ são definidos da seguinte forma:

$$\begin{aligned}
 \pi_i &= \frac{1}{1 + e^{X_i \vartheta}} \\
 \mu_i &= e^{X_i \beta}
 \end{aligned} \tag{3.15}$$

Como os parâmetros ϑ e β compõem os preditores lineares, que por sua vez serão transformados por meio das funções de ligação para corresponderem ao domínio da variável Y , suas distribuições *a priori* podem seguir o mesmo padrão apresentado anteriormente (por exemplo, *Normal*(0, 1) para dados padronizados). A previsão do modelo é então dada por:

$$\hat{y}_i \sim \left(1 - \text{Bernoulli}\left(\frac{1}{1 + e^{X_i \vartheta}}\right)\right) \times \text{NegBinom}(e^{X_i \beta}, \phi) \tag{3.16}$$

O primeiro fator da função de previsão corresponde ao resultado da Regressão Logística, enquanto o segundo fator corresponde ao resultado da Regressão Binomial Negativa. Assim, caso seja amostrado o valor 1 da distribuição de Bernoulli, isso representa que o modelo previu que ocorrerá a situação definida pelo *processo estrutural*, de modo que a previsão da quantidade esperada deverá ser 0, o que é obtido por meio da subtra-

ção $1 - 1 = 0$. A multiplicação do primeiro pelo segundo fator resultará também na anulação do segundo fator, de modo que, a despeito da quantidade esperada prevista para o *processo de contagem*, o valor final previsto será $\hat{y}_i = 0$. Complementarmente, caso a Regressão Logística preveja que não ocorrerá o *zero estrutural*, de modo que será amostrado o valor 0 da Distribuição Binomial, a subtração do primeiro fator resultará em 1, pois $1 - 0 = 1$. Dessa forma, a multiplicação do primeiro pelo segundo fator resultará na manutenção do valor previsto pela Regressão Binomial Negativa, que corresponde à quantidade esperada de clientes.

O retorno ao exemplo das lojas de sorvete permite esclarecer esse ponto. Imagine-se que é um belo dia de sol e é feriado, de modo que, como vimos nas seções 3.2.7.2 e 3.2.7.3, a quantidade esperada de clientes é bastante alta. No entanto, a loja não abriu pois estava em reformas. Assim, não importa qual o potencial de clientes que a loja teria naquele dia, ela *necessariamente* terá zero clientes caso não seja aberta ao público. Dessa forma, se o modelo prever que a loja não será aberta – ocorrência cuja chance é aumentada pelo fato de ela estar em reformas – então ele deverá prever também que ela não receberá nenhum cliente nesse dia. Caso o modelo preveja que a loja será aberta, então naturalmente o fato de ser um dia de sol e feriado resultarão em uma quantidade esperada de clientes maior.

Por fim, é possível também reespecificar o Modelo ZIP para que a Regressão Logística não mais preveja a ocorrência de um fenômeno que *impede* a efetivação da quantidade esperada, mas sim de um fenômeno que *permite*. Voltando ao exemplo das lojas de sorvete, é possível modelar igualmente tanto o fato de que a loja *não abrirá* – neste caso valor positivo da variável latente, $S = 1$, representa a não-abertura da loja – quanto o fato de que a loja *abrirá*, caso no qual $S = 1$ representará a abertura da loja. Isso requer no entanto uma reespecificação do modelo, invertendo os valores para a Regressão Logística, na seguinte forma:

$$\begin{aligned}
& \text{, se } y_i = 0, \\
& P(y_i = 0 \mid X_i) \sim \text{Bernoulli}(0 \mid \pi_i) + \text{Bernoulli}(1 \mid \pi_i) \times \text{NegBinom}(\mu_i, \phi) \\
& \text{, se } y_i = 1, \\
& P(y_i = y_i \mid y_i > 0, X_i) \sim \text{Bernoulli}(1 \mid \pi_i) \times \text{NegBinom}(\mu_i, \phi)
\end{aligned} \tag{3.17}$$

Logicamente, é necessário reespecificar também a função de previsão, que não mais necessitará da subtração no primeiro fator:

$$\hat{y}_i \sim \text{Bernoulli}\left(\frac{1}{1 + e^{X_i \theta}}\right) \times \text{NegBinom}(e^{X_i \theta}, \phi) \tag{3.18}$$

A intuição por trás dessas alterações é bastante simples. Uma vez que o *processo estrutural* agora representa a probabilidade de ocorrência de um fator que *possibilita* e não mais *impede* a contagem, quando $S = 1$ a contagem será portanto efetivada e não mais anulada. Assim, caso a Regressão Logística preveja que o processo estrutural ocorrerá, de modo que $\hat{S}_i \sim \text{Bernoulli}(\pi_i \mid \pi_i > 0, 5) = 1$, então o valor previsto pela Regressão Binomial Negativa, $\hat{y}_i \sim \text{NegBinom}(\mu_i, \phi)$ deverá ser mantido, seja ele igual ou maior que zero. A remoção da subtração na função preditora resulta na solução desta questão, pois quando a variável latente prevista para o processo estrutural $\hat{S}_i = 1$, a multiplicação de 1 pela quantidade prevista para o *processo de contagem* resultará na manutenção desse valor como valor previsto pelo modelo ZIP. Complementarmente, caso $\hat{S}_i = 0$, a multiplicação de zero pela quantidade prevista para o *processo de contagem* resultará em sua anulação.

Voltando ao exemplo das lojas de sorvete, uma vez que agora a Regressão Logística modela o fato de a loja ser efetivamente aberta ao público, caso ela seja aberta, $S_i = 1$, devemos levar em consideração a quantidade esperada de clientes naquele dia, dadas as características do dia. No entanto, caso a loja não seja aberta, $S_i = 0$, não importa quais as características do dia, a quantidade esperada de clientes deverá ser zero.

O Modelo de Zeros Inflados é bastante adequado ao objeto de estudo dessa dissertação. Pois, como apresentado na seção 3.2.4, e detalhado na seção D.4 do Anexo Único, há diversos fatores que determinam restrições à

edificação na cidade de São Paulo. Quando presentes em uma determinada quadra da cidade, esses fatores reduzem drasticamente a probabilidade de que essa quadra receba empreendimentos imobiliários verticais, independentemente do potencial construtivo que essa quadra possua. Assim, uma quadra que está situada em uma área de proteção ambiental bastante restritiva terá uma contagem esperada de unidades de tipo vertical produzidas equivalente a zero, mesmo que ela possua grandes áreas de terreno livres para edificação e esteja localizada em uma região com elevado valor de metro quadrado.

3.2.7.5 Modelo Multinível

Os modelos apresentados nas seções 3.2.7.2, 3.2.7.3 e 3.2.7.4 possuem no entanto uma limitação em comum: eles pressupõem que os efeitos¹⁴ das variáveis de controle e tratamento sobre a quantidade esperada de unidades de tipo vertical no período 2013-2024 em São Paulo são constantes para todas as quadras da cidade, independentemente da região da cidade em que elas se situam. Como apresentado na seção 3.2.5, com referência à obra de Villaça (2001), esse pressuposto no entanto não é justificável. Pois é de se esperar que a produção imobiliária na área nobre, região onde se concentram as residências da elite da cidade e que é historicamente priorizada pelo capital imobiliário, possua uma dinâmica distinta das áreas periféricas da cidade, regiões caracterizadas pela vulnerabilidade social e baixa acessibilidade ao mercado de trabalho. E que ambas essas dinâmicas não sejam idênticas à dinâmica das áreas de residência da classe média.

Considere-se por exemplo o valor do metro quadrado de terreno. É possível hipotetizar, sem maiores objeções, que em áreas periféricas terrenos com valores mais elevados resultem em um desincentivo à realização de empreendimentos imobiliários, pois eles implicam em um aumento dos custos fixos do empreendimento que terão que ser diluídos nos valores das unidades a serem vendidas, aumentando-os. Como as pessoas que residem nessas regiões da cidade tendem a possuir baixa renda e, conseqüentemente, acesso limitado a financiamentos, haverá maior dificuldade em vender as unidades produzidas. Com efeito, empreendimentos de habitação social subsidiados

¹⁴Tanto β quanto ϑ , para os modelos ZIP

por programas governamentais, que correspondem à maior parte da produção imobiliária de tipo vertical nessas regiões, tendem a ser construídos em grandes terrenos com baixo valor de metro quadrado.

Já nas áreas de residência da classe média, é possível hipotetizar que o valor do terreno indique regiões mais valorizadas, para as quais há maior demanda. Como as pessoas que residem nessas regiões possuem maior renda e, principalmente, maior acesso a financiamentos, há maior interesse em edificar nesses terrenos devido à possibilidade de lucro com o solo criado. Pois o valor total de venda da área construída das unidades, se vendido ao mesmo valor do metro quadrado de terreno original, pode ultrapassar em muito o valor do terreno adquirido mais os custos de edificação. O que não ocorreria nas áreas periféricas.

Nesse cenário, caso seja ajustada uma regressão que tenha como variável dependente a quantidade de metros quadrados de área construída produzida e como variável explicativa o valor médio do metro quadrado do terreno na quadra, é possível que esse exemplo incorra em uma situação que se convencionou denominar Paradoxo de Simpson. Em sua forma mais simples, como descrito por Blyth (1972), o Paradoxo de Simpson ocorre pois:

$$\begin{aligned}
 & \text{É possível que seja verdadeiro que:} \\
 & P(A | B) < P(A | B') \\
 & \text{E simultaneamente:} \\
 & P(A | BC) \geq P(A | B'C) \\
 & P(A | BC') \geq P(A | B'C')
 \end{aligned} \tag{3.19}$$

O que é contra-intuitivo, pois as pessoas tendem a conceber $P(A | B)$ como representando o valor médio de $P(A | BC)$ e $P(A | BC')$. Assim como tendem a conceber $P(A | B')$ como o valor médio de $P(A | B'C)$ e $P(A | B'C')$. Para ilustrar esse aparente paradoxo, considere-se novamente o exemplo das lojas de sorvete. A base de dados no Quadro 3.5 tabula a preferência por sorvete de morango para os clientes de duas unidades distintas da cadeia de lojas:

Quadro 3.5 – *Exemplo: paradoxo de Simpson*

id_loja	qtd_cli	qtd_mor	mulher	perc_mor
1	80	50	0	62,5%
1	20	10	1	50%
2	20	18	0	90%
2	80	60	1	75%

Fonte: Elaboração Própria

A variável *id_loja* identifica a unidade, *qtd_cli* representa o total de clientes recebidos pela loja no dia e *qtd_mor* identifica o total de clientes que escolheram sorvete de morango, dentre os clientes atendidos. Variável *mulher* identifica tratar-se de clientes do gênero feminino. Já a variável *perc_mor*, por sua vez, é um dado calculado que representa o percentual de clientes de mesmo gênero, para uma mesma unidade, que escolheram sorvete de morango, conforme a equação $perc_mor_i = ((\frac{qtd_mor_i}{qtd_cli_i}) \times 100 \mid mulher_i = k, id_loja_i = j)$.

Como se pode notar, os dados apresentam evidências de que clientes do gênero masculino têm uma maior preferência por sorvetes de morango. Pois em ambas as lojas, eles preferiram mais os sorvetes desse sabor do que as mulheres. Na unidade 1, 62,5% dos homens escolheram esse sabor e as mulheres optaram por ele apenas 50% das vezes. Na loja 2, onde há uma maior preferência geral por morango, os homens escolheram o sabor 90% das vezes, enquanto 75% das mulheres preferiram o morango. No entanto, caso os dados sejam agregados, não mais diferenciando a unidade, ocorre uma aparente transformação. Veja o Quadro 3.6 abaixo.

Quadro 3.6 – *Exemplo: paradoxo de Simpson*

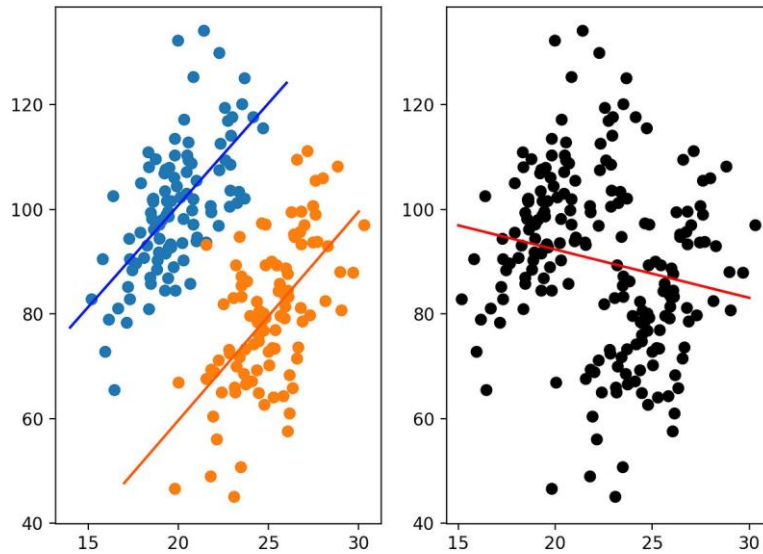
total_cli	total_mor	mulher	perc_mor
100	68	0	68%
100	70	1	70%

Fonte: Elaboração Própria

Os dados agregados mostram uma tendência diferente da identificada anteriormente, indicando que as mulheres parecem preferir mais o morango

do que os homens – 70% das vezes para elas, contra 68% das vezes para eles. Este é o (aparente) Paradoxo de Simpson. Ele pode ocorrer também em modelos de regressão, como mostra a Figura 3.11:

Figura 3.11 – *Paradoxo de Simpson.*



Fonte: Elaboração própria

O painel à esquerda da Figura 3.11 representa uma Regressão Linear Múltipla que contém duas variáveis explicativas. A primeira é contínua, e será representada aqui por X_1 . Ela está plotada no eixo horizontal do gráfico de dispersão. A segunda delas, identificada por X_2 , possui caráter binário e é representada pela cor dos pontos, azuis para $X_2 = 1$ e laranjas para $X_2 = 0$. As retas indicam o melhor ajuste formado pela regressão, conforme a equação $Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2}$, na qual Y corresponde à variável do eixo vertical do gráfico. Note-se que o coeficiente angular para β_1 é o mesmo em ambas as retas, que são paralelas. Mas, para os casos em que $X_{i2} = 1$, a reta ajustada em azul cruza o eixo vertical em um valor acima que a reta da cor laranja. Isso indica que X_2 possui um efeito linear positivo, de modo que $\beta_2 > 0$. A variável explicativa contínua X_1 também possui efeito positivo, que pode ser notado pela orientação da angulação das retas, de modo que $\beta_1 > 0$.

O painel à direita da Figura 3.11 representa um gráfico de dispersão com os mesmos dados. No entanto, a variável X_2 não está mais presente, não havendo portanto mais diferenciação entre os pontos azuis e laranjas. Note-se que, neste caso, a reta em vermelho representada pela equação $Y_i = \alpha + \beta_1 X_{i1}$ cruza o eixo vertical em um valor superior às retas azul e laranja do painel à esquerda. Mas a principal diferença reside no coeficiente de angulação desta reta, que mudou de sentido. Surpreendentemente, X_1 agora possui efeito negativo, de modo que $\beta_1 < 0$.

Em termos mais gerais, o Paradoxo de Simpson ocorre quando a inclusão de uma variável de controle não considerada inicialmente – no exemplo do sorvete de morango, a identificação da loja, e na Figura 3.11, a variável X_2 – altera o sentido de um ou mais coeficientes do modelo. Em geral, essa variável identifica um agrupamento nos dados. A inclusão de uma *dummie* para identificar o grupo permite um ajuste no intercepto, possibilitando uma taxa basal distinta para cada grupo. No entanto, essa medida ainda assim obriga que o coeficiente β para as demais variáveis seja idêntico entre os grupos. Isso pode ser verificado na Figura 3.11, na qual as linhas de regressão azul e laranja possuem interceptos distintos mas um mesmo coeficiente de angulação, sendo portanto paralelas.

Os Modelos Multinível – também conhecidos como Modelos Hierárquicos – dão um passo além, permitindo que não apenas os interceptos variem entre os grupos, mas também os coeficientes de angulação. Esses modelos possibilitam identificar assim diferentes tendências internas aos grupos. Como argumenta Courgeau (2003) em sua introdução geral à metodologia e epistemologia da análise multinível, os modelos multinível possibilitam abandonar o pressuposto, presente nos modelos até então considerados nesta dissertação, de que os registros são independentes e igualmente distribuídos (IID). Em termos mais gerais, segundo o autor, esse pressuposto, operacionalizado em uma estrutura de equação única, resulta no fato de que os modelos GLM são insensíveis às maneiras pelas quais o pertencimento ao grupo pode afetar o comportamento dos indivíduos.

Um contexto clássico em que isso se torna um problema são os estudos voltados à qualidade escolar que utilizam a performance dos alunos como variável dependente. Sabe-se que há uma série de variáveis relacionadas ao aluno que devem ser empregadas como controle, pois estão associadas

à performance, como a habilidade individual, a renda *per capita* familiar e a escolaridade dos pais. No entanto, quem já esteve em uma sala de aula sabe que a relação de ensino-aprendizagem ocorre em um grupo – a turma – e que diferentes características da turma enquanto grupo podem afetar a aprendizagem¹⁵. As turmas, por sua vez, estão situadas em uma escola e, do mesmo modo, as características da escola também possuem efeitos sobre a aprendizagem dos alunos (COURGEAU, 2003, cap. 2). Para Courgeau, ignorar esta relação significa elaborar análises incorretas sobre o comportamento dos indivíduos e, igualmente, sobre o comportamento dos grupos: somente o reconhecimento destas influências recíprocas permite a análise correta dos fenômenos.

Conforme a notação apresentada por Goldstein (1999), os modelos multinível são definidos por um sistema de equações relacionadas da seguinte forma¹⁶:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \epsilon_{ij} \quad (3.20)$$

Onde:

- Y_{ij} representa o valor da variável dependente Y para um indivíduo i pertencente ao grupo j ;
- X_{ij} representa os valores da matriz de variáveis explicativas para um indivíduo i pertencente ao grupo j ;
- β_{0j} representa uma equação específica para cada grupo j que modela o intercepto do grupo (cada grupo possui um intercepto próprio);
- β_{1j} representa uma equação específica para cada grupo j que modela o efeito da matriz de variáveis dependentes X sobre Y_j no contexto de cada grupo;
- ϵ_{ij} representa o termo de erro (resíduo) que captura a variância não explicada pelo modelo.

¹⁵Por isso os professores costumam dizer que há “turmas boas” e “turmas ruins”.

¹⁶Com o objetivo de simplificar a explanação, estamos pressupondo modelos lineares, nos quais a função de ligação é a identidade

O termo β_{0j} , que possibilita a variação no intercepto para cada grupo, é definido pela equação:

$$\beta_{0j} = \gamma_0 + u_{0j} \quad (3.21)$$

Onde:

- γ_{00} representa o intercepto global;
- u_{0j} é o efeito aleatório que captura a variação do intercepto interna a cada grupo;

O termo β_{1j} , que permite que os coeficientes angulares variem para cada grupo, por sua vez, é definido pela equação:

$$\beta_{1j} = \gamma_1 + u_{1j} \quad (3.22)$$

Onde:

- γ_1 representa o valor fixo do coeficiente angular para todos os grupos;
- u_{1j} é o efeito aleatório que captura a variação do coeficiente angular interna a cada grupo;

A substituição das equações que definem os termos β_{0j} e β_{1j} na equação original, resulta na seguinte equação unificada:

$$\begin{aligned} Y_{ij} &= (\gamma_0 + u_{0j}) + (\gamma_1 + u_{1j})X_{ij} + \epsilon_{ij} \\ Y_{ij} &= \gamma_0 + \gamma_1 X_{ij} + u_{0j} + u_{1j} X_{ij} + \epsilon_{ij} \end{aligned} \quad (3.23)$$

Uma especificação *bayesiana* para esse mesmo modelo poderia ocorrer da seguinte forma, na qual ambos os coeficientes angulares e os interceptos – construídos como vetores de valor 1 – estão incluídos nas matrizes γ e u :

$$\begin{aligned} \gamma_j &\sim N(\mu, \tau^2) \\ \mu &= \gamma + u_j X_{ij} \\ \gamma_{ij} &\sim N(\gamma_j, \epsilon^2) \end{aligned} \quad (3.24)$$

As distribuições *a priori* para os parâmetros γ e u foram omitidas. Nesse modelo, o valor da variável Y de um indivíduo i pertencente ao

grupo j , Y_{ij} , é modelado como advindo de uma distribuição normal que possui como média o parâmetro ϑ_j e desvio padrão ϵ . O parâmetro ϑ_j , por sua vez, provém de outra distribuição normal, indexada para o grupo j , com média μ e desvio-padrão τ . Note-se que, nessa especificação, o valor médio previsto para um indivíduo i provém não apenas de suas características individuais, mas de seu pertencimento ao grupo j . Isto pois ϑ_j é ajustado para cada grupo j . Esse parâmetro tem como média o parâmetro μ , que por sua vez é definido pela combinação linear de um coeficiente de angulação comum a todos os grupos, γ , com o desvio específico nesse coeficiente de angulação para aquele grupo u_j em relação aos dados do indivíduo pertencente ao grupo X_{ij} . Nesse cenário, o parâmetro ϵ^2 representa a variação geral do modelo, a nível dos indivíduos, enquanto o parâmetro τ^2 representa a variação entre os grupos. Dessa forma, o modelo é capaz de aprender sobre os indivíduos no mesmo movimento que aprende também sobre cada grupo, um informando o outro de forma recíproca, como defende Courgeau (2003).

Como mostra McElreath (2020, cap. 13), em modelos multinível *bayesianos* é possível definir a distribuição *a priori* dos coeficientes de angulação dos grupos (u_j na equação acima) como uma distribuição Normal Multivariada para cada grupo, de modo que $u_j \sim N_k(\mu, \Sigma)$. Nessa especificação, k representa a quantidade de colunas na matriz de dados X do modelo (ou seja, o total de variáveis explicativas), μ é um vetor de médias e Σ é uma matriz de covariância. Por exemplo, considerando apenas duas variáveis explicativas, de modo que u_j é um vetor composto por duas variáveis aleatórias β_0 e β_1 amostradas de uma distribuição normal bivariada:

$$(u_j | j \in \{1, 2\}) \sim N_2 \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma^2_1 & \sigma_{12} \\ \sigma_{12} & \sigma^2_2 \end{pmatrix} \quad (3.25)$$

A distribuição *a priori* da matriz de covariância Σ – neste caso, também chamada *hyper-priori* – é especificada pela distribuição LLKJ (LEWANDOWSKI; KUROWICKA; JOE, 2009). Segundo McElreath (2020, cap. 13) a especificação da distribuição *a priori* dos coeficientes de angulação de grupo como uma normal multivariada possui duas principais vantagens, decorrentes do fato de que ela permite que esses coeficientes covariem:

- Possibilita que o modelo aprenda melhor com os dados, pois o ajuste de um coeficiente informa os outros coeficientes para aquele grupo (técnica denominada *partial pooling*);
- É adequada para situações nas quais há razões para acreditar que os efeitos das variáveis explicativas estão associados de formas específicas dentro de cada grupo;

Esse último ponto é bastante adequado ao problema de pesquisa dessa dissertação. Considere-se novamente o exemplo do efeito do valor de metro quadrado de terreno nas áreas periféricas e de classe média da cidade. Vimos que nas áreas periféricas há razões para acreditar que essa variável possui efeito negativo em relação à quantidade esperada de unidades verticais produzida por quadra no período 2013-2024. O mesmo pode ser dito da proporção de área edificada de tipo horizontal. Pois as favelas possuem elevada densidade construtiva, de modo que nas áreas periféricas valores altos nessa variável para uma quadra indicam que ela está situada em um loteamento irregular. Por conta dos problemas fundiários e das questões sociais envolvidas, isso resulta em uma redução drástica da probabilidade de que a quadra receba um empreendimento imobiliário vertical, diminuindo assim a quantidade esperada de unidades produzida. Os efeitos do valor de metro quadrado de área construída e da proporção de área edificada de tipo horizontal podem estar portanto positivamente correlacionados nas áreas periféricas.

Já nas áreas de classe média, um valor elevado para a proporção de área edificada de tipo horizontal pode indicar bairros com casas geminadas, com alta densidade construtiva e populacional. Esses imóveis em geral possuem terrenos reduzidos, de modo que é necessário adquirir uma quantidade elevada de lotes, demolir as casas neles situadas e realizar o remembramento dos lotes para viabilizar o empreendimento (vide o processo descrito na seção [D.1](#) do Anexo Único). O custo de aquisição desses imóveis levará em conta não apenas o valor da área de terreno, mas também o valor agregado da área construída. A isso se somam os custos de demolição e incorporação, o que pode resultar em uma elevação significativa dos custos fixos do empreendimento, reduzindo os lucros. Além disso, há a complexidade negocial e burocrática envolvida na incorporação de grande quantidade de imóveis,

assim como o risco de inviabilizar o projeto caso um proprietário se recuse a vender. Dessa forma, nas áreas de classe média pode-se hipotetizar que a proporção da área edificada por construções horizontais está negativamente associada à quantidade esperada de unidades verticais produzidas. O coeficiente dessa variável pode estar, portanto, negativamente correlacionado ao coeficiente do valor do metro quadrado de terreno nessas áreas da cidade.

Os benefícios dos modelos multinível têm um custo, porém. Pois a relação de dependência entre os parâmetros nos diferentes níveis requer que eles sejam estimados em conjunto. A autocorrelação entre os parâmetros dependentes pode resultar na dificuldade do amostrador hamiltoniano em explorar o espaço amostral na medida em que a região de maior densidade de probabilidade conjunta pode ser definida por uma amplitude reduzida de valores possíveis para ambos os parâmetros, assumindo o formato de um funil, como exemplifica a Figura 3.12. Esse cenário resulta em uma maior probabilidade de autocorrelação elevada entre as Cadeias de Markov, convergência lenta e ocorrência de transições divergentes no amostrador.

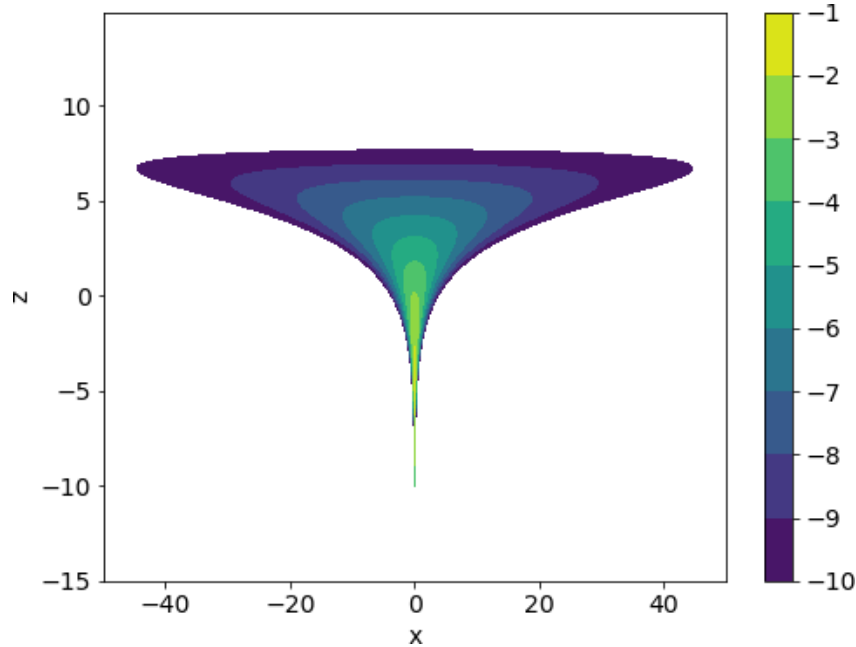
Caso isso ocorra, a parametrização não-centrada é uma solução possível (MCELREATH, 2020). Pois essa técnica permite amostrar os parâmetros de forma independente e, em seguida, recombina-los. Considere por exemplo o seguinte modelo hierárquico, no qual a distribuição de Y_{ij} depende do parâmetro ϑ_j .

$$\begin{aligned} \vartheta_j &\sim N(\mu, \tau^2) \\ Y_{ij} &\sim N(\vartheta_j, \sigma^2) \end{aligned} \quad (3.26)$$

Esse modelo pode ser reparametrizado de forma não-centrada da seguinte maneira:

$$\begin{aligned} \vartheta_j &= \mu + \tau\eta_j \\ \eta_j &\sim N(0, 1) \\ Y_{ij} &\sim N(\vartheta_j, \sigma^2) \end{aligned} \quad (3.27)$$

Essa reparametrização permite que os parâmetros μ e τ (que também possuem suas próprias distribuições *a priori*) sejam amostrados independentemente, como valores globais que representam, respectivamente, a média geral e a variabilidade entre os grupos. O parâmetro η_j , cuja

Figura 3.12 – Densidade de probabilidade conjunta em formato de funil.

Fonte:

https://beanmachine.org/docs/overview/tutorials/Neals_funnel/NealsFunnel/

distribuição *a priori* é uma Normal Padrão, passa então a representar a variabilidade de cada grupo em relação aos parâmetros globais μ e τ . O parâmetro de grupo, ϑ_j , é então reconstruído por meio da translação e escala do parâmetro η_j , obtida pela combinação linear de η_j com μ e τ , conforme a equação $\vartheta_j = \mu + \tau\eta_j$. Note-se que o parâmetro η_j tem como distribuição *a priori* uma Normal Padrão, pois a translação da Normal Padrão por μ e τ resultaria em uma distribuição Normal com parâmetros de média μ e variância τ^2 , $N(\mu, \tau^2)$. Dessa forma, caso a distribuição *a posteriori* de η_j seja similar à sua distribuição *a priori*, isso indica que o coeficiente de grupo é similar ao coeficiente global, não havendo assim nenhuma divergência significativa do grupo em relação à população.

Há ainda um outro custo associado aos modelos multinível, pois a parametrização das distribuições *a priori* dos coeficientes de grupo como distribuições Normais Multivariadas pode resultar em dificuldades computacionais associadas à matriz de covariância. Uma técnica que permite

contornar esse problema é a transformação da matriz de covariância usando a Decomposição de Cholesky. Essa técnica permite representar a matriz de covariância Σ , que é uma matriz simétrica, positiva definida no produto de uma matriz triangular inferior L e sua transposta L^T :

$$\Sigma = LL^T \quad (3.28)$$

Isso permite acelerar o cálculo da inversão da matriz de covariância Σ , que pode ser bem custoso computacionalmente sobretudo se o modelo multinível possuir muitos parâmetros. O que acontece com facilidade, basta que o número de grupos ou de variáveis explicativas seja elevado, pois a quantidade de parâmetros será o produto cartesiano entre ambos. O cálculo da inversão da matriz de covariância pode ser feito na forma mais eficiente com o emprego da Decomposição de Cholesky, na forma $\Sigma^{-1} = (LL^T)^{-1} = L^{-T} L^{-1}$. Além disso, a Decomposição de Cholesky permite realizar a parametrização não-centrada dos coeficientes angulares, de forma similar à apresentada anteriormente. Pois é possível reparametrizar a Normal Multivariada da seguinte forma:

$$\mathbf{N}_k(\mu, \Sigma) = \mu + L\mathbf{N}_k(0, I) \quad (3.29)$$

Nessa reparametrização, a Normal Multivariada passa a ser construída a partir da translação e escala de uma Normal Multivariada Padrão por meio da equação $\mu + L\mathbf{N}_k(0, I)$, onde I é a matriz identidade e L é a Decomposição de Cholesky da matriz Σ . Assim como no caso da parametrização não-centrada apresentada anteriormente, isso permite amostrar separadamente os valores para a média e desvio-padrão do parâmetro e recombina-los em seguida.

As técnicas descritas nessa seção foram utilizadas para permitir a convergência do modelo multinível desenvolvido neste estudo. Esse modelo consolida a discussão apresentada nas seções anteriores, em relação aos modelos para dados de contagem (3.2.7.2) com hiperdispersão (3.2.7.3) e zeros inflados (3.2.7.4). Com efeito, trata-se de um modelo de contagem multinível para zeros inflados com função de verossimilhança binomial negativa que combina, dessa forma, todos os modelos aqui descritos. Com isso, é encerrada a apresentação dos materiais e métodos deste capítulo.

3.3 RESULTADOS

Nesta seção, apresentamos os resultados obtidos a partir dos materiais e métodos apresentados na seção anterior, 3.2. Iniciamos a apresentação com os resultados da análise para identificação das áreas nobre e periféricas da cidade. Em seguida, são apresentados os resultados dos modelos estatísticos desenvolvidos. Todos os modelos têm como variável dependente a quantidade de unidades de tipo vertical produzidas por quadra na cidade de São Paulo no período 2013-2024. Os modelos são apresentados de acordo com a ordem estabelecida pela metodologia do *vorkflow bayesiano*, apresentada na seção 3.2.7: iniciamos com os modelos mais simples e, a partir da análise e comparação de seus resultados, progredimos até os mais complexos.

Esta seção está portanto organizada da seguinte forma:

- Seção 3.3.1 – **Segregação social e agrupamentos espaciais**: apresenta os resultados da identificação das áreas nobre e periféricas da cidade por meio da metodologia de análise LISA descrita nas seções 3.2.5.1 e 3.2.5.2.
- Seção 3.3.2 – **Produção imobiliária e modelos de contagem**: apresenta os resultados da Regressão de Poisson para dados de contagem descrita na seção 3.2.7.2.
- Seção 3.3.3 – **Empreendimentos de grande porte e superdispersão**: apresenta os resultados da Regressão Binomial Negativa para dados de contagem com hiperdispersão descrita na seção 3.2.7.3.
- Seção 3.3.4 – **Restrições construtivas e zeros inflados**: apresenta os resultados do modelo de mistura para dados de contagem com zeros inflados (ZIP) descrito na seção 3.2.7.4.
- Seção 3.3.5 – **Desigualdade e modelos multinível**: apresenta os resultados do modelo multinível descrito na seção 3.2.7.5.

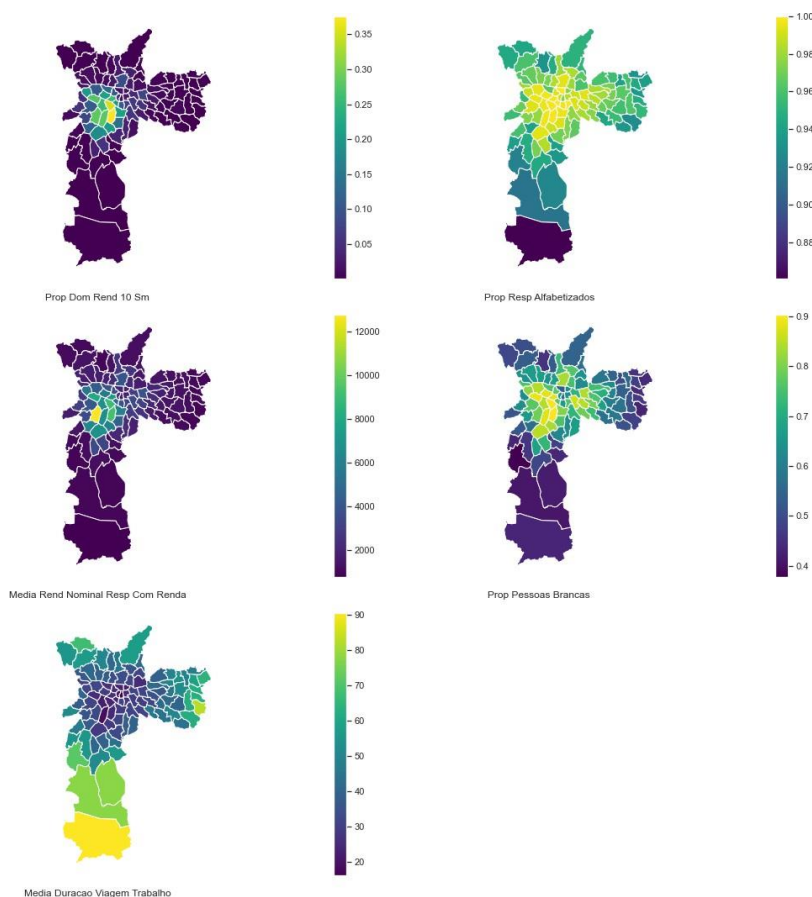
3.3.1 Segregação social e agrupamentos espaciais

Nesta seção, apresentamos os resultados da análise de agrupamentos (*clusters*) que tem por objetivo identificar a estrutura espacial da segregação social na cidade de São Paulo. As variáveis analisadas caracterizam os distritos do município a partir de diferentes dimensões que constituem a segregação social. Essas dimensões são a segregação racial, o acesso à educação básica, a concentração de renda e a acessibilidade ao mercado de trabalho. A listagem abaixo retoma a descrição das variáveis, assim como da dimensão da segregação que cada uma delas tem por objetivo mensurar:

- Proporção de pessoas brancas residentes no distrito (segregação racial);
- A proporção de domicílios com renda superior a 10 salários mínimos no distrito (concentração de renda);
- A média da renda média dos responsáveis por domicílio no distrito (concentração de renda);
- A proporção de responsáveis por domicílio alfabetizados (acesso à educação básica);
- A duração média das viagens a trabalho (acessibilidade ao mercado de trabalho);

A metodologia de cálculo de cada uma dessas variáveis está detalhada na seção [D.6](#) e suas fontes de dados são descritas nas seções [C.1.1](#) e [C.2.2](#) do Anexo Único. A metodologia da análise aqui realizada, por sua vez, está descrita na seção [3.2.5](#) deste capítulo.

A Figura [3.13](#) contém mapas coropléticos que plotam a distribuição espacial dessas variáveis no território da cidade. Todos os mapas utilizam o mesmo esquema de cores, baseado no mapeamento de cores *viridis*, no qual tons mais escuros e próximos do roxo indicam valores reduzidos e tons mais claros, próximos do amarelo, indicam valores elevados. No entanto, cada um dos mapas foi produzido de acordo com a escala da própria variável, o que pode ser verificado na legenda à sua direita.

Figura 3.13 – *Distribuição espacial das dimensões da segregação.*

Fonte: Elaboração própria

Como se pode notar, ambas as variáveis associadas à concentração de renda possuem distribuição espacial bastante similar. Os distritos com maior concentração de renda situam-se no setor sudoeste da cidade. Os distritos de Moema e Jardim Paulista destacam-se no que diz respeito à proporção de domicílios com renda superior a 10 salários mínimos (chegando a mais de 30%). Os distritos do Itaim Bibi, Morumbi e Pinheiros, por sua vez, destacam-se quanto à média da renda nominal média do responsável pelo domicílio, que chega a mais de 12 mil reais – valor que, se corrigido pelo IPCA considerando como data base janeiro de 2010 e como data de destino setembro de 2024 equivale a mais de 27 mil reais.

A proporção de responsáveis por domicílio alfabetizados possui distri-

buição mais igualitária pela cidade, havendo no entanto uma maior concentração de proporções mais elevadas nas regiões mais centrais e menores taxas de alfabetização nas áreas mais periféricas, com destaque para Parelheiros, no Extremo Sul da cidade, que é uma área rural.

Já a proporção de pessoas brancas possui concentração nos distritos do setor sudoeste, além dos distritos do Tatuapé, Mooca e Água Branca na Zona Leste e Santana e Tucuruvi na Zona Norte. A comparação com os mapas que plotam a concentração de renda mostra que esses distritos estão entre os distritos com renda mais elevada em suas respectivas regiões.

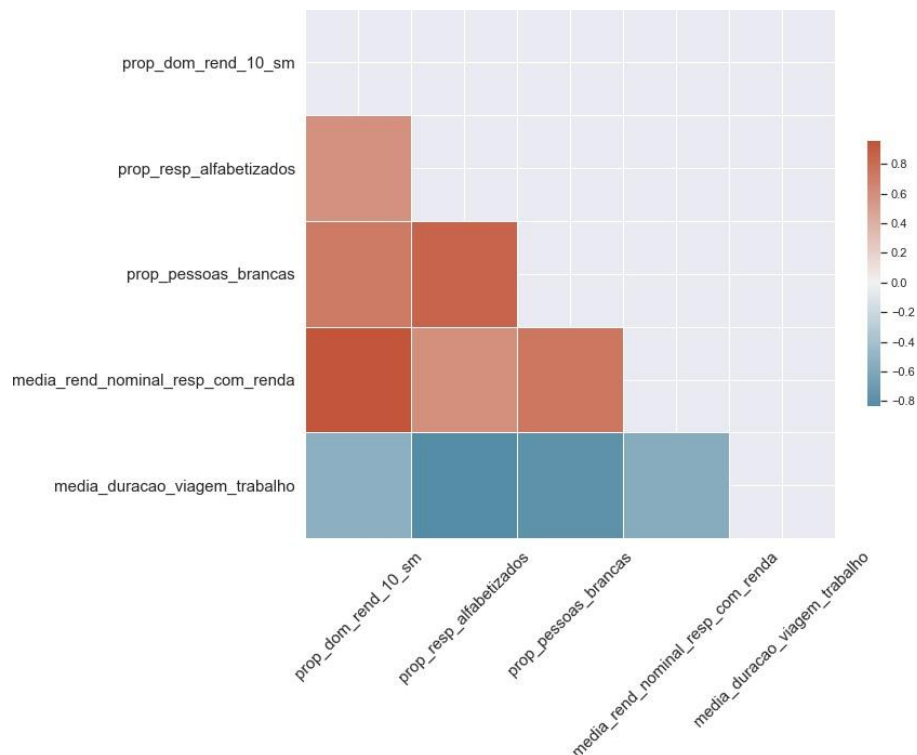
Por fim, a duração média das viagens a trabalho possui uma clara distribuição centro-periferia. Assim as pessoas residentes nos distritos mais periféricos, distantes do centro da cidade tendem a realizar viagens pendulares significativamente mais longas. Note-se que a amplitude dos dados varia de viagens com duração média de menos de 20 minutos nos distritos ricos situados na região central da cidade, como Moema, até uma hora e meia, como é o caso de Parelheiros e Cidade Tiradentes, nos extremos Sul e Leste respectivamente.

De modo geral, esses mapas indicam haver correlação entre as variáveis aqui consideradas. A Figura 3.14 permite visualizar a matriz de correlação de Pearson entre elas, calculada de acordo com a equação abaixo. Valores de ρ próximos a 1 indicam elevada correlação positiva, valores próximos a -1 indicam elevada correlação negativa e valores próximo a 0 indicam a ausência de correlação.

$$\rho_{X,Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (3.30)$$

Como esperado, ambas as variáveis associadas à concentração de renda possuem alta correlação positiva, com $\rho = 0,95$. A proporção de pessoas brancas, por sua vez, também possui correlação positiva elevada em relação à proporção de domicílios com renda maior que 10 salários mínimos ($\rho = 0,72$) e à média da renda nominal média do responsável pelo domicílio ($\rho = 0,74$).

Já a taxa de alfabetização do responsável por domicílio – que, como mostra o mapa da Figura 3.13 possui distribuição territorial mais igualitária – está menos associada à renda, com $\rho = 0,60$ para a média da renda

Figura 3.14 – Matriz de correlação – segregação social.

Fonte: Elaboração própria

média do responsável pelo domicílio e $\rho = 0,58$ para a proporção de domicílios com renda maior que 10 salários mínimos. Ainda assim são correlações significativas. A correlação entre a taxa de alfabetização e a proporção de pessoas brancas é, no entanto, bastante elevada, com $\rho = 0,86$.

Note-se que a correlação entre todas as variáveis descritas até aqui é consistentemente alta e positiva, indicando que é possível que elas mensurem uma mesma variável latente. A situação se inverte quando se analisa a correlação dessas variáveis com o tempo médio de deslocamento ao trabalho: ainda que todas as correlações mantenham-se significativas, todas elas são *negativas*.

O tempo médio de deslocamento a trabalho possui $\rho = -0,54$ para a proporção de domicílios com renda do responsável maior que 10 salários mínimos, $\rho = -0,57$ para a média da renda média do responsável por domicílio, $\rho = -0,78$ para a proporção de pessoas brancas e $\rho =$

−0,83 para a taxa de responsáveis por domicílios alfabetizados. O fato de que esta alteração de sinal ocorreu para a correlação do tempo médio de deslocamento com todas as outras variáveis – que estão positivamente correlacionadas entre si – reforça os indícios de que há uma variável latente.

Todas as correlações da matriz são significativas e fortes, sendo que a mais próxima de 0 de todas elas é $\rho = -0,54$, o que é um valor relativamente elevado. Em resumo, os dados apresentados mostram que, em São Paulo, os distritos que possuem a maior concentração de pessoas brancas, ricas e alfabetizadas são também aqueles nos quais residem as pessoas com os menores tempos de deslocamento a trabalho.

Com objetivo de identificar a variável latente que subjaz a essas variáveis, foi realizada a Análise de Componentes Principais, cuja metodologia foi descrita na seção 3.2.5.1. O Quadro 3.7 abaixo apresenta os resultados dessa análise.

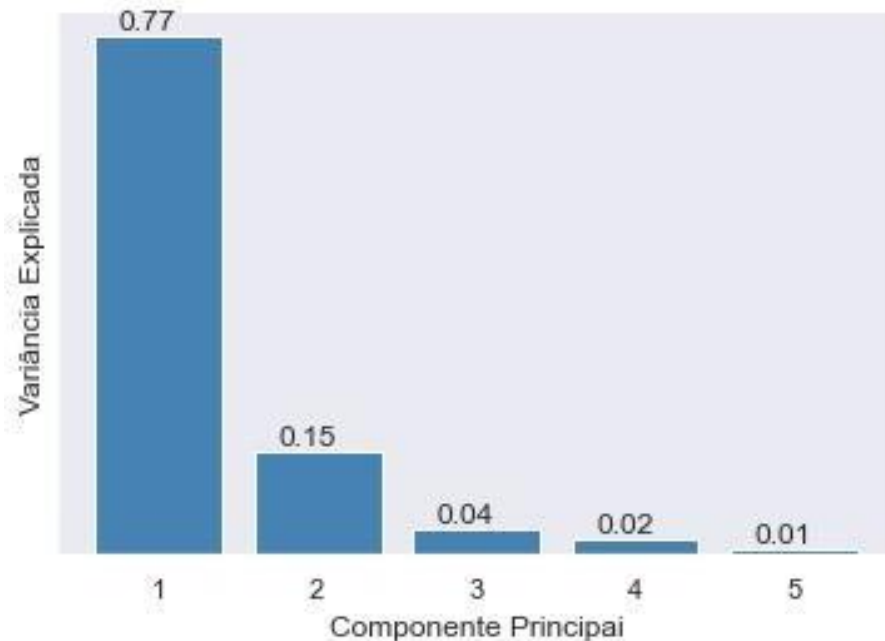
Quadro 3.7 – Análise de Componentes Principais

Componente	Valores singulares	Variância explicada
0	19.283	77.47%
1	8.604	15.42%
2	4.306	3.86%
3	3.322	2.3%
4	2.133	0.95%

Fonte: Elaboração Própria

Como era esperado pela correlação elevada entre as variáveis, o Primeiro Componente, também chamado de Componente Principal, explica um alto percentual da variabilidade dos dados: aproximadamente 80%. A Figura 3.15 apresenta o gráfico de declividade (*scree plot*) da variância explicada pelos componentes. Como se pode notar, há uma queda brusca na declividade a partir do segundo componente, indicando que a partir do primeiro componente a variabilidade explicada pelos demais é residual.

Por fim, o Quadro 3.8 apresenta os pesos (*loadings*) do Componente Principal. Esses valores correspondem à contribuição de cada variável na composição do componente. Note-se que as variáveis contribuem de forma

Figura 3.15 – Gráfico de declividade – PCA

Fonte: Elaboração própria

aproximadamente equivalente com o componente. Além disso, os sinais dos pesos estão de acordo com os resultados apresentados nos parágrafos anteriores. Pois as variáveis socioeconômicas possuem pesos positivos enquanto a variável que mede a duração dos tempos de deslocamento possui um peso negativo. Esse padrão, associado à alta proporção da variabilidade explicada pelo Componente Principal indica que ele possui representatividade em relação ao fenômeno que se está analisando.

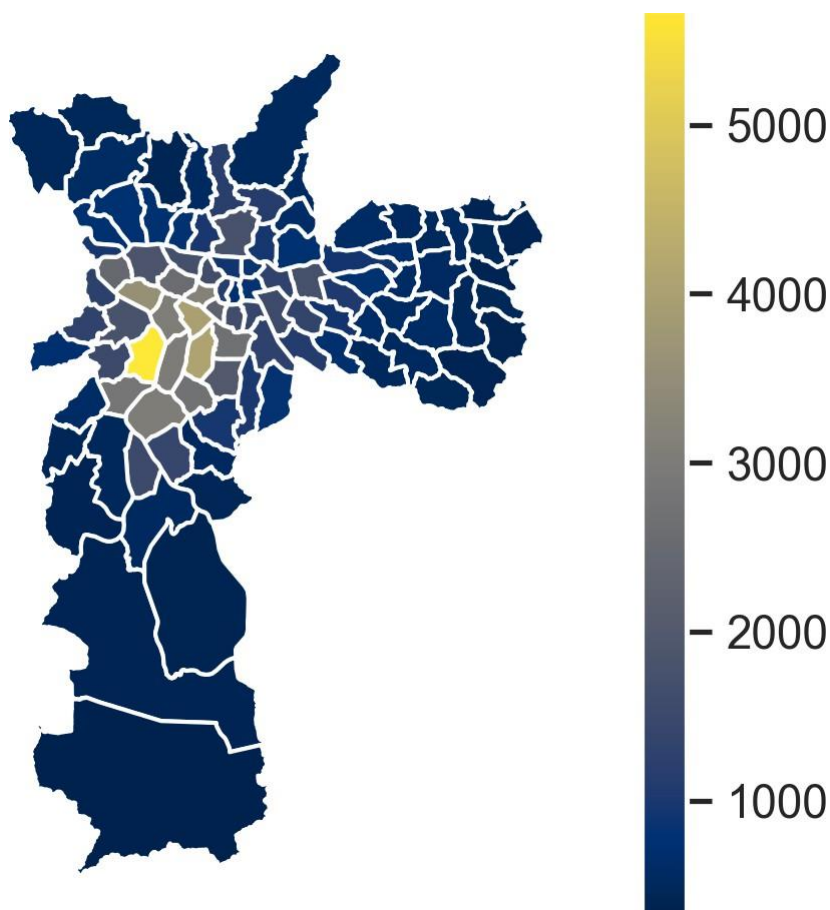
Quadro 3.8 – Pesos do Componente Principal

Variável	Pesos
prop_dom_rend_10_sm	0.44
prop_resp_alfabetizados	0.45
prop_pessoas_brancas	0.48
media_rend_nominal_resp	0.45
media_duracao_viagem	-0.43

Fonte: Elaboração Própria

As variáveis foram então transformadas pela [PCA](#) e foi realizada a redução de dimensionalidade por meio da seleção da primeira dimensão, correspondente à transformação pelo Componente Principal. A variável latente resultante é interpretada neste estudo como retratando o privilégio social, pois os sinais dos pesos apresentados no [Quadro 3.8](#) indicam que, quanto maior o valor dessa variável latente, maior a proporção de pessoas brancas, ricas e alfabetizadas e menores os tempos de deslocamento no distrito. A [Figura 3.16](#) apresenta a distribuição espacial dessa variável latente.

Figura 3.16 – *Distribuição Espacial da Variável Latente*



Fonte: Elaboração própria

O Quadros [3.9](#) e [3.10](#) apresentam cada um os 5 distritos com os maiores e menores valores para a variável latente, respectivamente. Os distritos com

valores mais elevados estão situados no setor sudoeste da cidade e estão na região com os maiores valores por metro quadrado, bastante procurada pelo mercado imobiliário. Eles estão contidos nas listagens dos *bairros nobres* da cidade referenciadas na seção 3.2.5. Já os distritos com os menores valores para a variável latente estão situados nos extremos Leste e Sul da cidade, em áreas distantes do centro com alta vulnerabilidade social. Esses resultados confirmam que a variável latente mensura o privilégio social do distrito.

Quadro 3.9 – *Variável Latente – maiores valores.*

Nome do Distrito	Valor
MORUMBI	5678.747
MOEMA	4102.527
JARDIM PAULISTA	4012.740
ALTO DE PINHEIROS	3570.324
CONSOLACAO	3188.458

Fonte: Elaboração Própria

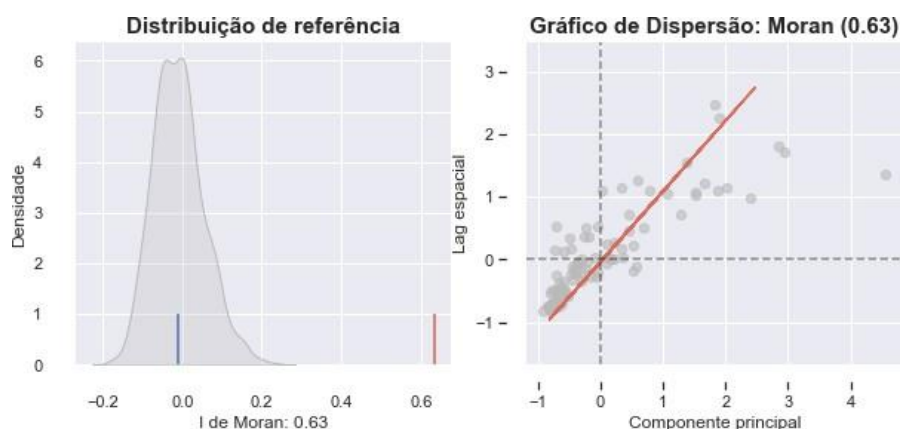
Quadro 3.10 – *Variável Latente – menores valores.*

Nome do Distrito	Valor
PARELHEIROS	408.987
JARDIM ANGELA	400.475
GRAJAU	391.226
CIDADE TIRADENTES	382.397
MARSILAC	310.266

Fonte: Elaboração Própria

Em termos gerais, a distribuição espacial da variável latente apresentada pela Figura 3.16 parece seguir um padrão de autocorrelação espacial positiva, no qual os valores dos distritos próximos entre si são similares. Com o objetivo de testar essa hipótese, foi calculada a estatística I de Moran, obtida conforme metodologia apresentada na seção 3.2.5. A Figura 3.17 apresenta o resultado dessa estatística.

O valor positivo e elevado de 0,62 para a estatística confirma a autocorrelação positiva. Como se pode observar pelo quadro à esquerda da

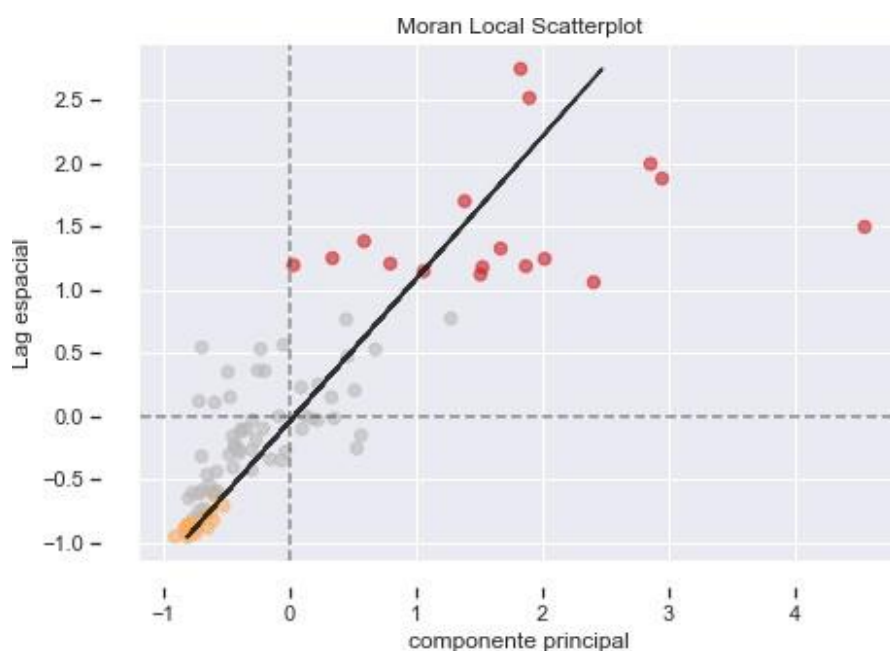
Figura 3.17 – Correlação espacial global – *I* de Moran

Fonte: Elaboração própria

Figura 3.16, esse resultado é significativo, estando bastante distante da média e da área de concentração da densidade de probabilidade da distribuição de referência obtida pela permutação aleatória da variável latente na matriz de vizinhança dos distritos. A linha de tendência linear do Gráfico de Dispersão de Moran, apresentado no quadro à direita da Figura 3.16 e produzido conforme a metodologia descrita na seção 3.2.5, também confirma essa hipótese.

Além disso, a análise da distribuição espacial da variável apresentada pela Figura 3.16 indica haver uma área de concentração da variável latente situada no Setor Sudoeste da cidade, com alguns valores relativamente mais elevados nas regiões de Santana e do Tatuapé. As áreas periféricas da cidade possuem também valores consistentemente reduzidos. Para confirmar essa hipótese de existência de agrupamentos espaciais definidos (*clusterização espacial*), foi realizada uma análise LISA (ANSELIN, 1995), conforme a metodologia descrita na seção 3.2.5.

A Figura 3.18 detalha o Gráfico de Dispersão de Moran para a variável latente. Os pontos destacados em vermelho no quadrante superior direito do gráfico de dispersão indicam distritos pertencentes a *clusters* Alto-Alto, que possuem valores elevados tanto para a variável latente quanto para sua defasagem espacial (*spatial lag*). Em outras palavras, trata-se de distritos com valores elevados circundados de distritos que também possuem valores

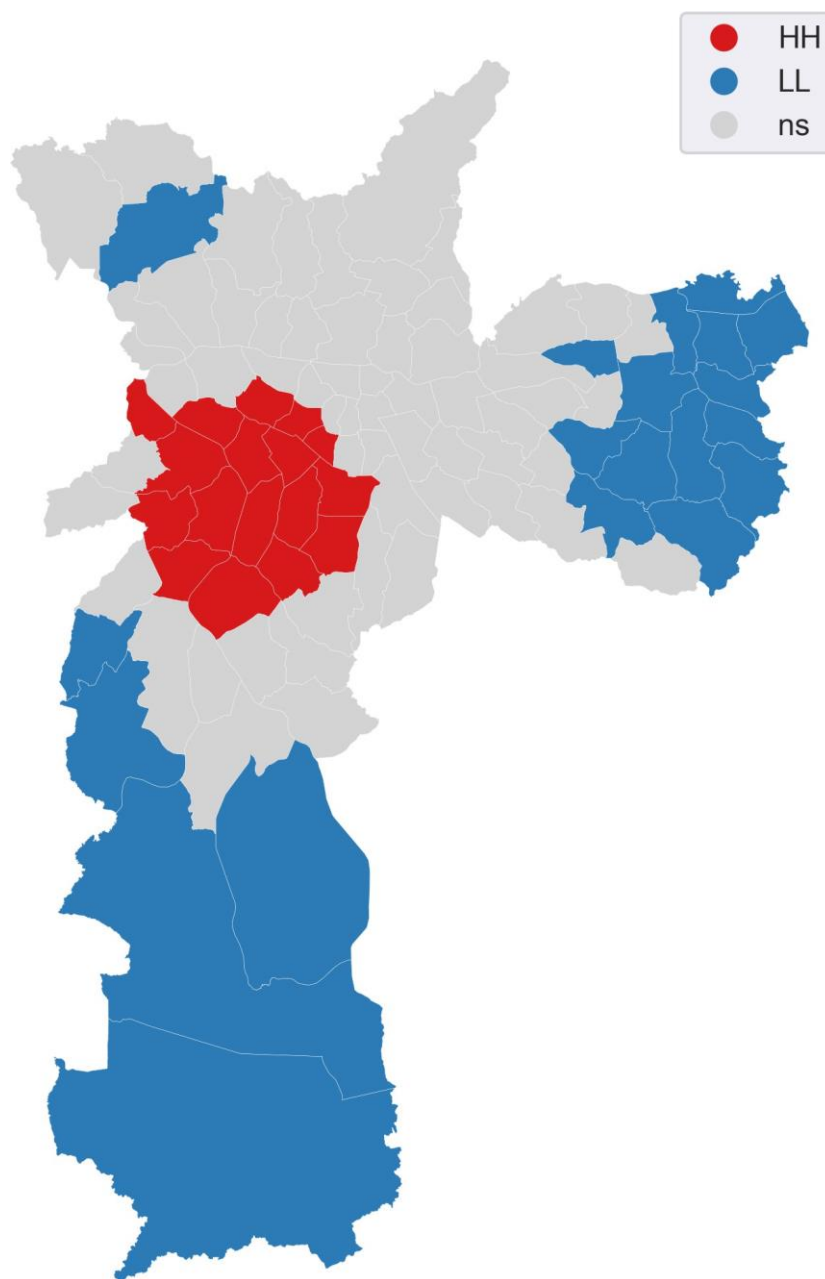
Figura 3.18 – Gráfico de Dispersão de Moran

Fonte: Elaboração própria

elevados.

Já os pontos em amarelo no quadrante inferior esquerdo indicam distritos pertencentes a *clusters* Baixo-Baixo, que possuem valores reduzidos para a variável latente e para a defasagem espacial. Ou seja, são distritos com valores reduzidos circundados de distritos que também possuem valores reduzidos (ANSELIN, 1996). Note-se que estão destacados apenas os distritos com autocorrelação espacial local significativa, isto é, cujos valores para a estatística Moran Local possuem menos de 5% de probabilidade de terem sido gerados pela permutação aleatória dos valores dos registros situados em sua vizinhança (ANSELIN, 1995).

A Figura 3.19 permite visualizar a localização espacial dos *clusters* Alto-Alto e Baixo-Baixo encontrados pela análise LISA. Os distritos em azul correspondem aos pontos em amarelo destacados no Gráfico de Dispersão de Moran da Figura 3.18 – trata-se dos *clusters* Baixo-Baixo. Esses *clusters* estão consistentemente situados na periferia da cidade, nos extremos Leste, Sul e Noroeste. Note-se que os distritos ausentes do *cluster*

Figura 3.19 – *Agrupamentos espaciais*

Fonte: Elaboração própria

Baixo-Baixo na região Noroeste possuem grande proporção de suas áreas

em preservação ambiental¹⁷, além de conterem ainda áreas de ocupação de caráter rural, o que explica a aparente quebra de padrão. Já os distritos em vermelho correspondem aos pontos destacados na mesma cor no Gráfico de Dispersão de Moran, que identificam o pertencimento a *clusters* Alto-Alto. Como se pode notar, eles formam um único *cluster* Alto-Alto com fronteiras contínuas situado em uma região central no quadrante Sudoeste da cidade.

Os resultados da análise demonstram assim que os distritos da cidade com a maior concentração de pessoas brancas, ricas, alfabetizadas e com maior acessibilidade ao mercado de trabalho estão concentrados no quadrante Sudoeste da cidade. Há, além disso, agrupamentos espaciais definidos nas áreas periféricas da cidade nos quais essas variáveis possuem valores reduzidos. Esses resultados confirmam em larga medida a análise de Villaça (2001) sobre a estrutura de segregação social no espaço intraurbano da cidade de São Paulo.

O pertencimento dos distritos a cada um desses tipos de *clusters* foi caracterizado de acordo com sua posição relativa na estrutura de segregação social aqui identificada. Os distritos situados no *cluster* Alto-Alto no quadrante Sudoeste da cidade foram identificados como pertencentes à *área nobre* da cidade, termo que foi empregado para referenciar o interesse histórico do mercado imobiliário nessa região, que costuma denominá-la dessa maneira, como apresentado na seção 3.2.5. Esse termo foi escolhido em preferência ao termo *centro* pois a localização do *cluster* não corresponde ao centro histórico ou mesmo geométrico da cidade.

Já os distritos situados nos *clusters* Baixo-Baixo localizados nos extremos Sul, Leste e Noroeste da cidade foram identificados como pertencentes à *área periférica*, termo escolhido com o objetivo de referir a lógica espacial Centro-Periferia por meio da qual se manifesta a segregação social identificada pela análise aqui desenvolvida. Note-se que apesar da *área nobre* não estar situada exatamente no centro geográfico da cidade, ela ainda assim está em uma posição central. Por fim, os distritos que não pertencem a nenhum desses *clusters* foram identificados como pertencentes às *áreas de classe média*, nas quais, conforme afirma Marques (2014) há maior heterogeneidade social. Essas categorias foram então codificadas como variáveis

¹⁷O Pico do Jaraguá está situado nessa região.

dummies, tendo sido o pertencimento às *áreas de classe média* codificado como a categoria de referência.

Essas variáveis foram então incluídas na base de dados deste estudo – descrita na seção 3.2.6 – por meio da identificação do distrito a que pertencem as quadras da cidade. Elas serão incluídas nos modelos estatísticos que têm por objetivo analisar a capacidade do planejamento urbano em direcionar a produção imobiliária na cidade, cujos resultados são apresentados nas seções que seguem. Como será demonstrado, a localização das quadras na área nobre ou nas áreas periféricas da cidade tem efeitos significativos sobre essa capacidade de direcionamento.

3.3.2 Produção imobiliária e modelos de contagem

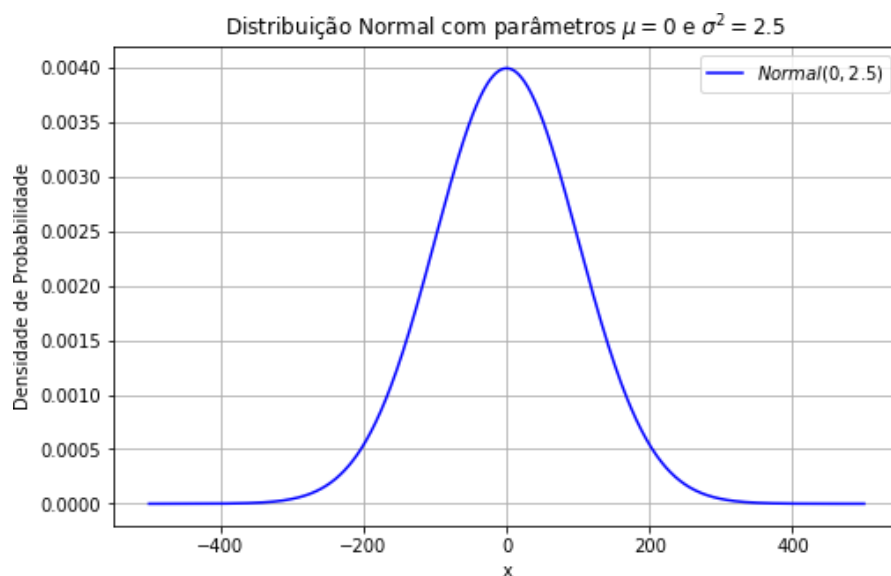
Como discutido na seção 3.2.7.1, a variável dependente deste estudo – a quantidade de unidades de tipo vertical produzidas por quadra na cidade de São Paulo no período 2014-2023 – representa dados de contagem. Dessa forma, ela tem como domínio o conjunto dos números naturais, incluindo o zero, $Y \in \mathbb{N}_0$. Modelos GLM com função de ligação logarítmica são adequados para modelar esse tipo de variável. Com o objetivo de seguir a metodologia do fluxo de trabalho de análise de dados *bayesianos* (*bayesian workflow*) apresentada na seção 3.2.7, a descrição dos resultados dos modelos estatísticos desenvolvidos no presente estudo é iniciada com a Regressão de Poisson, que consiste em uma das especificações mais simples da família de modelos GLM para dados de contagem.

Conforme a metodologia exposta na seção 3.2.7.2, a Regressão de Poisson foi especificada de forma a modelar a quantidade de unidades de tipo vertical produzidas na *i-ésima* quadra da cidade de São Paulo no período 2013-2014, y_i , como distribuída de acordo com uma distribuição de Poisson. O parâmetro único λ dessa distribuição é definido pela exponenciação do preditor linear $X_i\beta$, que contém as variáveis explicativas do modelo, incluindo a variável de tratamento. O modelo segue, portanto, a equação abaixo:

$$\begin{aligned} Y_i &\sim \text{Poisson}(\lambda_i) \\ \lambda_i &= e^{\alpha + X_i\beta} \end{aligned} \quad (3.31)$$

A matriz de dados X do preditor linear foi construída de modo a conter a variável de tratamento, que identifica se a quadra encontra-se uma Zona de Estruturação Urbana (ZEU), assim como as variáveis de controle associadas ao potencial construtivo da quadra apresentadas no Quadro 3.1 da seção 3.2.6. Foi especificada uma mesma distribuição *a priori* pouco informativa para cada um dos coeficientes β : a distribuição Normal com média 0 e desvio-padrão 2,5, $N(0, 2, 5^2)$. A função de densidade de probabilidade dessa distribuição está representada pela Figura 3.20. Como se pode notar, ela possui suporte para valores bastante extremos dos coeficientes, ao mesmo tempo em que concentra a distribuição em torno do zero.

Figura 3.20 – Distribuição *a priori* – $Normal(0, 2.5^2)$



Fonte: Elaboração própria

Previamente ao ajuste do modelo aos dados, foi realizada a checagem da adequação da especificação das distribuições *a priori* por meio do ajuste do modelo a dados sintéticos com parâmetros conhecidos (*prior predictive check*). Para isso, foram gerados dados sintéticos de acordo com a seguinte equação:

$$\lambda_i = e^{\alpha' + \beta' X_i} \quad (3.32)$$

$$y_i \sim \text{Poisson}(\lambda_i)$$

O coeficiente α' , que representa o intercepto, corresponde ao valor 1. O vetor de coeficientes sintéticos β' , por sua vez, corresponde a $\beta' = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$. Já a matriz de dados simulados X' possui 100 linhas e 3 colunas, na seguinte forma:

$$X' = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ \vdots & \vdots & \vdots \\ a_{100,1} & a_{100,2} & a_{100,3} \end{pmatrix} \quad (3.33)$$

Os valores de X' consistem em amostras aleatórias extraídas de uma distribuição Normal Padrão, $X'_{ij} \sim N(0, 1)$. O modelo foi então ajustado para esses dados, parametrizado com 2 Correntes de Markov e 2.000 iterações, das quais mil foram reservadas para a fase de aquecimento (*varmup*) do amostrador **NUTS**. O modelo convergiu sem dificuldades, sendo que a maior distância em relação ao valor 1 para a estatística \hat{R} correspondeu a 1,004.

Essa estatística corresponde à raiz quadrada da proporção entre a média das variâncias internas de cada Cadeia de Markov e variância total estimada pelo modelo, W , que inclui a variância interna às cadeias mas também entre elas. Ela é calculada de acordo com a equação $\hat{R} = \sqrt{\frac{\bar{v}}{W}}$. Um valor próximo de 1 significa que as cadeias não variaram entre si numa proporção superior à sua variação interna, tratando-se assim de um indicador de convergência do modelo. A Tabela 3.1 a seguir apresenta o resultado dos coeficientes estimados:

As colunas HDI 0.5% e HDI 99.5% representam, respectivamente, os valores inicial e final do menor intervalo que contém 99% da densidade de probabilidade da distribuição *a posteriori* estimada para o coeficiente. Em outras palavras, esses valores definem o intervalo de confiabilidade do coeficiente, a 99%. Como se pode notar pela leitura da Tabela 3.1 a especificação do modelo logrou sucesso em estimar os parâmetros dos dados sintéticos. As

Tabela 3.1 – Resultados da simulação - Poisson

Coeficiente	Média	Desvio-padrão	HDI 0.5%	HDI 99.5%
α'	1.005	0.014	0.967	1.039
β'_1	1.014	0.008	0.992	1.033
β'_2	2.016	0.011	1.988	2.044
β'_3	2.995	0.003	2.986	3.004

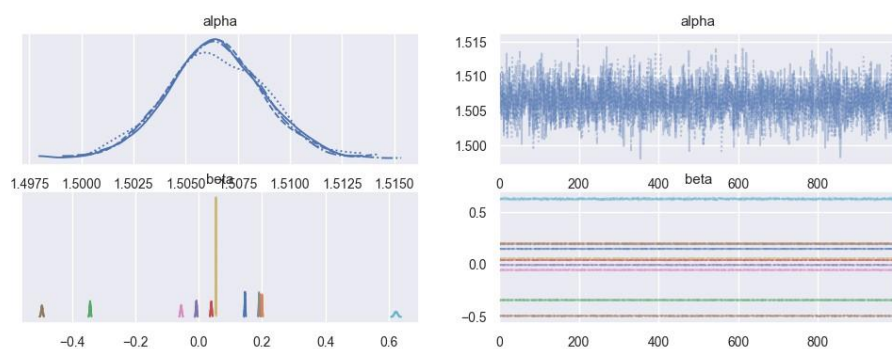
Fonte: Elaboração Própria

médias das distribuições *a posteriori* estão bastante próximas dos valores verdadeiros e os desvios-padrão são pequenos. Além disso, os valores inicial e final dos intervalos de confiabilidade não cruzam o zero, de modo que a estimativa do sinal dos coeficientes é consistente. Dessa forma, os resultados da simulação indicam que o modelo está especificado adequadamente e que é possível seguir com o ajuste para os dados reais.

O modelo foi parametrizado para 2.000 iterações, das quais mil correspondem ao aquecimento do amostrador, e 4 Cadeias de Markov. A Figura 3.21 representa o gráfico de rastreamento (*traceplot*) das Cadeias de Markov do modelo ajustado. Esse gráfico permite visualizar os parâmetros amostrados pelas cadeias ao longo das iterações. Ele permite identificar visualmente se houve convergência, que é representada pelo fato de as Cadeias de Markov terem se estabilizado em torno de um valor ao longo das iterações. Além disso, permite identificar também a mistura entre as cadeias, que pode ser notada pela correlação entre as amostras das cadeias para um mesmo parâmetro. Como se pode notar pela Figura 3.21, tanto a convergência quanto a mistura foram alcançadas com sucesso pelo modelo. Consistentemente, as estatísticas \hat{R} para todos os coeficientes são satisfatórias, sendo que o valor 1,004 é aquele mais se distanciou de 1.

A Tabela 3.2 apresenta os resultados das distribuições *a posteriori* dos coeficientes amostrados pelo modelo. Os valores são apresentados na escala logarítmica, de modo que valores positivos indicam associação positiva entre a variável explicativa e a variável dependente e, complementarmente, valores negativos indicam associação negativa. Todos os coeficientes do modelo são significativos a um intervalo de confiabilidade de 99%.

O coeficiente α , referente ao intercepto, representa a taxa basal da quantidade esperada de unidades de tipo vertical produzidas na quadra no

Figura 3.21 – Traceplot - Poisson

Fonte: Elaboração própria

Tabela 3.2 – Descrição da tabela

Coefficiente	Média	Desvio-padrão	HDI 0.5%	HDI 99.5%
α	1.506	0.002	1.500	1.513
val_m2_terreno	0.146	0.001	0.143	0.149
idade_imoveis	0.199	0.001	0.196	0.203
idade_contribuintes	-0.344	0.002	-0.349	-0.340
media_pavimentos	0.039	0.002	0.035	0.044
prop_maior_terreno	-0.008	0.002	-0.012	-0.004
prop_area_hor	-0.497	0.002	-0.503	-0.491
prop_area_verti	-0.056	0.002	-0.062	-0.051
prop_area_galpoes	0.191	0.001	0.188	0.193
qd_area	0.054	0.000	0.053	0.055
zeu	0.624	0.005	0.611	0.638

Fonte: Elaboração Própria

período 2014-2023. Após exponenciação, esse coeficiente corresponde a 4,5 unidades por quadra. Esse valor, aparentemente baixo, é relevante pois, como será apresentado na seção 3.3.4 há uma proporção significativa de quadras da cidade que não recebeu nenhum empreendimento imobiliário no período.

Dentre as variáveis de controle, o maior tamanho de efeito positivo é dado pela média da idade dos imóvel da quadra em 2014 (idade_imoveis). Após exponenciação, um aumento de um desvio-padrão nessa variável corresponde a um aumento de 22% na quantidade esperada de unidades de

tipo vertical produzidas, as demais variáveis mantidas iguais. Isso indica que há um fator relevante de substituição do estoque de edifícios na cidade, com edificações mais antigas sendo substituídas em uma proporção superior.

A proporção da área total de terrenos com uso do solo referente a terrenos baldios, galpões e usos similares em relação à área da quadra (*prop_area_galpoes*) representa o segundo maior efeito positivo. Após exponenciação, um aumento de um desvio-padrão nessa variável corresponde a um aumento de 21% na quantidade esperada de unidades produzida, tudo o mais mantido igual. Esse efeito positivo é esperado, pois indica que quadras nas quais há imóveis disponíveis com grandes terrenos e baixos custos de demolição são mais propensas a receber empreendimentos imobiliários.

O valor médio do metro quadrado de terreno na quadra (*val_m2_terreno*) está na terceira posição, com um tamanho de efeito equivalente a aumento de aproximadamente 16% na quantidade esperada de unidades produzida, *ceteris paribus*. Mais uma vez, o efeito positivo é esperado pois, como argumenta (BERTAUD, 2022b) o valor da terra deve ser suficientemente alto para que seja viável a realização de empreendimentos verticais. Pois é necessário que o valor de venda da área construída total do empreendimento, que é determinado em grande parte pelo valor do metro quadrado na região, ultrapasse os custos de aquisição de terreno e de edificação para que o empreendimento seja lucrativo.

A média da quantidade de pavimentos das edificações existentes na quadra (*media_pavimentos*) também possui efeito positivo, ainda que mais reduzido, equivalente a um aumento de 4% na quantidade esperada. Esse efeito indica que quadras já possuem edificações são mais propensas a receber empreendimentos imobiliários. Isso é esperado por duas principais razões. Em primeiro lugar, o fato de que a quadra já possui edificações de maior porte indica que ela não incorre nas restrições à construção descritas na seção 3.2.4 e detalhadas na seção D.4 do Anexo Único, o que aumenta a probabilidade de receber um novo empreendimento. Em segundo lugar, há que se considerar que os processos de verticalização em geral estão associados a ciclos de valorização da terra que são localizados na cidade, de forma que o fato da quadra ter recebido empreendimentos no passado pode indicar que ela está situada em uma região que está se valorizando,

o que também resulta em um aumento na probabilidade de receber novos empreendimentos.

A proporção do maior terreno não vertical em relação à área da quadra (*prop_maior_terreno*) tem efeito negativo, mas muito próximo de zero, correspondendo a uma redução de apenas 0,7% na quantidade esperada de unidades produzida. Isso pode ter ocorrido pois a variável *prop_area_galpoes* está positivamente correlacionada, pois áreas de galpão tendem a ocupar grandes terrenos, de modo que o efeito de facilidade de incorporação foi capturado por essa variável.

A média da idade dos proprietários dos imóveis em 2014 (*idade_contribuintes*) possui um efeito negativo relevante, correspondendo a uma redução de 29% na quantidade esperada de unidades produzida na quadra. Esse efeito surpreende, pois esperava-se que houvesse uma tendência maior de verticalização em áreas ocupadas por uma população mais idosa, indicando a renovação do estoque residencial mencionada na análise do efeito da variável *idade_imoveis*, que está associada à transformação do perfil demográfico da população residente ocorrida em processos de verticalização. Será necessário um estudo posterior para compreender melhor esse resultado, mas é possível que ele resulte da correlação entre a idade dos proprietários e a idade das edificações.

A proporção da área construída total de edificações horizontais em relação à área da quadra (*prop_area_hor*) possui forte efeito negativo, correspondendo a uma redução de 39% na quantidade esperada de unidades. Como assinalado na seção 3.2.7.5, o efeito provavelmente está associado ao fato de áreas de favela e regiões periféricas com loteamentos irregulares possuírem alta densidade construtiva. A proporção da área construída total de edificações de tipo vertical em relação à área da quadra (*prop_area_vert*) também possui efeito negativo, mas com menor magnitude, correspondendo a uma redução de aproximadamente 5% na quantidade de unidades esperada. O efeito captura o fato de que quadras saturadas, com muitos edifícios, são menos propensas a receber novos empreendimentos devido aos custos de demolição e dificuldades de incorporação.

Um aumento de um desvio-padrão na área total da quadra (*qd_area*) corresponde a um aumento de 5% na quantidade esperada de unidades produzidas, tudo o mais mantido igual. Esse efeito positivo é esperado, uma

vez que quadras maiores, pelo simples fato de possuírem maior área, têm maior chance de receber empreendimentos. O fato de que o efeito é relativamente pequeno, ademais, indica que os fatores que regem a verticalização da cidade não estão associados apenas à área urbanizada disponível, o que se coaduna aos demais resultados aqui descritos.

Por fim, a variável de tratamento, *zeu*, que identifica se a quadra está situada em Zona de Estruturação Urbana tem efeito positivo. Há que se ressaltar que trata-se do efeito de maior magnitude do modelo. De acordo com o modelo, o fato de a quadra estar situada em uma *ZEU* resulta em um impressionante aumento médio de 86% na quantidade esperada de unidades de tipo vertical produzidas no período de vigência do Plano Diretor, 2014-2023. Como será demonstrado nas próximas seções, esse aparente sucesso do planejamento urbano em direcionar a produção imobiliária no entanto não pode ser tomado como verdadeiro de forma precipitada.

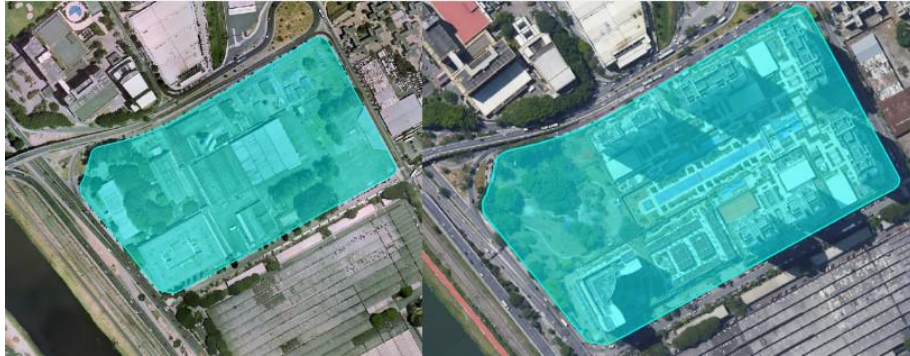
3.3.3 Empreendimentos de grande porte e superdispersão

Como apresentado na seção 3.2.7.3, uma das limitações da Regressão de Poisson é que esse modelo pressupõe que a variável prevista y^* possui média e variância idênticas. Esse pressuposto não se sustenta para domínios nos quais o processo gerativo dos dados pode resultar em valores raros e extremos. A produção imobiliária enquadra-se nessa situação, pois há casos relativamente raros de empreendimentos imobiliários de grande porte que ultrapassam significativamente a média de unidades produzida por outros empreendimentos. Esses empreendimentos em alguns casos chegam até mesmo a ocupar a totalidade da área da quadra, cujo padrão de uso do solo é inteiramente modificado por eles.

A Figura 3.22 ilustra um desses casos. Ela contém fotos aéreas de uma mesma quadra da cidade de São Paulo para os anos de 2004 e 2020 retiradas do Portal GeoSampa. A análise da imagem permite identificar que, no período, foi construído um empreendimento de grande porte que abarcou toda a área da quadra. Foram construídas 6 torres, com pelo menos 2 blocos cada uma. A quadra, que era anteriormente ocupada por galpões, teve seu uso do solo completamente modificado pelo empreendimento. Como afirmado na seção 3.2.7.3, casos extremos como esse, que divergem signi-

ficativamente em relação à média, acarretam em um aumento da distância entre a média e a variância dos dados.

Figura 3.22 – *Empreendimento de grande porte*



Fonte: Portal GeoSampa

Com efeito, a variável dependente deste estudo, que mensura a quantidade de unidades de tipo vertical produzida por cada quadra da cidade de São Paulo no período 2014-2023 possui variância muito superior à sua média. A média é de 6.95 unidades por quadra, enquanto o valor da variância é 3008,55. Há, portanto, uma situação de superdispersão, em relação à qual os pressupostos dos modelos de Poisson são pouco adequados para lidar.

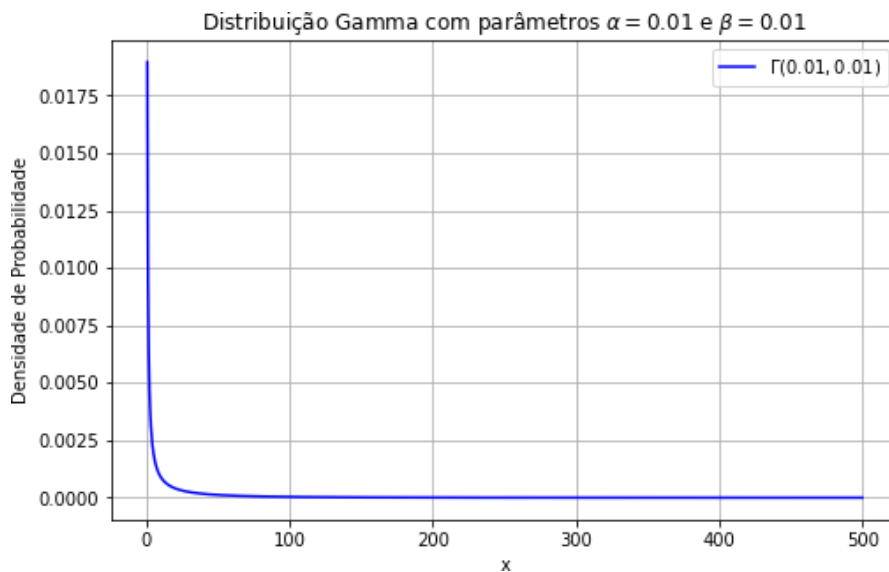
Com o objetivo de contornar essa limitação, foi desenvolvido um modelo para dados de contagem com função de verossimilhança definida pela distribuição Binomial Negativa. Como apresentado na seção 3.2.7.3, a distribuição Binomial Negativa possui um parâmetro de superdispersão que permite que ela se ajuste melhor a casos extremos, aqui representado por ϕ , que tem como domínio os números reais positivos, $\phi \in \mathbb{R}^+$. Quando esse parâmetro possui um valor elevado, $\phi \rightarrow \infty$, a variância e a média da distribuição Binomial Negativa se aproximam, de modo que ela se comporta como a distribuição de Poisson.

O modelo desenvolvido foi especificado da seguinte forma:

$$\begin{aligned} y_{ij} &\sim \text{NegBinom}(\mu_i, \phi) \\ \mu_i &= e^{\alpha + \beta x_i} \end{aligned} \quad (3.34)$$

Para os coeficientes α e β foi especificada a mesma distribuição *a priori* pouco informativa $N(0, 2, 5^2)$ representada pela Figura 3.20. O parâmetro de superdispersão ϕ foi especificado inicialmente como seu inverso, $\frac{1}{\phi}$ e em seguida transformado novamente para ϕ na computação da função de verossimilhança. Essa medida permite tornar a especificação da distribuição *a priori* do parâmetro mais intuitiva, pois quando o parâmetro transformado *phi* inverso se aproxima de zero, $\frac{1}{\phi} \rightarrow 0$, *phi* por sua vez se aproximará do infinito positivo, $\phi \rightarrow \infty$. Assim, para especificar uma distribuição *a priori* que é pouco informativa quanto à superdispersão dos dados, é possível definir uma distribuição de probabilidade com domínio no conjunto dos reais positivos e que tenha maior concentração da densidade de probabilidade próxima a zero (mas que não torne impossíveis valores mais extremos). A distribuição *a priori* escolhida foi a Gamma com os parâmetros de forma $\alpha = 0.01$ e de escala $\frac{1}{\theta} = 0.01$. A Figura 3.23 plota a função de densidade de probabilidade dessa distribuição.

Figura 3.23 – Distribuição *a priori* – $\text{Gamma}(0.01, 0.01)$



Fonte: Elaboração própria

O modelo foi ajustado inicialmente para os mesmos dados sintéticos apresentados na seção 3.3.2. A Tabela 3.3 apresenta os coeficientes resultantes.

Tabela 3.3 – Resultados da simulação - Binomial Negativa

Coeficiente	Média	Desvio-padrão	HDI 0.5%	HDI 99.5%
α'	0.969	0.023	0.900	1.014
β'_1	1.021	0.015	0.981	1.060
β'_2	2.027	0.017	1.986	2.069
β'_3	3.006	0.009	2.987	3.035
ϕ'	27547.830	84977.697	164.846	433256.000

Fonte: Elaboração Própria

Como se pode notar pela tabela, o modelo logrou sucesso em estimar os coeficientes utilizados para gerar os dados. Há no entanto maior incerteza em relação aos mesmos, com um aumento de quase todos os desvios-padrão em comparação à Regressão de Poisson. O parâmetro de dispersão ϕ foi estimado com um valor bastante alto, mas também com bastante incerteza em torno dele. Essa maior incerteza do modelo se manifestou também na estatística \hat{R} , cujo valor mais distante de 1 equivale a 1,187, o que é relativamente elevado, indicando dificuldade de convergência no modelo.

Os modelos de Poisson e Binomial Negativa ajustados aos dados sintéticos foram comparados com o uso da técnica de validação cruzada *leave-one-out*. Essa técnica consiste em ajustar o modelo para $N - 1$ registros e realizar a verificação preditiva em relação ao registro faltante, repetindo o processo até realizar N validações. Ela permite identificar de forma bastante segura a performance preditiva do modelo. O resultado corresponde à somatória do logaritmo da densidade preditiva para todas as observações testadas pela validação cruzada, de acordo com a seguinte equação:

$$\text{LPD-LOO} = \sum_{i=1}^n \log p(y_i \mid \text{dados}_{-i}) \quad (3.35)$$

Dessa forma, resultados mais elevados para a validação cruzada indicam que o modelo teve uma performance melhor no ajuste aos dados, pois a densidade preditiva para as observações foi maior. Para reduzir os custos computacionais na amostragem dos coeficientes, foi utilizada a técnica de suavização de Pareto na amostragem (*Pareto smooth importance sampling*) (VEHTARI; GELMAN; GABRY, 2016). O resultado da comparação entre os modelos está disposto na Tabela 3.4.

Tabela 3.4 – Comparação com Dados Simulados – Poisson e Binomial Negativa

Modelo	Posicao	LOO
Poisson	0	-203.544
Binomial Negativa	1	-208.676

Fonte: Elaboração Própria

Como se pode notar, a Regressão de Poisson teve performance superior à Regressão Binomial Negativa. Esse resultado é esperado, pois os dados sintéticos foram produzidos de forma a ajustarem-se a uma distribuição de Poisson, sem superdispersão. Por isso o valor elevado de ϕ que foi amostrado pela Regressão Binomial Negativa está correto, o que indica que o modelo passou na checagem.

Após a realização dessa checagem, o modelo de Regressão de Binomial Negativa foi então ajustado aos dados reais do estudo. A Figura 3.24 apresenta o gráfico de rastreamento (*traceplot*) das Cadeias de Markov do modelo. Como se pode notar, houve convergência entre as cadeias, o que é confirmado também pela estatística \hat{R} , cujo valor mais distante de 1 foi igual a 1,003¹⁸.

A Tabela 3.5 apresenta os resultados da comparação das Regressões de Poisson e Binomial Negativa ajustadas aos dados reais. Como se pode verificar, dessa vez a Regressão Binomial Negativa teve performance superior à regressão de Poisson, o que é esperado devido ao caráter de superdispersão da variável dependente.

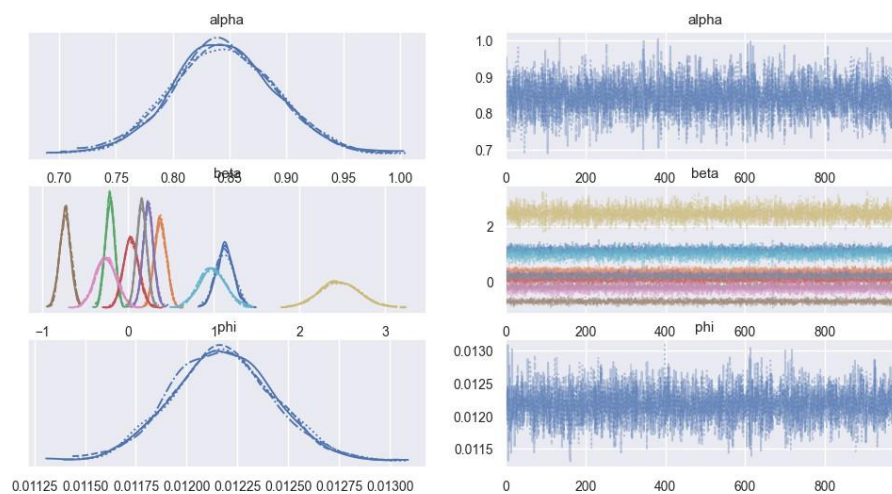
Tabela 3.5 – Comparação com Dados Reais – Poisson e Binomial Negativa

Modelo	Posicao	LOO
Binomial Negativa	0	-2.603e+04
Poisson	1	-1.004e+06

Fonte: Elaboração Própria

A Tabela 3.6 apresenta os resultados das distribuições *a posteriori* dos coeficientes amostrados pelas Cadeias de Markov do modelo de Regressão

¹⁸O amostrador gerou alguns avisos (*warnings*) pois houve valores amostrados da distribuição Gamma inferiores a 0. Todos os valores são muito próximos a zero e são devidos a questões computacionais relacionadas à precisão de pontos flutuantes.

Figura 3.24 – Traceplot – Binomial Negativa

Fonte: Elaboração própria

Binomial Negativa. Serão comentadas as alterações mais significativas nos parâmetros.

Tabela 3.6 – Descrição da tabela

Coeficiente	Média	Desvio-padrão	HDI 0.5%	HDI 99.5%
α	0.845	0.045	0.723	0.960
val_m2_terreno	1.131	0.102	0.874	1.385
idade_imoveis	0.379	0.068	0.221	0.559
idade_contribuintes	-0.212	0.055	-0.350	-0.077
media_pavimentos	0.037	0.087	-0.184	0.255
prop_maior_terreno	0.237	0.060	0.075	0.384
prop_area_hor	-0.732	0.064	-0.886	-0.566
prop_area_verti	-0.258	0.119	-0.538	0.034
prop_area_galpoes	0.160	0.058	0.012	0.306
qd_area	2.469	0.230	1.905	3.047
zeu	0.983	0.149	0.626	1.353
ϕ	0.012	0.000	0.012	0.013

Fonte: Elaboração Própria

Houve uma redução significativa no intercepto, de modo que a taxa basal esperada de unidades produzidas por quadra reduziu das aproximadamente 4,5 unidades estimada pela Regressão de Poisson para 2,3 unidades

em média estimadas pela Regressão Binomial Negativa. Essa redução é positiva, pois indica que há uma maior porção da variabilidade dos dados explicada pelas variáveis de controle e pela variável tratamento. Ademais, ela aproxima os resultados do modelo aos dados reais, pois, como veremos na seção 3.3.4 boa parte das quadras da cidade não recebeu nenhum empreendimento imobiliário no período do estudo. É possível, no entanto, que a área da quadra tenha assumido o papel do intercepto, devido ao aumento substancial na magnitude de seu efeito. Isso indica que há ainda variáveis de controle a serem consideradas.

A variável que representa o valor de metro quadrado de terreno teve um aumento substancial em seu efeito, de modo que um aumento de um desvio padrão no valor do metro quadrado de terreno em uma determinada quadra passou a acarretar um aumento de mais de 200% na quantidade esperada de unidades produzidas. Isso mostra que a Regressão de Poisson não estava tendo sucesso em capturar os empreendimentos de grande porte, para os quais é de grande importância que o valor total de venda da área construída ultrapasse os custos de construção, como mencionado anteriormente.

As variáveis que medem a verticalização na quadra – proporção da área construída de edificações de tipo vertical e a média da quantidade de pavimentos – deixaram de ser significantes. É possível que isso tenha se dado devido à nova importância assumida pelo valor do metro quadrado do terreno, pois áreas mais verticalizadas tendem a estar situadas em regiões com maior valor por metro quadrado. Em relação às demais variáveis de controle, não houve alterações dignas de nota.

A variável de tratamento teve um aumento relevante. De acordo com o novo modelo, o fato da quadra estar situada em ZEU acarreta um aumento de 167% na quantidade esperada de unidades de tipo vertical produzidas no período. Por fim, o parâmetro ϕ possui um valor reduzido e com uma distribuição de probabilidade bastante concentrada. Isso indica que há de fato superdispersão nos dados, conforme exposto nesta seção, e que o modelo foi capaz de capturar essa característica com sucesso.

3.3.4 Restrições construtivas e zeros inflados

Contudo, à luz dos resultados da análise da estrutura de segregação espacial na cidade de São Paulo apresentados na seção 3.3.1, cumpre verificar se o efeito da variável de tratamento apresentado na seção 3.3.3 na realidade não está se confundindo com a localização das Zonas de Estruturação Urbana, que estão situadas em maior proporção na *área nobre* da cidade, para a qual está voltado, historicamente, o interesse do capital imobiliário.

Para ilustrar essa questão, a Figura 3.25, apresentada anteriormente na seção 3.2.2 é retomada. Note-se que há uma maior concentração dos Eixos de Estruturação da Transformação Urbana (que deram origem às ZEU) no setor Sudoeste, onde se situa o *cluster* Alto-Alto que identifica a área nobre da cidade.

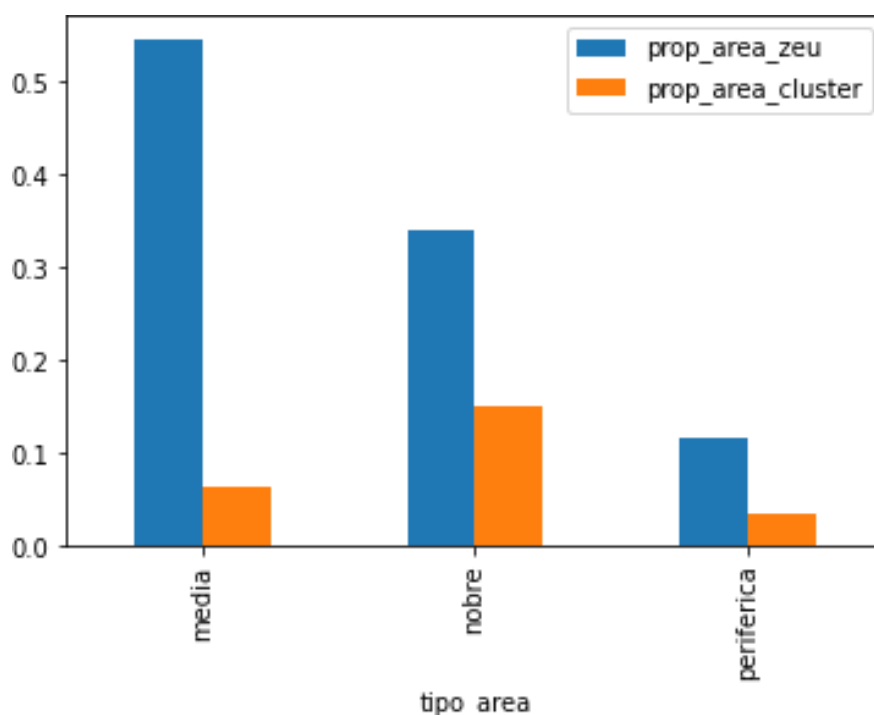
Figura 3.25 – Localização dos Eixos de Estruturação da Transformação Urbana



Fonte: Portal GeoSampa

A Figura 3.26 apresenta os dados de distribuição da área das quadras situadas em ZEU com relação às diferentes regiões que compõem a estrutura de segregação social da cidade (área nobre, periférica e de classe média). As barras em laranja correspondem à proporção da área total das quadras classificadas como ZEU em relação à área total da região da cidade em que elas estão situadas. Como se pode notar, 15% da área total das quadras na área nobre da cidade estão classificadas como ZEU. Esse valor está muito acima das demais regiões: 6% da área total das quadras situadas nas áreas de classe média da cidade corresponde a ZEU, contra apenas 3% da área total das quadras situadas nas áreas periféricas.

Figura 3.26 – Distribuição da área das quadras em ZEU



Fonte: Elaboração Própria

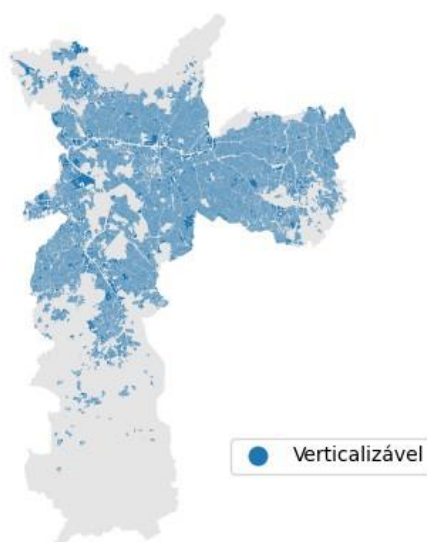
As barras em azul na Figura 3.26, por sua vez, apresenta a distribuição da área total das quadras classificadas como ZEU entre as diferentes regiões da cidade. Assim, 54% da área total das quadras em ZEU está situada na região de classe média da cidade, 34% dessa área total está na área nobre e apenas 12% se localiza na área periférica. Se considerarmos

que a área total das quadras situadas na área nobre da cidade corresponde a apenas 16% da área total das quadras do município, pode-se concluir com segurança que há uma sobrerepresentação das ZEU nessa região.

Além disso, conforme discutido na seção 3.2.4, há que se considerar que vigoram no município uma série de medidas que restringem a edificação das quadras. Essas medidas não estão distribuídas equitativamente nas diferentes regiões da cidade. O mapa da Figura 3.27 apresenta a distribuição das quadras da cidade que não são afetadas por essas restrições, que foram identificadas em azul. O mapa foi gerado de acordo com a metodologia apresentada na seção D.5 do Anexo Único.

Figura 3.27 – *Distribuição espacial da área verticalizável*

Quadras Urbanas Verticalizáveis - São Paulo



Fonte: Elaboração Própria

Considerando-se a distribuição das áreas edificáveis nas diferentes regiões que compõem a estrutura de segregação urbana em São Paulo, dois principais pontos advém da análise da distribuição espacial apresentada nesse mapa. Em primeiro lugar, algumas áreas periféricas, sobretudo aquelas localizadas nos extremos Noroeste e Sul da cidade são desproporcionalmente afetadas por essas restrições. Isso se deve sobretudo ao fato de

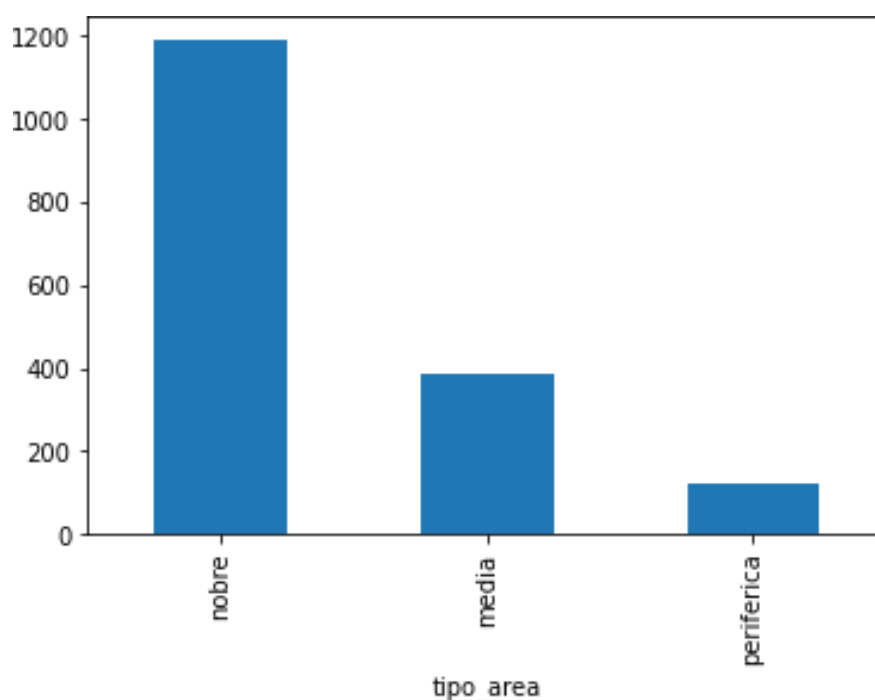
que essas áreas contém os principais remanescentes de vegetação nativa da cidade, além dos mananciais. Elas estão portanto recobertas por diferentes formas de proteção ambiental e de proteção às reservas de água do município.

Em segundo lugar, nota-se que há porções significativas de quadras com restrições construtivas na área nobre da cidade. Trata-se de bairros tradicionais de residência da elite nos quais foram impostas diversas medidas que têm por objetivo impedir a verticalização. Dentre essas medidas, destacam-se as Zonas Exclusivamente Residenciais Unifamiliares (ZER-1). Trata-se do zoneamento conhecido na literatura internacional como *single family detached housing*, no qual há uma série de parâmetros urbanísticos que impedem a construção de edifícios residenciais. Dentre os parâmetros instituídos em São Paulo para esse fim, destaca-se a definição de um gabarito de altura máximo de 10 metros, o coeficiente de aproveitamento máximo igual a um e, principalmente, a proibição de usos condominiais (as habitações devem ser unifamiliares). A presença dessas restrições tem por efeito reduzir significativamente a área disponível para empreendimentos imobiliários na área nobre da cidade, como se pode notar pelo mapa na Figura 3.27.

Há que se considerar, por fim, que as áreas nobre, de classe média e periférica da cidade possuem diferenças significativas em relação ao valor da terra. A Figura 3.28 apresenta os valores médios do valor médio do metro quadrado de terreno das quadras situadas nessas regiões da cidade. Note-se que trata-se de valores nominais para o ano de 2014, tal qual registrados na base do IPTU, de modo que eles não refletem os valores de mercado atuais, que são muitos superiores¹⁹.

A análise do gráfico mostra que os valores de terreno na área nobre são substancialmente superiores do que aqueles situados nas demais regiões da cidade. Além disso, de forma consistente, as áreas de classe média possuem valores superiores aos das áreas periféricas. No entanto, essa diferença é pequena se comparada à disparidade de valores entre a área nobre e a área

¹⁹É possível, inclusive, que eles não reflitam também os valores de mercado no período em que foram coletados, pois a atualização da Planta Genérica de Valores apenas nos últimos anos passou a utilizar bases de dados mais representativas. Dessa forma, é possível que a disparidade entre as áreas nobre e periférica seja ainda maior, devido à subvalorização dos terrenos na área nobre.

Figura 3.28 – Valor nominal do metro quadrado de terreno

Fonte: Elaboração Própria

de classe média.

Como afirmado anteriormente, o valor da terra é um dos principais determinantes para viabilizar o lucro de empreendimentos imobiliários (BERTAUD, 2022b) – o que foi corroborado pelos modelos desenvolvidos nas seções 3.3.2 e 3.3.3. Dessa forma, a disparidade entre a área nobre e as demais regiões da cidade assinalada pelo gráfico contido na Figura 3.28 fornece evidências confirmatórias sobre a afirmação de que a área nobre concentra o interesse do capital imobiliário, realizada na seção 3.2.5 com base em Villaça (2001). Se considerarmos essa informação à luz da distribuição da área edificável na área nobre que, como foi exposto, é bastante reduzida, decorre que há uma concentração ainda maior do interesse do capital imobiliário sobre as quadras que podem receber empreendimentos verticais.

Em resumo, os dados até aqui expostos nesta seção assinalam que:

- As Zonas de Estruturação Urbana estão desigualmente distribuídas

nas áreas nobre, de classe média e periférica da cidade, havendo uma concentração significativa na área nobre;

- Há evidências de que a área nobre concentra o interesse do mercado imobiliário pois é a região com o maior valor de metro quadrado;
- As restrições à verticalização na área nobre resultam em uma concentração ainda maior desse interesse;
- As demais regiões da cidade são afetadas de forma desigual pelas restrições construtivas.

Depreende-se disso que é provável que os incentivos construtivos estabelecidos pelo Plano Diretor com o objetivo de direcionar a produção imobiliária para as Zonas de Estruturação Urbana tenham efeitos distintos nas diferentes regiões da cidade. Em especial, é possível que esses efeitos sejam muito superiores na área nobre, na qual se concentra o interesse do mercado imobiliário. Note-se que é nessa área também que esse interesse encontra-se restringido pelo Zoneamento Exclusivamente Residencial Unifamiliar e as demais medidas que impedem a verticalização. É possível, portanto, que devido à maior possibilidade de edificação, as quadras em **ZEU** canalizem o interesse do capital imobiliário na área nobre.

Mas não há garantia que o mesmo ocorra nas demais regiões da cidade, em especial na área periférica, em relação à qual há evidências de que o interesse do capital imobiliário não é significativo. Uma das principais evidências para isso é o fato de que as áreas periféricas são, até hoje, muito pouco verticalizadas. Houvesse interesse do capital imobiliário sobre elas, teriam sido construídos edifícios nas últimas décadas nas quadras que não estão afetadas por restrições construtivas – o que não ocorreu. Como pode ser conferido por meio da Figura 2.2, as áreas periféricas possuem uma densidade construtiva de tipo vertical bastante reduzida.

Nesse cenário, uma vez que há uma maior proporção de quadras em **ZEU** na área nobre, é possível que o efeito substancial para a variável de tratamento identificado na seção 3.3.3 mascare uma inefetividade nas demais áreas da cidade. Além disso, é preciso também levar em consideração a distribuição desigual das restrições construtivas, que afetam em maior proporção as áreas periféricas. O modelo estatístico desenvolvido nesta seção

endereça essas duas questões.

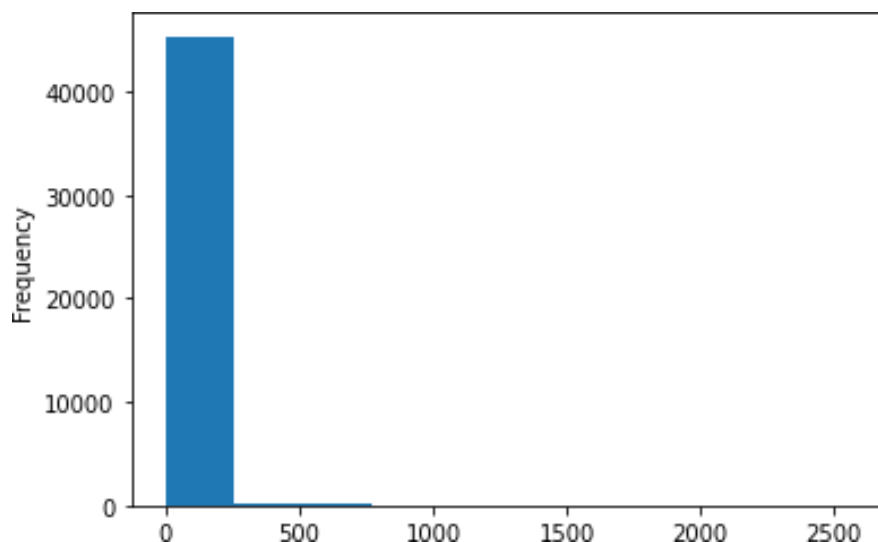
Com o objetivo de analisar as diferenças nos efeitos das ZEU nas regiões que compõem a estrutura espacial de segregação urbana em São Paulo, foram criadas duas variáveis *dummie* que identificam a interação entre a localização em ZEU e a região da cidade em que a quadra está situada. Essas variáveis são *zeu_area_nobre*, que registra as quadras com zoneamento do tipo ZEU situadas na área nobre e *zeu_periferia*, que identifica as quadras em ZEU localizadas nas áreas periféricas da cidade. As áreas de classe média foram codificadas como a categoria de referência, de modo que a variável *zeu* passará a identificar o efeito das ZEU sobre a quantidade esperada de unidades de tipo vertical produzidas nesta região da cidade.

As restrições construtivas, por sua vez, foram modeladas como correspondendo a um processo gerativo distinto, que pode levar à anulação da quantidade esperada de unidades verticais produzidas na quadra independentemente de seu potencial construtivo. Isso é evidente caso se considere, por exemplo, uma quadra situada em uma região com alto valor do metro quadrado de terreno e com um grande terreno pouco edificado mas que está tombado. Neste caso, pouco importa o potencial construtivo da quadra, pois o tombamento impede a realização de empreendimentos imobiliários. Mas essa não é a única razão pela qual uma dada quadra pode ter uma quantidade esperada de unidades produzidas igual a zero, pois é possível, por exemplo, que ela se situe em uma região com um valor de metro quadrado tão reduzido que inviabilize a realização de empreendimentos verticais.

Como demonstrado na seção 3.2.7.4 esse cenário, no qual há ocorrência tanto de *zeros estruturais* quanto de *zeros de contagem* é adequadamente modelado por Modelos de Zeros Inflados (ZIP). A necessidade de ajuste de um modelo de zeros inflados é reforçada pela consideração da distribuição da variável dependente deste estudo, que possui elevada concentração de valores zero. O que é de se esperar, pois a verticalização de uma proporção elevada das quadras da cidade seria muito improvável, de forma que é mais provável que uma dada quadra não tenha recebido um empreendimento imobiliário. A Figura 3.29 apresenta o histograma dessa variável.

Antes de prosseguir com a apresentação dos resultados do modelo, é preciso expor uma última consideração metodológica. Pois a inclusão das

Figura 3.29 – *Histograma – Quantidade de unidades verticais produzidas por quadra 2014-2023*



Fonte: Elaboração Própria

variáveis *dummie* que identificam o pertencimento às diferentes regiões da cidade implicam a necessidade de abrir mão de alguns dos controles estatísticos que foram utilizados nos modelos expostos nas seções anteriores. Isso é necessário pois as diferentes regiões que compõem a estrutura espacial de segregação social da cidade possuem padrões de uso e ocupação do solo distintos. Esses padrões, assim como as razões históricas e sociais para sua conformação, são descritos em detalhes por Villaça (2001). Para as finalidades deste estudo, basta que se considere novamente o gráfico apresentado na Figura 3.28, que mostra como o valor do metro quadrado de terreno varia consideravelmente entre as áreas nobre, de classe média e periférica da cidade. Ou ainda a Figura 2.2 que mostra como a densidade construtiva de tipo vertical varia consideravelmente no território da cidade.

O valor médio do metro quadrado de terreno, por exemplo, está substancialmente correlacionado a dois outros controles utilizados nos modelos anteriores, a média da quantidade de pavimentos das edificações situadas na quadra e a proporção da área construída total de edificações verticais em relação à área da quadra. Essas variáveis possuem correlação de Pearson positiva em relação ao valor médio do metro quadrado de terreno, com

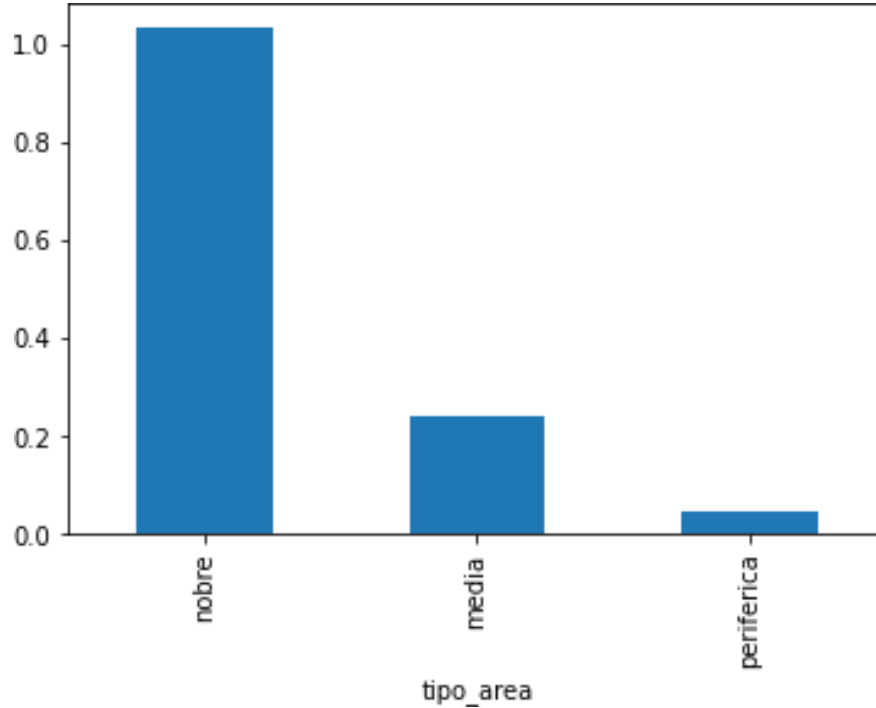
ρ igual a 0,52 e 0,62, respectivamente. Ambas as variáveis – área edificada vertical e média de pavimentos – estão também fortemente correlacionadas entre si, com $\rho = 0,80$. Considerando-se que o valor médio do metro quadrado de terreno está distribuído de forma desproporcional entre as áreas nobre, de classe média e periférica da cidade, é de se esperar que as variáveis aqui consideradas tenham a mesma distribuição desigual.

Dessa forma, empregar as variáveis utilizadas nos modelos anteriores como controles para o potencial construtivo pode acarretar na identificação das regiões da cidade em que as quadras estão situadas. Isso se torna evidente caso se considere o gráfico de barras na Figura 3.30. Como se pode notar, a proporção média da área construída de tipo vertical em relação à área das quadras varia significativamente nas diferentes regiões da cidade. As quadras na área nobre, por exemplo, têm em média mais de 100% de suas áreas de terreno edificadas por empreendimentos verticais. McElreath (2020, cap. 6)

As variáveis de controle que estão associadas ao padrão de uso e ocupação solo e ao valor da terra são portanto *maus controles* (CINELLI; FORNEY; PEARL, 2022a). A inclusão dessas variáveis no modelo poderia resultar na subestimação do tamanho de efeito das variáveis tratamento, que agora estão sendo medidas em sua interação com as diferentes regiões da cidade (McElreath (2020, cap. 6) desenvolve em detalhes essa questão). Dessa forma, o modelo aqui desenvolvido utiliza menor quantidade de controles que os modelos anteriores.

Como afirmado anteriormente, o modelo cujos resultados serão apresentados nesta seção é uma Regressão Binomial Negativa para Zeros Inflados. Trata-se de um modelo de mistura que combina um processo gerativo de contagem, modelado por meio de uma distribuição Binomial Negativa, com um processo gerativo estrutural, modelado por meio de uma distribuição de Bernoulli, conforme detalhado na seção 3.2.7.4. O modelo foi especificado da seguinte forma:

Figura 3.30 – Média da proporção da área construída vertical por região da cidade



Fonte: Elaboração Própria

$$\begin{aligned}
 & \text{se } y_i = 0, \\
 & P(y_i = 0 \mid X_i, M_i) = \text{Bernoulli}(0 \mid \tau + M_i \vartheta) \\
 & + \text{Bernoulli}(1 \mid \tau + M_i \vartheta) \times \text{NegBinom}(\alpha + X_i \beta, \phi) \\
 \forall i \in N, \\
 & \text{se } y_i > 0, \\
 & P(y_i = y \mid y > 0, X_i, M_i) = \text{Bernoulli}(1 \mid \tau + M_i \vartheta) \\
 & \times \text{NegBinom}(\alpha + X_i \beta, \phi)
 \end{aligned} \tag{3.36}$$

Nessa especificação, ϑ corresponde a um vetor de variáveis explicativas associadas ao *processo de zeros estruturais*, modelado por uma Regressão Logística com distribuição de Bernoulli. O parâmetro τ corresponde ao intercepto do preditor linear dessa Regressão Logística. Note-se que o modelo

está parametrizado da segunda forma descrita na seção 3.2.7.4, na qual o sucesso da distribuição de Bernoulli sinaliza que o *processo de contagem* não foi anulado. No contexto da pesquisa aqui desenvolvida, probabilidades maiores que 0,5 amostradas da distribuição de Bernoulli a partir da Regressão Logística indicam que a quadra pode ser verticalizada. Assim, coeficientes negativos para os parâmetros ϑ indicam uma redução na probabilidade de que a quadra tenha recebido um empreendimento imobiliário – por exemplo devido à existência de um tombamento na quadra, identificado por ϑ_i .

A modelagem do *processo de contagem* segue os mesmos parâmetros da Regressão Binomial Negativa exposta na seção 3.3.3. Os interceptos α e τ , assim como os vetores de coeficientes angulares β e ϑ tiveram suas distribuições *a priori* especificadas com a mesma distribuição Normal pouco informativa apresentada na seção 3.3.2: $N(0, 2.5)^2$. Assim como o parâmetro de dispersão da Distribuição Binomial Negativa, que foi especificado por meio de sua inversão e recebeu a mesma *distribuição a priori* apresentada na seção 3.3.3.

Como para os modelos anteriores, foi realizada a checagem preditiva das distribuições *a priori*. Para isso, foram gerados dados sintéticos de acordo com a seguinte equação:

$$\forall i \in N, \quad y_i' = \text{Bernoulli}(\tau' + M_i' \vartheta') \times \text{NegBinom}(\alpha' + X_i' \beta', \phi') \quad (3.37)$$

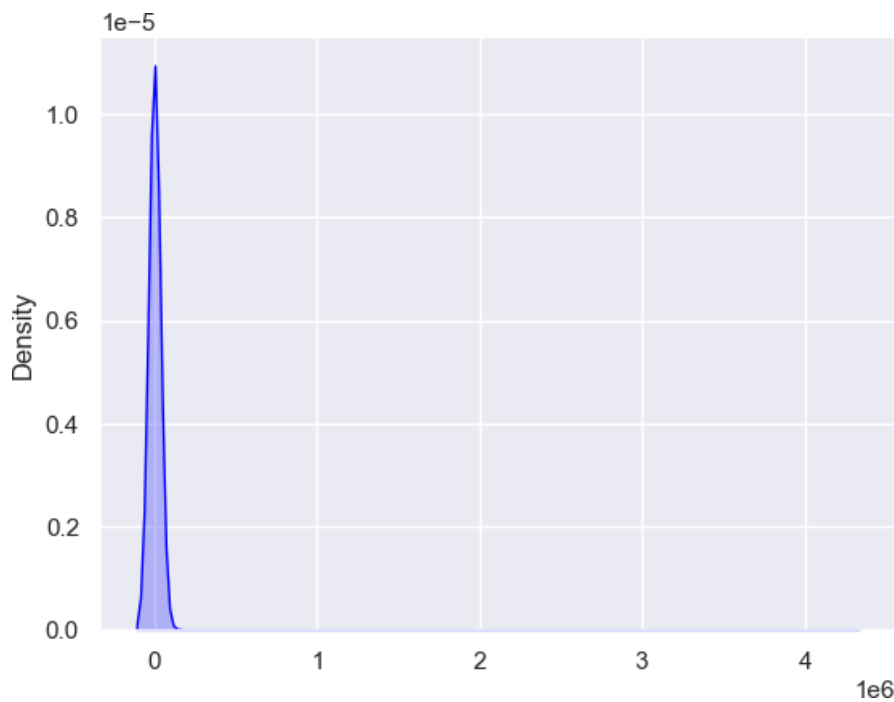
Onde:

- $\tau' = 0,5$
- $\alpha' = 1$
- $\beta' = \begin{matrix} 1 & & 2 \\ 1 & 2 & 3 \end{matrix}$
- $\vartheta' = \begin{matrix} 1 & & 2 \\ 3 & 5 & -2 \end{matrix}$
- $\phi' = 0,1$

As matrizes de dados sintéticos X' e M' são matrizes 100×3 amostradas de uma distribuição Normal Padrão. A Figura 3.31 apresenta o gráfico

de densidade de Kernel da distribuição da variável simulada y' . Como se pode notar, a simulação teve sucesso em produzir uma distribuição com zeros inflados similar à da variável dependente deste estudo.

Figura 3.31 – Distribuição dos dados simulados - Zeros Inflados



Fonte: Elaboração Própria

A Tabela 3.7 apresenta as distribuições *a posteriori* dos coeficientes amostrados pelo modelo. Como se pode notar, o modelo logrou sucesso em identificar os coeficientes simulados β referentes ao processo de contagem, mas teve maior dificuldade em relação aos coeficientes ϑ referentes ao processo estrutural. Isso provavelmente ocorreu pois trata-se de um modelo com maior quantidade de graus de liberdade que os anteriores, que possui além disso uma estrutura funcional complexa, de modo que os apenas 100 registros sintéticos gerados não foram suficientes para um ajuste pleno. O modelo no entanto convergiu com sucesso, sendo que o \hat{R} que mais se distanciou de 1 equivale a 1,002.

Por essas razões, optou-se por prosseguir com o ajuste do modelo para os dados reais. Como discutido anteriormente, foram utilizados apenas as

Tabela 3.7 – Descrição da tabela

Coeficiente	Média	Desvio-padrão	HDI 0.5%	HDI 99.5%
<i>alpha</i>	3.229	0.031	3.170	3.291
<i>beta</i> ₀	1.033	0.025	0.982	1.080
<i>beta</i> ₁	1.937	0.027	1.884	1.989
<i>beta</i> ₂	2.981	0.030	2.922	3.040
<i>phi</i>	5.406	0.662	4.122	6.692
<i>tau</i>	-0.222	0.133	-0.491	0.027
<i>theta</i> ₀	2.327	0.241	1.867	2.815
<i>theta</i> ₁	3.866	0.356	3.181	4.577
<i>theta</i> ₂	-1.731	0.195	-2.120	-1.362

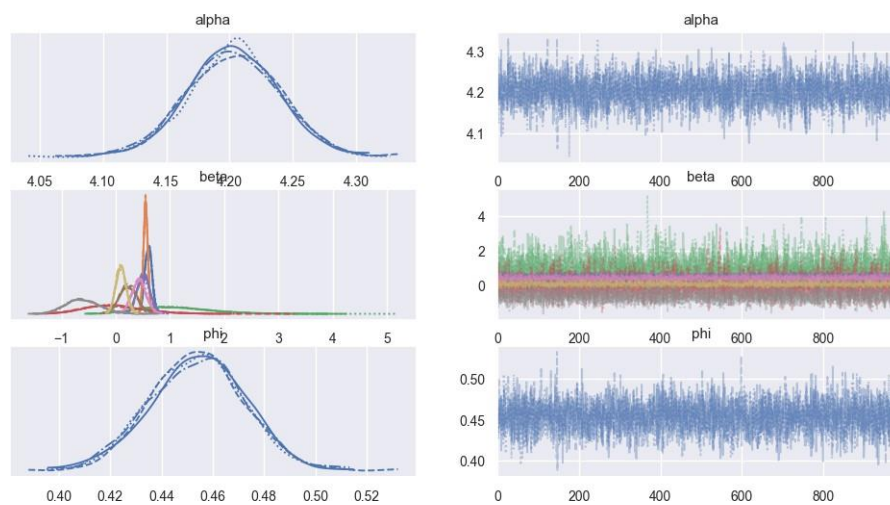
Fonte: Elaboração Própria

variáveis de controle que não estão diretamente relacionadas ao padrão de uso e ocupação do solo. Além disso, foram acrescentadas como controles as variáveis relativas às restrições construtivas apresentadas no Quadro 3.1 da seção 3.2.6. Uma vez que algumas dessas restrições não são absolutas, sendo possível edificar nas quadras que incorrem nesses casos, optou-se por incluir essas variáveis como controles no processo de contagem. Este é o caso, por exemplo, da identificação de que a quadra está sob rampa de pouso e decolagem de helipontos. Pois, neste caso, a restrição não é total: ela se aplica apenas à altura máxima do empreendimento, para evitar colisões com as aeronaves.

O modelo convergiu com sucesso, com todos os \hat{R} satisfatórios – o mais distante de 1 corresponde a 1.002. A Figura 3.32 apresenta o *traceplot* para as Cadeias de Markov dos parâmetros associados ao processo de contagem. A Tabela 3.8 apresenta os resultados das distribuições *a posteriori* dos coeficientes que modelam esse processo, ajustados para os dados reais do estudo. Trata-se dos parâmetros de intercepto α e de dispersão ϕ , além do vetor θ , cujas variáveis foram identificadas.

A Figura 3.33, por sua vez, contém o *traceplot* para as Cadeias de Markov dos parâmetros associados ao processo estrutural. E a Tabela 3.9 apresenta os resultados das distribuições *a posteriori* dos coeficientes que modelam esse processo, o intercepto τ e o vetor de coeficientes angulares ϑ , cujas variáveis também foram identificadas.

Uma vez que esse modelo não contém as mesmas variáveis explicativas

Figura 3.32 – Traceplot ZIP - Contagem

Fonte: Elaboração Própria

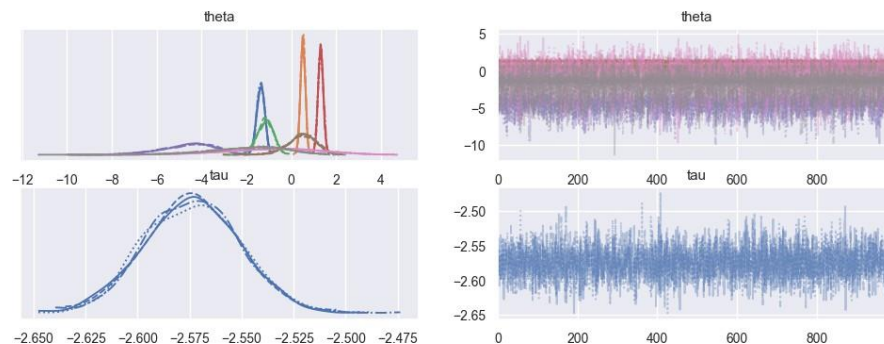
Tabela 3.8 – Descrição da tabela

Coefficiente	Média	Desvio-padrão	HDI 2.5%	HDI 97.5%
α	4.203	0.036	4.134	4.275
ϕ	0.454	0.018	0.417	0.488
qd_area	0.603	0.065	0.478	0.727
prop_maior_terreno	0.551	0.041	0.472	0.633
area_protec_amb	1.086	0.676	-0.161	2.383
rest_geotec	-0.009	0.553	-1.009	1.110
envol_tombamento	0.510	0.114	0.295	0.739
rampa_heliponto	0.226	0.157	-0.075	0.538
ZEU_area_nobre	0.427	0.129	0.166	0.673
ZEU_periferia	-0.607	0.308	-1.191	-0.000
ZEU	0.095	0.093	-0.080	0.283

Fonte: Elaboração Própria

que os modelos ajustados nas seções anteriores, 3.3.2 e 3.3.3, não é possível comparar diretamente os coeficientes das variáveis em comum.

Em primeiro lugar, sobre o processo de contagem, nota-se que o coeficiente amostrado para o parâmetro de dispersão ϕ é próximo de zero, de forma que o modelo segue captando a superdispersão dos dados. A taxa basal da quantidade esperada de unidades produzida por quadra, α ,

Figura 3.33 – Traceplot ZIP - Estrutural

Fonte: Elaboração Própria

Tabela 3.9 – Descrição da tabela

Coeficiente	Média	Desvio-padrão	HDI 2.5%	HDI 97.5%
τ	4.203	0.036	4.134	4.275
ZER	-1.365	0.175	-1.697	-1.021
tombo	0.518	0.108	0.297	0.716
bairro_ambiental	-1.165	0.361	-1.838	-0.446
area_contaminada	1.312	0.117	1.092	1.546
protecao_aos_mananciais	-4.640	1.215	-7.097	-2.442
risco_geologico	0.459	0.645	-0.757	1.717
rppn	-0.867	2.096	-5.025	3.018
unidade_de_conservacao	-1.997	1.711	-5.269	1.129

Fonte: Elaboração Própria

corresponde a 66 unidades. O valor elevado explica-se pois a Regressão Binomial Negativa está modelando agora apenas as quadras que o modelo ZIP identificou que possuem maior probabilidade de receber empreendimentos imobiliários. Como os zeros estruturais são modelados pela Regressão Logística, o intercepto da Regressão Binomial Negativa tende assim a se elevar. A área da quadra e a proporção do maior terreno não verticalizado em relação à área da quadra possuem efeitos positivos significativos, como esperado.

As variáveis que identificam se a quadra está em área de restrição geotécnica, *rest_geotec* e rampa de heliponto, *rampa_heliponto* não possuem efeitos significativo no intervalo de confiabilidade de 95%. Trata-se,

em ambos os casos, de restrições à edificação que foram modeladas como controles para o processo de contagem por não terem caráter absoluto. Em uma próxima iteração, elas serão incluídas no processo estrutural. A área envoltória de tombamento, por sua vez, possui efeito positivo. Isso pode ocorrer devido ao fato de que há maior concentração dessas áreas envoltórias no centro histórico da cidade, de modo que elas podem estar capturando o efeito dessa localização mais central.

Prosseguindo à análise para as restrições construtivas associadas aos *zeros estruturais*, cumpre reforçar que um efeito positivo, neste caso, indica que há maior probabilidade de que o imóvel receba um empreendimento vertical, de modo que é esperado que todas as variáveis de controle possuam efeitos negativos, pois todas elas refletem restrições à edificação. Isto dito, nota-se que, assim como as áreas envoltórias, também a presença de imóvel tombado na quadra tem efeito positivo. Esse resultado provavelmente ocorre pela mesma razão que as áreas envoltórias.

Além disso, cumpre ressaltar que o tombamento impõe restrições diretamente apenas ao imóvel que foi tombado, permanecendo edificáveis os demais imóveis da quadra. Em uma nova oportunidade, seria interessante ponderar essa restrição para que ela reflita a proporção da área tombada em relação à área da quadra. É possível que o mesmo ocorra com a restrição associada às áreas contaminadas, que também tem efeito positivo. Pois, assim como o tombamento, a área contaminada gera restrições apenas para o imóvel com o terreno contaminado, permanecendo edificáveis os demais imóveis da quadra.

As variáveis relativas às unidades de conservação, [RPPN](#) e risco geológico não possuem efeito significativo no intervalo de confiabilidade de 95%. Já a localização em área de proteção aos mananciais tem forte efeito negativo, reduzindo em aproximadamente 99% a chance de que a quadra receba um empreendimento imobiliário vertical. Note-se, porém, que é possível que parte desse efeito negativo esteja associado ao valor da terra e à localização em área periférica, pois uma porção significativa das áreas de manancial na Zona Sul estão contidas em um dos *clusters* Baixo-Baixo identificados na seção [3.3.1](#).

As variáveis relacionadas a medidas protetivas contra a verticalização dos bairros de residência da elite, nomeadamente o zoneamento de tipo

ZER e o tombamento dos bairros ambientais, popularmente conhecidos como *bairros jardim*, têm efeito negativo significativo ²⁰. Elas reduzem a chance de a quadra receber um empreendimento vertical em 25% e 31% respectivamente. Esse resultado indica que as restrições construtivas presentes na área nobre são efetivas, fornecendo evidências para confirmar a hipótese de concentração do interesse do mercado imobiliário para as quadras edificáveis nessa região, dentre as quais destaca-se as quadras em **ZEU**.

Em relação às variáveis tratamento, nomeadamente *zeu*, *zeu_periferia* e *zeu_area_nobre*, sua interpretação requer um cuidado metodológico adicional por se tratarem de interações. Como argumenta McElreath (2020, cap. 7), os efeitos das variáveis que indicam interações devem ser interpretados em relação à categoria de referência. Dessa forma, o efeito da variável *zeu_area_nobre*, aqui representado por ζ , é definido pela combinação linear de seu coeficiente, θ_1 , com o coeficiente da variável *zeu*, θ_2 , conforme a equação $\zeta = e^{\theta_1 + \theta_2}$.

Como foi exposto, o coeficiente da variável *zeu* representa o efeito dos incentivos construtivos fornecidos pelo Plano Diretor nas áreas dos Eixos de Estruturação da Transformação Urbana localizadas na região da cidade definida como a categoria de referência, isto é, a área de classe média. Essa variável não possui efeito significativo no intervalo de confiança de 95%. Dessa forma, segundo o modelo, na região de classe média da cidade, a localização da quadra em **ZEU** não surte efeito definido sobre a quantidade esperada de unidades de tipo vertical produzidas no período de vigência do **PDE**. Note-se que o fato de o coeficiente não ser significativo não inviabiliza a interpretação das demais interações.

A análise do coeficiente para a variável *zeu_periferia*, que é significativo no intervalo de confiabilidade de 95%, indica que uma quadra situada em **ZEU** e localizada em área periférica possui uma quantidade esperada de unidades verticais produzidas no período 45% menor do que a quantidade esperada para uma quadra com as mesmas características, mas localizada na área de classe média. Em termos gerais, não mais em comparação a uma quadra situada na área de classe média, o fato de uma quadra estar localizada em **ZEU** na área periférica tem uma redução de 40% na quantidade esperada de unidades.

²⁰A seção **D.4** apresenta maiores detalhes sobre essas variáveis

O coeficiente para a variável `zeu_area_nobre` também é significativo. Assim, uma quadra localizada na área nobre e situada em [ZEU](#), de acordo com o modelo, tem um aumento de 53% na quantidade esperada de unidades verticais produzidas no período em relação a uma quadra situada em [ZEU](#) e localizada em área de classe média, tudo o mais mantido igual. Em termos gerais, esse aumento é de 69%.

Os resultados do modelo aqui desenvolvido apresentam, portanto, evidências de que os incentivos construtivos fornecidos pelo [PDE](#) para direcionar a produção imobiliária para as áreas próximas aos eixos de transporte da cidade tiveram efeitos distintos nas diferentes regiões que compõem a estrutura espacial de segregação urbana de São Paulo.

Por fim, a Tabela 3.10 apresenta os resultados da comparação, com o uso da validação cruzada *leave-one-out*, entre o modelo de Zeros Inflados desenvolvido nesta seção e a Regressão Binomial Negativa ajustada na seção anterior. Como se pode notar, apesar de ter sido necessário abrir mão de diversas variáveis de controle para as quais os modelos anteriores já havia fornecido evidências de que são fatores relevantes na explicação da variável dependente, ainda assim a Regressão Binomial Negativa com Zeros Inflados teve melhor performance.

Tabela 3.10 – Comparação – Binomial Negativa e Zeros Inflados

Modelo	Posicao	LOO
Binomial Negativa com Zeros Inflados	-25759.51	-2.603e+04
Binomial Negativa	1	-26032.20

Fonte: Elaboração Própria

3.3.5 Desigualdade e modelos multinível

Como discutido na seção 3.2.7.5, o modelo de Regressão Binomial Negativa para Zeros Inflados desenvolvido na seção anterior, apesar de apresentar evidências sobre o comportamento distinto dos incentivos construtivos nas diferentes regiões que compõem a estrutura espacial de segregação na cidade, possui no entanto uma limitação. Pois ele considera as quadras individualmente, não sendo capaz de capturar as tendências internas a cada uma das regiões da cidade no que diz respeito à dinâmica da produção imo-

bilíaria. Além disso, como discutido na seção 3.3.3, essa mesma limitação acarretou a necessidade de abrir mão de uma série de variáveis de controle que já haviam se mostrado significativas, pois elas poderiam identificar a região da cidade e assim resultar na subestimação do efeito das variáveis de tratamento.

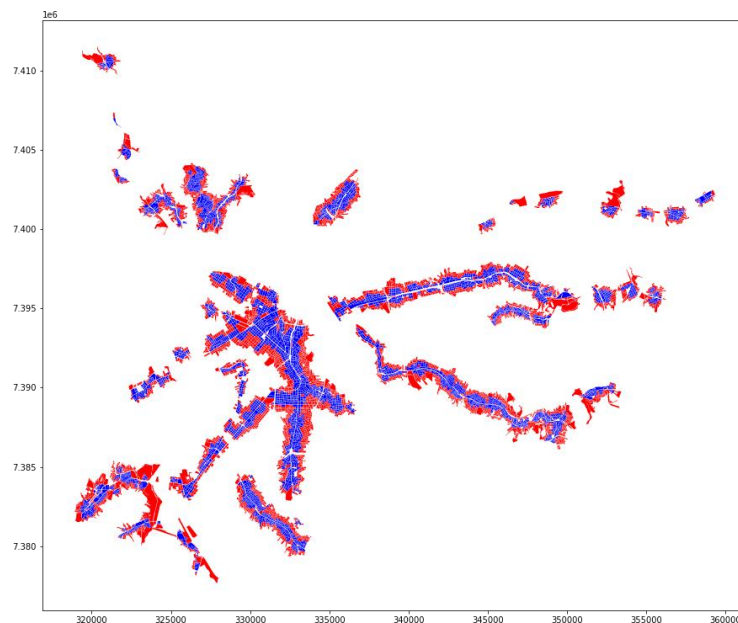
O desenvolvimento de um modelo multinível que tem como primeiro nível as quadras da cidade e como segundo nível as regiões nas quais elas se situam – área nobre, de classe média ou periférica – possibilita incorporar como controles todas as variáveis listadas no Quadro 3.1 da seção 3.2.6. Pois neste modelo o coeficiente angular do valor do metro quadrado de terreno, por exemplo, é ajustado *internamente* à região. Como exposto na seção 3.2.7.5 essa abordagem permite identificar não apenas como esse coeficiente pode se alterar dependendo da região, mas também é sensível às correlações entre os demais parâmetros, resultando em um melhor ajuste e uma representação mais fiel da dinâmica da produção imobiliária nessa região.

Contudo, dada a complexidade da função de verossimilhança do modelo multinível aqui desenvolvido, que estende o modelo de Zeros Inflados ajustado na seção 3.3.4, não foi possível gerar dados sintéticos para realizar a checagem preditiva das distribuições *a priori*. Optou-se por outra alternativa para verificar a adequação da parametrização do modelo: treiná-lo com um subconjunto definido da base de dados antes de prosseguir para a totalidade dos registros.

Para construir o subconjunto de dados, foi desenvolvida uma estratégia de amostragem que consistiu em identificar, para cada uma das quadras localizadas em ZEU da cidade, a quadra mais próxima que não está em ZEU e que não foi selecionada anteriormente. O mapa contigo na Figura 3.34 apresenta os resultados dessa amostragem. Nele estão plotadas, em azul, as quadras localizadas em ZEU e, em vermelho, as demais quadras amostradas segundo o método aqui descrito. Como se pode notar, as quadras em ZEU estão consistentemente circundadas pelas quadras amostradas, não havendo nenhuma quadra em vermelho que não seja contígua ou que se situe excessivamente distante de uma quadra em ZEU. A base de dados final gerada por esse método possui 7998 registros, representando assim uma redução significativa em relação aos aproximadamente 47 mil registros da

base de dados completa.

Figura 3.34 – *Quadras mais próximas às quadras em ZEUs*



Fonte: Elaboração Própria

Além de possibilitar um ajuste mais ágil do modelo – note-se que o modelo com a base de dados completa demorou mais de 5 dias para convergir – essa estratégia de amostragem possui uma motivação metodológica. Pois ela tem duas principais vantagens:

- Produz uma base de dados balanceada, de modo que 50% dos registros receberam o tratamento e 50% não;
- Aumenta a capacidade de inferência causal do modelo;

Esse último ponto é explorado em pesquisa econométrica que utilizam o método de regressões com base em descontinuidades espaciais, *Spa-*

tial Regression Discontinuity Design (RDD). Esse método opera a partir do pressuposto que elementos situados próximos aos limites de uma fronteira arbitrária são similares. Assim, segundo o método, caso o tratamento seja aplicado apenas aos elementos de um dos lados da fronteira, é possível inferir o efeito causal do tratamento a partir da comparação entre os elementos situados no lado da fronteira que recebeu o tratamento e aqueles também situados no limite, mas que não receberam o tratamento. A inferência se baseia no fato de que esses elementos tendem a ser similares, de forma que a única diferença entre eles consistiria na aplicação do tratamento (IMBENS; LEMIEUX, 2008).

No contexto do problema de pesquisa desta dissertação, essa fronteira arbitrária foi descrita na seção 3.2.2. Trata-se dos *buffers* aplicados sobre as estações de metrô, trem e corredores de ônibus que foram utilizados para definir os Eixos de Estruturação da Transformação Urbana no PDE e que posteriormente definiram as áreas de ZEU. Note-se que a fronteira é arbitrária pois o valor definido para os *buffers* é idêntico em todo o território da cidade, não levando em conta qualquer modulação baseada em condições geográficas ou fatores sociodemográficos e de uso e ocupação do solo. Como esses *buffers* são bastante reduzidos, todas as quadras situadas em ZEU já se encontram na região limítrofe dessa fronteira arbitrária. A metodologia de amostragem aqui descrita permitiu selecionar as demais quadras que se encontram do outro lado desse limite.

A equação a seguir descreve a função de verossimilhança do modelo multinível desenvolvido. Como se pode notar, ele consiste em uma expansão da Regressão Binomial Negativa para Zeros Inflados desenvolvida na seção 3.3.4, inserindo-a em uma estrutura hierárquica que tem as quadras como primeiro nível e as regiões da cidade (área nobre, periférica e de classe média) onde essas quadras estão situadas como segundo nível.

$$\begin{aligned}
& \text{, se } y_{ij} = 0, \\
& P(y_{ij} = 0 \mid X_{ij}, M_{ij}) = \text{Bernoulli}(0 \mid \tau_j + M_{ij}\vartheta_j) \\
& \quad + \text{Bernoulli}(1 \mid \tau_j + M_{ij}\vartheta_j) \times \text{NegBinom}(\alpha_j + X_{ij}\beta_j, \phi) \\
& \text{, se } y_{ij} > 0, \\
& P(y_{ij} = y_{ij} \mid y_{ij} > 0, X_{ij}, M_{ij}) = \text{Bernoulli}(1 \mid \tau_j + M_{ij}\vartheta_j) \\
& \quad \times \text{NegBinom}(\alpha_j + X_{ij}\beta_j, \phi)
\end{aligned} \tag{3.38}$$

Nessa equação:

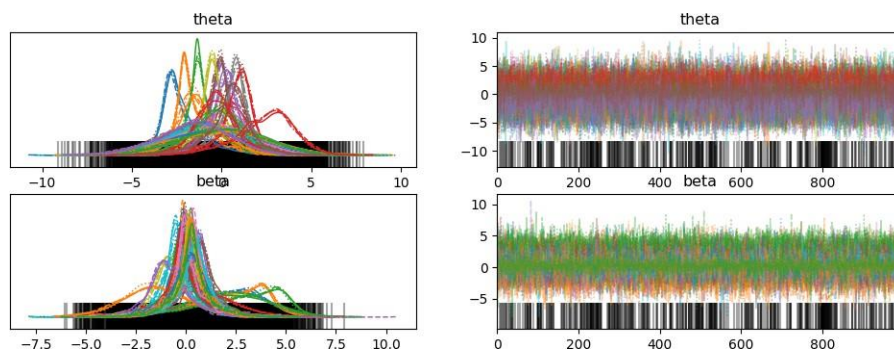
- N representa a quantidade de registros;
- J é o vetor de índices de grupo;
- $\alpha \in \mathbf{R}^J$ é o vetor de J interceptos, um para cada grupo, que integra o preditor linear da Regressão Binomial Negativa;
- $\beta \in \mathbf{R}^{K \times J}$ é a matriz de coeficientes angulares para as K variáveis explicativas que compõem a Regressão Binomial Negativa e indexada para J grupos;
- ϕ é o parâmetro de dispersão da distribuição Binomial Negativa;
- $X \in \mathbf{R}^{N \times K}$ é a matriz de dados para as variáveis explicativas que compõem o preditor linear da Regressão Binomial Negativa;
- $\tau \in \mathbf{R}^J$ é o vetor de J interceptos, um para cada grupo, que integra o preditor linear da Regressão Logística;
- $\vartheta \in \mathbf{R}^{H \times J}$ é a matriz de coeficientes angulares para as H variáveis explicativas que compõem a Regressão Logística e indexada para J grupos;
- $M \in \mathbf{R}^{N \times H}$ é a matriz de dados para as variáveis explicativas que compõem o preditor linear da Regressão Logística;

Foram mantidas as mesmas distribuições *a priori* para os coeficientes da Regressão Binomial Negativa para Zeros Inflados descrita na seção 3.3.4,

com a diferença de que as distribuições dos coeficientes angulares ϑ e β passaram a ser Normais Multivariadas, o que permite modelar as covariações entre os coeficientes internas a cada grupo. Para aumentar a probabilidade de convergência, o modelo foi especificado com o uso da técnicas de parametrização não-centrada e da Decomposição de Cholesky descritas na seção 3.2.7.5. As distribuições *a priori* das matrizes de covariância das Normais Multivariadas foram especificadas pela distribuição *LKJ*(2).

Apesar das técnicas empregadas, o amostrador ainda assim resultou em algumas transições divergentes. No entanto, as estatísticas \hat{R} foram satisfatórias para todos os parâmetros – o \hat{R} mais distante de 1 equivale a 1,005 –, indicando que apesar das transições divergentes as Cadeias de Markov convergiram com sucesso. A Figura 3.35 apresenta o *traceplot* dos coeficientes β e ϑ (omitimos os demais coeficientes pois a quantidade elevada de parâmetros impediria a leitura do gráfico).

Figura 3.35 – *Traceplot - Modelo Multinível*



Fonte: Elaboração Própria

A apresentação dos resultados está organizada em tabelas que apresentam os coeficientes para cada grupo do modelo e para cada processo gerativo modelado. Há assim, por exemplo, uma tabela para os coeficientes que modelam o processo de contagem na área periférica e outra para os mesmos coeficientes na área média. Ao todo, são 6 tabelas. Devido à quantidade elevada de coeficientes, eles não serão discutidos em sua completude. O foco da análise residirá nos coeficientes diretamente relacionados ao problema de pesquisa desta dissertação, que tem por objetivo mensurar a capacidade do planejamento urbano em direcionar a produção imobiliária

na cidade de São Paulo no período 2014-2023.

A análise dos resultados será iniciada com a área de classe média.

Tabela 3.11 – *Processo de contagem - Área Classe Média*

Coefficiente	Média	Desvio-padrão	HDI 5%	HDI 95%
α	3.917	0.094	3.759	4.072
valor_m2_terreno	0.623	0.1353	0.402	0.848
idade_imoveis	0.252	0.112	0.067	0.433
idade_contrib	-0.195	0.073	-0.315	-0.072
maior_terreno	0.391	0.091	0.243	0.546
area_horizont	-0.548	0.0969	-0.707	-0.387
area_vertical	-0.304	0.134	-0.523	-0.076
area_galpao	-0.043	0.064	-0.147	0.063
media_pavimentos	0.183	0.122	-0.018	0.3912
qd_area	0.191	0.091	0.070	0.361
zeu	0.264	0.111	0.084	0.451

Fonte: Elaboração Própria

Tabela 3.12 – *Processo de zeros estruturais - Área Classe Média*

Coefficiente	Média	Desvio-padrão	HDI 5%	HDI 95%
τ	-2.077100	0.051299	-2.161210	-1.992490
envolt_tomb	0.007099	0.257871	-0.430848	0.413472
tombo	0.864601	0.263815	0.423201	1.288980
bairro_ambiental	-2.444780	3.461570	-7.312770	0.923220
area_contaminada	1.161640	0.304743	0.648463	1.645050
protecao_ambiental	-8.400750	88.512100	-17.769700	2.292750
protecao_mananciais	-6.476880	39.788200	-17.916600	1.132870
restricao_geotecnica	-0.278742	1.118380	-2.244160	1.404780
risco_geologico	0.947718	65.259800	-10.448400	10.137100
rppn	-1.047860	72.850100	-10.501900	9.306250
unidade_conserv	-0.221598	20.940900	-8.818830	8.546150
rampa_heliponto	-2.462220	2.646130	-7.170210	0.837358

Fonte: Elaboração Própria

Os resultados para essa região da cidade confirmam alguns dos pontos principais levantados pelos modelos desenvolvidos anteriormente com destaque para os efeitos positivos do valor do metro quadrado de terreno,

que está associado à lucratividade dos empreendimentos, e da idade dos imóveis, que indica o processo de renovação do estoque de edifícios. Interessantemente, a área de galpões não tem efeito significativo, o que pode estar associado ao fato de que os antigos bairros industriais que estão situados nessa região ainda não se valorizaram o suficiente para que ocorra a mudança de uso do solo.

Em relação ao processo de zeros estruturais, é importante notar que, devido à técnica de amostragem utilizada, a depender da área da cidade considerada, pode haver uma quantidade de registros muito reduzida afetados por uma determinada restrição. Isso explica o número elevado de variáveis que não são significativas. Assim, por exemplo, o tombamento dos bairros ambientais e as áreas de proteção aos mananciais, que estão concentrados respectivamente nas áreas nobre e periférica da cidade, praticamente não incidem sobre as quadras próximas aos Eixos de Estruturação da Transformação Urbana na área de classe média. O mesmo ocorre para as variáveis relacionadas à proteção ambiental. Isso pois os EETU são definidos pelos terminais de transporte coletivo, que se localizam nas áreas mais centrais dos bairros, nas quais há pouco ou nenhum remanescente de vegetação nativa protegido.

Quanto à variável de tratamento, nota-se que há um efeito positivo que corresponde a um aumento substancial de 30% na quantidade esperada de unidades de tipo vertical produzidas na quadra. Esse efeito, no entanto, deve ser considerado apenas em comparação às quadras localizadas na área de classe média que estão próximas aos Eixos de Estruturação e Transformação Urbana mas não estão localizadas em ZEU. O resultado apresenta assim evidências que o planejamento urbano logrou sucesso em direcionar a produção imobiliária entre as quadras situadas próximas aos EETU na área de classe média – e não em relação à todas as quadras desta região.

As tabelas a seguir apresentam os resultados para a área periférica.

Em relação ao processo de zeros estruturais, nota-se a mesma questão ocorrida para a área de classe média. Pois mesmo as áreas de proteção ambiental situadas no Extremo Sul estão distantes dos EETU. No entanto, é possível comparar os interceptos τ para ambas as áreas da cidade. Esse parâmetro indica a taxa basal, isto é, a chance média de que uma dada quadra receba um empreendimento imobiliário em cada região. Como se pode no-

Tabela 3.13 – Processo de contagem - Área Periférica

Coeficiente	Média	Desvio-padrão	HDI 5%	HDI 95%
α	3.590	0.77	2.325	4.865
valor_m2_terreno	0.557	1.664	-2.119	3.336
idade_imoveis	0.155	0.357	-0.415	0.754
idade_contrib	-0.573	0.414	-1.268	0.103
maior_terreno	0.774	0.258	0.368	1.211
area_horizont	-0.867	0.429	-1.568	-0.872
area_vertical	-1.954	1.368	-4.23	0.362
area_galpao	-1.031	0.505	-1.802	-0.132
media_pavimentos	2.619	2.008	-0.805	5.839
qd_area	0.185	0.243	-0.212	0.597
zeu	0.237	0.384	-0.396	0.887

Fonte: Elaboração Própria

Tabela 3.14 – Processo de zeros estruturais - Área Periférica

Coeficiente	Média	Desvio-padrão	HDI 5%	HDI 95%
τ	-2.864060	0.157472	-3.132050	-2.608680
envol_tomb	0.250800	0.986753	-1.517140	1.764520
tombo	0.066715	1.348200	-2.235700	2.156360
bairro_ambiental	0.074543	4.631190	-5.596540	6.017070
area_contaminada	-1.424940	2.644910	-5.772470	2.033960
protecao_ambiental	-9.964700	86.315600	-18.126200	0.462844
protecao_mananciais	2.171250	188.359000	-9.454920	9.418310
restricao_geotecnica	0.104250	3.417530	-5.026740	5.320130
risco_geologico	-3.419700	211.647000	-9.657160	10.208200
rppn	0.326523	32.847600	-11.093500	10.568700
unidade_conserv	-3.344230	32.174900	-10.924800	4.740750
rampa_heliponto	3.432000	1.021770	1.991590	5.056680

Fonte: Elaboração Própria

tar, essa taxa basal é bastante reduzida na área periférica, correspondendo a apenas 6%, contra 12% para a área de classe média.

É interessante notar também que na área periférica diversas variáveis relevantes para a produção imobiliária não têm efeito significativo. Esse é o caso do valor do metro quadrado de terreno e da idade das edificações, o que indica que a dinâmica de verticalização nessa região é regida por outros

fatores. Note-se que a proporção do maior terreno não verticalizado tem efeito positivo, o que fornece evidências para a hipótese aventada na seção 3.2.7.5 de que nessa região a verticalização ocorre sobretudo por meio de empreendimentos de habitação social que são prioritariamente alocados em grandes terrenos não edificadas (pois isso reduz o custo de desapropriação).

Quanto à variável de tratamento, nota-se que o fato de uma dada quadra localizada na área periférica da cidade estar situada em ZEU não possui efeito significativo sobre a quantidade esperada de unidades de tipo vertical produzidas no período 2014-2023.

A seguir, tabulamos os resultados para a área nobre.

Tabela 3.15 – Processo de contagem - Área Nobre

Coefficiente	Média	Desvio-padrão	HDI 5%	HDI 95%
α	4.723140	0.153235	4.478720	4.985220
valor_m2_terreno	-0.085428	0.050821	-0.165981	-0.001100
idade_imoveis	0.401180	0.142474	0.166775	0.634270
idade_contrib	-0.146712	0.087491	-0.293900	-0.004811
maior_terreno	0.106749	0.138299	-0.115080	0.339099
area_horizont	-0.317226	0.115692	-0.506859	-0.132361
area_vertical	-0.249453	0.091630	-0.401723	-0.098584
area_galpaoc	-0.067535	0.088894	-0.210030	0.081666
media_pavimentos	0.037833	0.104629	-0.135282	0.215035
qd_area	0.481167	0.226619	0.116496	0.868369
ZEU	0.252421	0.123683	0.048355	0.457489

Fonte: Elaboração Própria

Quanto às restrições à edificação, note-se em primeiro lugar que a taxa basal τ possui um valor elevado. Como mencionado, essa taxa representa a chance média de que uma dada quadra localizada em uma dada área, próxima ao EETU, receba um empreendimento imobiliário. Na área nobre, essa chance média equivale a 25%, valor substancialmente maior que as áreas média (12%) e periférica (6%). Nota-se que restrições relacionadas à proteção arquitetônica e do traçado urbano têm efeito significativo. Já as restrições ambientais não possuem efeito, provavelmente devido à distância dos EETU em relação às áreas de preservação ambiental.

Em relação ao processo de contagem, note-se que, surpreendentemente, o valor do metro quadrado de terreno contribui negativamente para a quan-

Tabela 3.16 – *Processo de zeros estruturais - Área Nobre*

Coeficiente	Média	Desvio-padrão	HDI 5%	HDI 95%
τ	-1.360670	0.061245	-1.459930	-1.259140
envolt_tombamento	-0.107787	0.195505	-0.435555	0.205869
tombo	-0.544521	0.239305	-0.950422	-0.161533
bairro_ambiental	-1.697650	0.420286	-2.435590	-1.051490
area_contaminada	0.255800	0.375615	-0.384651	0.860674
protecao_ambiental	1.264680	57.674900	-9.421060	9.036530
protecao_mananciais	-0.548165	112.672000	-9.982280	9.598670
restricao_geotecnica	-0.384864	0.638976	-1.472540	0.591873
risco_geologico	1.472220	72.340700	-9.695940	10.124100
rppn	0.708209	53.033700	-9.239630	10.141800
unidade_conserv&0.399146	32.741000	-7.607380	7.808670	
rampa_heliponto	0.680246	0.322614	0.130323	1.196570

Fonte: Elaboração Própria

tidade esperada de unidades verticais produzidas na quadra no período. O efeito não possui grande magnitude, no entanto, correspondendo a uma redução de 7% na quantidade esperada de unidades. É possível que esse resultado também seja decorrente da seleção apenas das quadras próximas aos [EETU](#), o que poderá ser verificado pelo modelo ajustado para os dados completos.

Quanto à variável de tratamento, assim como no caso da área de classe média, há efeito positivo substancial. O fato de uma dada quadra localizada na área nobre estar situada em [ZEU](#) aumenta em 28% a quantidade esperada de unidades de tipo verticais produzida no período 2014-2023, em comparação com uma quadra localizada na mesma região, situada próximo ao [EETU](#) e com as demais características similares.

A partir da discussão dos resultados apresentados, conclui-se que, apesar de que a amostragem seguindo o método da descontinuidade espacial (*Spatial RDD*) ter sido descrita na literatura econométrica como uma forma mais efetiva para mensurar os efeitos causais de tratamentos que ocorrem no espaço, no que diz respeito ao problema de pesquisa deste estudo essa metodologia possui limitações. Ao selecionar apenas as quadras localizadas próximas aos Eixos de Estruturação da Transformação Urbana, a metodologia de amostragem resulta em um enviesamento. Pois, uma vez que os

limites dos EETU são definidos pela proximidade com as estações de trem, metrô e corredores de ônibus, e que a localização destes elementos não é aleatória pela cidade, decorre-se que eles estão situados em áreas com características de uso e ocupação do solo específicas: trata-se, em quase todos os casos, de áreas de centralidade.

Disso resulta que a dinâmica da produção imobiliária que está sendo avaliada por meio do modelo ajustado a esses dados é específica a essas áreas de centralidade. Dessa forma, não é possível extrapolar com segurança os resultados do modelo para a cidade como um todo. Para contornar essa questão, foi desenvolvido um modelo multinível com a mesma especificação do modelo aqui descrito, mas ajustado para a totalidade das mais de 46 mil quadras da cidade.

O ajuste do modelo para a base de dados completa teve performance superior ao modelo ajustado para o subconjunto dos dados. Houveram menos transições divergentes do amostrador (6% contra 11,8% do modelo anterior), o que indica que ele foi capaz de explorar o espaço amostral com maior efetividade. As estatísticas \hat{R} foram todas satisfatórias, sendo que o valor 1,005 é o que mais se distanciou de 1. A seguir, são apresentados os resultados desse modelo, novamente organizados por cada uma das regiões da cidade.

As Tabelas 3.17 e 3.18 contém os resultados do modelo para a área de classe média da cidade.

De início, cumpre ressaltar que diversas variáveis de controle que não foram significativas ao intervalo de confiabilidade de 90% no ajuste do modelo para as quadras próximas aos EETU passaram a ser significativas uma vez que a totalidade das quadras da cidade passou a ser considerada. Isso se deu sobretudo em relação aos coeficientes da Regressão Logística que modela o processo de zeros estruturais.

Assim, a proteção aos mananciais voltou a ter um efeito negativo substantivo, praticamente anulando a chance de que uma quadra localizada na área de classe média da cidade e que está em área de proteção aos mananciais receba um empreendimento imobiliário formal. O zoneamento exclusivamente residencial também passou a ter efeito negativo significativo. Esses efeitos não puderam ser captados pelo modelo anterior por conta do viés amostral para quadras em áreas de centralidade. Uma vez que todas

Tabela 3.17 – *Processo de contagem - Área Classe Média*

Coeficiente	Média	Desvio-padrão	HDI 5%	HDI 95%
α	3.895090	0.042824	3.825830	3.965930
valor_m2_terreno	0.433525	0.082285	0.301166	0.571377
idade_imoveis	0.263843	0.055184	0.172848	0.357675
idade_contrib2014	-0.135654	0.042399	-0.206009	-0.067274
qtd_pavimentos	0.255504	0.072674	0.140714	0.377008
maior_terreno	0.413269	0.050737	0.331345	0.499710
area_horizontal	-0.679737	0.052108	-0.766421	-0.594551
area_vertical	-0.305248	0.080261	-0.438372	-0.171567
qd_area	0.252667	0.055441	0.166145	0.350325
area_galpoes	0.038466	0.030848	-0.011375	0.088374
zeu	0.329106	0.087187	0.188696	0.476980

Fonte: Elaboração Própria

Tabela 3.18 – *Processo de zeros estruturais - Área Classe Média*

Coeficiente	Média	Desvio-padrão	HDI 5%	HDI 95%
τ	-2.567300	0.027995	-2.613560	-2.521780
envolt_tombamento	0.061192	0.125649	-0.148534	0.270079
tombo	0.534867	0.150563	0.282907	0.779796
bairro_ambiental	-1.636320	0.749486	-2.992170	-0.540220
area_contaminada	1.215000	0.140825	0.981283	1.442260
protecao_ambiental	-0.590289	1.082180	-2.543780	0.958247
protecao_mananciais	-25.623600	104.865000	-84.033900	-2.782990
restricao_geotec	0.336777	1.168970	-1.736740	2.046420
risco_geologico	0.469712	0.678571	-0.727815	1.507780
rppn	-6.619420	112.677000	-10.249600	2.623010
unidade_conservacao	-0.969556	50.064700	-9.070050	8.226750
rampa_heliponto	-0.120223	0.480641	-0.953841	0.612886
zer	-1.442640	0.348438	-2.052780	-0.895309

Fonte: Elaboração Própria

as quadras da cidade estão sendo consideradas, o modelo pôde ser treinado com os dados de quadras situadas em áreas residenciais e próximas aos mananciais, entre outras características às quais não havia sido exposto adequadamente.

A área de galpões permaneceu com um efeito não significativo, o que

fornece evidências para corroborar a hipótese de que nas áreas de classe média da cidade há ainda antigos bairros industriais que não se valorizaram o suficiente para que sejam verticalizados. O maior terreno não edificado no entanto passou a ter um efeito positivo significativo, o que mostra que a facilidade de incorporação e a redução dos custos de demolição são fatores relevantes para a verticalização nessa região da cidade. A idade dos imóveis e o valor do metro quadrado de terreno também mantiveram seus efeitos positivos.

Em relação à variável de tratamento, que indica se a quadra está situada em [ZEU](#), o efeito positivo significativo se manteve, tendo sido inclusive ampliado. Assim, tudo o mais mantido igual, o fato de que uma quadra localizada na área de classe média da cidade está situada em Zona de Estruturação Urbana acarreta um aumento de 39% na quantidade esperada de unidades de tipo vertical produzidas no período 2013-2024 em relação às demais quadras localizadas na mesma região.

A seguir, são analisados os resultados do modelo para a área periférica da cidade, que constam nas Tabelas [3.19](#) e [3.20](#).

Tabela 3.19 – Processo de contagem - Área Periférica

Coeficiente	Média	Desvio-padrão	HDI 5%	HDI 95%
α	3.165510	0.288923	2.700990	3.646530
idade_contrib	-0.393957	0.099105	-0.561330	-0.232205
maior_terreno	0.376126	0.099189	0.218343	0.543906
area_horizontal	-1.023860	0.166865	-1.294620	-0.750699
area_vertical	-1.002190	0.555496	-1.903010	-0.077022
qd_area	0.302933	0.145525	0.080747	0.561665
valor_m2_terreno	-0.059171	0.574854	-1.004310	0.822348
idade_imoveis	0.040972	0.088791	-0.098598	0.190598
qtd_pavimentos	0.467170	0.368752	-0.114262	1.095470
area_total_galpoes	0.094390	0.148916	-0.139460	0.351237
zeu	-0.042279	0.301577	-0.528713	0.483644

Fonte: Elaboração Própria

Dos resultados obtidos para a área periférica, destaca-se o fato de que ambos os interceptos, α e τ possuem valores inferiores aos amostrados para a área de classe média. Isso indica que a quantidade média esperada de unidades de tipo vertical produzidas em uma quadra situada nessa região

Tabela 3.20 – *Processo de zeros estruturais - Área Periférica*

Coeficiente	Média	Desvio-padrão	HDI 5%	HDI 95%
τ	-3.415760	0.057077	-3.509520	-3.321330
envoltoria_tomb	0.778084	0.564083	-0.191524	1.635750
tombo	-0.767393	1.062640	-2.620870	0.777263
bairro_ambiental	0.007291	3.546290	-5.061240	5.034790
area_contaminada	1.684570	0.512778	0.776412	2.465720
protecao_ambiental	-1.433590	0.620561	-2.530400	-0.518543
protecao_mananciais	-29.111700	125.109000	-86.991400	-1.858630
restricao_geotec	0.048082	3.376400	-5.326220	5.120310
risco_geologico	0.260258	4.895320	-5.552180	6.187240
rppn	-3.804570	57.217900	-9.766960	4.572870
unidade_conservacao	-6.848630	41.421900	-18.003900	1.451480
rampa_heliponto	3.505480	0.744699	2.327760	4.734850
ZER	-3.191000	2.583950	-7.582430	-0.331528

Fonte: Elaboração Própria

da cidade, assim como a chance de que essa quadra esteja apta a receber um empreendimento imobiliário são menores quando comparadas às áreas de classe média (e também à área nobre).

Também para esta região da cidade houve aumento da quantidade de variáveis de controle com efeitos significativos. Destaca-se, novamente, o forte efeito negativo da localização em área de proteção aos mananciais. O valor médio do metro quadrado de terreno permaneceu, no entanto, não significativo, o que tende a confirmar a hipótese aventada de que este não é o principal fator que rege a dinâmica de verticalização nas áreas periféricas.

O efeito positivo do maior terreno não verticalizado também foi mantido.

Em relação à variável de tratamento, verifica-se que ela permanece sem efeito significativo. Dessa forma, o fato de que uma quadra localizada na área periférica da cidade está localizada em [ZEU](#) não está associado à quantidade esperada de unidades de tipo vertical produzidas nessa quadra durante o período 2014-2023.

Os resultados do modelo para a área nobre da cidade são apresentados nas Tabelas [3.21](#) e [3.22](#) que seguem.

Os interceptos α e τ indicam que a propensão à verticalização é significativamente superior na área nobre da cidade. Em relação ao processo

Tabela 3.161 – *Processo de 161ontagem - Área*

Coeficiente	Média	Desvio-padrão	HDI 5%	HDI 95%
α 4.687380	0.086446	4.543830	4.828280	
valor_m2_terreno	-0.117948	0.040447	-0.183737	-0.049899
idade_imoveis	0.165886	0.061148	0.068512	0.267202
idade_contribuintes	-0.104695	0.060220	-0.203975	-0.004592
area_horizontal	-0.231568	0.078702	-0.361024	-0.102327
area_vertical	-0.168115	0.067885	-0.283322	-0.060572
qtd_area	0.449436	0.144185	0.225465	0.701037
qtd_pavimentos	0.011350	0.061585	-0.088436	0.113012
maior_terreno	0.109132	0.088171	-0.035727	0.256248
area_galpoes	0.045879	0.055229	-0.041641	0.137996
zeu	0.390372	0.111038	0.207174	0.573436

Fonte: Elaboração Própria

Tabela 3.22 – *Processo de zeros estruturais - Área Nobre*

Coeficiente	Média	Desvio-padrão	HDI 5%	HDI 95%
τ	-1.745300	0.045477	-1.820060	-1.670240
envoltoria_tomb	-0.031491	0.150523	-0.278793	0.209069
tombo	-0.358253	0.218665	-0.720115	-0.006874
bairro_ambiental	-1.615120	0.416421	-2.329680	-0.964382
area_contaminada	0.707514	0.243365	0.300405	1.100340
protecao_ambiental	-0.014568	3.281700	-4.997490	4.833680
protecao_mananciais	2.613520	125.694000	-35.191900	35.842200
restricao_geotec	-0.246034	0.411584	-0.941386	0.382780
risco_geologico	0.024432	4.559030	-5.566120	5.633000
rppn	-2.170540	99.160900	-6.802270	7.130820
unidade_conservacao	-2.185130	70.877300	-9.039660	8.045060
rampa_heliponto	0.694177	0.137061	0.466890	0.920710
ZER	-2.183170	0.209093	-2.529180	-1.845970

Fonte: Elaboração Própria

de contagem, o coeficiente de α equivale a um valor médio esperado de mais de 100 unidades produzidas por quadra no período. Já em relação ao processo de zeros estruturais, o coeficiente de τ corresponde a uma chance média de 17% de que uma quadra localizada na área nobre receba um empreendimento imobiliário vertical. Note-se esse último coeficiente, apesar de

ainda bastante elevado em relação às demais áreas da cidade (que equivale a 7% para a área de classe média e apenas 3% para a área periférica) foi reduzido substancialmente em relação à estimativa do modelo anterior.

Mais uma vez, essa diferença provavelmente está relacionado ao viés de seleção da amostragem realizada anteriormente, pois é possível que as quadras localizadas na área nobre que estão em áreas de centralidade possuam menos restrições à edificação. Ademais, em relação às restrições à edificação na área nobre, ressalta-se o efeito negativo substancial da localização em Zoneamento Exclusivamente Residencial. Tudo o mais mantido igual, uma quadra localizada na área nobre que está situada em ZER tem uma redução de 88% na probabilidade de que ela receba um empreendimento vertical em relação a outra quadra localizada na mesma região. O tombamento dos bairros ambientais (conhecidos como *bairros jardim*) também acarreta uma redução significativa, equivalente a uma queda de 80% na chance de receber um empreendimento.

Por fim, em relação à variável de tratamento, constata-se que ela possui efeito positivo significativo. Assim, uma quadra localizada na área nobre que está situada em ZEU tem um aumento de 47% na quantidade esperada de unidades de tipo vertical produzidas na quadra no período 2014-2023 em comparação a uma quadra de características similares localizada na mesma região.

3.4 CONSIDERAÇÕES FINAIS

Os resultados da análise de agrupamentos espaciais realizada na seção 3.3.1 confirmam em larga medida o argumento de Villaça (2001) sobre a estrutura espacial de segregação social que vigora na cidade de São Paulo. A área nobre, região com a maior proporção de pessoas brancas, ricas e alfabetizadas e que possui a maior acessibilidade ao mercado de trabalho está situada no quadrante sudoeste, como previu Villaça (2001) e corroborando também os achados de outros autores que se voltaram sobre o tema da segregação, em espacial Marques (2014).

Os resultados da modelagem estatística realizada nesse capítulo demonstram que o planejamento urbano deve levar em conta essa estrutura de segregação. Assim, em um primeiro momento os resultados dos modelos

ajustados nas seções 3.3.2 e 3.3.3 apontaram para uma forte capacidade do planejamento urbano em direcionar a produção imobiliária na cidade para as Zonas de Estruturação da Transformação Urbana.

No entanto, os resultados da inclusão, no modelo desenvolvido na seção 3.3.3, das interações entre o tratamento e a localização da quadra em uma das três regiões que conformam a estrutura de segregação urbana na cidade alteraram esse quadro. Pois eles mostraram que os incentivos construtivos fornecidos pelo Plano Diretor nas ZEU possuem efeitos díspares nas áreas nobre, de classe média e periféricas da cidade. O modelo desenvolvido na seção 3.3.5 aprofunda esse entendimento, mostrando como as dinâmicas da produção imobiliária ocorrem de formas distintas em cada uma dessas áreas.

Os resultados deste último modelo fornecem evidências de que o direcionamento da produção imobiliária na cidade para as áreas de ZEU efetivou-se apenas nas áreas nobre e de classe média da cidade, não havendo efeito significativo nas áreas periféricas. Ademais, destaca-se a elevada propensão das áreas nobres a receber empreendimentos imobiliários, em muito superior às demais regiões da cidade. Esse resultado assinala que há fatores, não devidamente considerados pelo planejamento urbano, que estão determinando os padrões de adensamento na cidade. No capítulo 4, que segue, analisaremos um desses fatores: a acessibilidade ao mercado de trabalho.

Por fim, destaca-se efeitos negativos substanciais do tombamento do traçado urbano dos *bairros jardim*, assim como do zoneamento exclusivamente residencial unifamiliar (*single detached housing*) sobre possibilidade de verticalização na área nobre da cidade. Essas restrições, que vigoram sobre os bairros de residência da elite da cidade, têm por principal efeito reforçar ainda mais a segregação social nessa região.

Os dados indicam, assim, que o levantamento dessas restrições à verticalização na área nobre da cidade poderia ter efeitos muito mais substanciais sobre o adensamento urbano do que aquele promovido pelos Eixos de Estruturação da Transformação Urbana. Uma vez que esses bairros possuem acesso privilegiado aos bens, serviços e infraestrutura urbana, caso o planejamento urbano apenas permitisse que se efetivasse a verticalização que atualmente ele impede, os objetivos do Plano Diretor relacionados à redu-

ção dos tempos de deslocamento e promoção do direito à cidade seriam alcançados com maior efetividade.

4 RETORNO AO MODELO MONOCÊNTRICO

Resumo do capítulo

Este capítulo examina a relação entre adensamento construtivo de tipo residencial vertical e acessibilidade ao mercado de trabalho na cidade de São Paulo (2007-2017). Por meio de um modelo estatístico com alta significância e capacidade explicativa ($R^2 = 0.57$), revela forte associação positiva entre acessibilidade e adensamento. Esse resultado contribui para a literatura empírica sobre estrutura urbana ao apresentar evidências que corroboram a principal contribuição teórica do modelo urbano padrão, formulado na década de 1960. Isso permite avaliar criticamente propostas recentes do urbanismo que retomam o tema da relação entre acessibilidade e estrutura urbana, mas que negligenciam dinâmicas socioespaciais mais amplas, o que eleva os riscos de efeitos adversos, como a segregação e a gentrificação.

Objetivo: Analisar a relação entre acessibilidade ao mercado de trabalho e adensamento urbano.

4.1 INTRODUÇÃO

Como apresentado no Capítulo 2, diversos movimentos do urbanismo nas últimas décadas se voltaram sobre os temas da densidade e da proximidade entre os locais de moradia e trabalho como formas de combater os efeitos adversos do espraiamento da mancha urbana. O Plano Diretor Estratégico da Cidade de São Paulo (PDE), inspirado em um desses movimentos (BALBIM; KRAUSE, 2016), estabeleceu o objetivo de promover o adensamento nas áreas próximas aos terminais de transporte público da cidade, denominadas Eixos de Transformação da Estruturação Urbana.

Como exposto no Capítulo 3, essa medida surtiu efeitos apenas nas áreas da cidade para as quais já se voltava o interesse do mercado imobiliário, sem lograr efetividade nas áreas periféricas. Dessa forma, o PDE não teve êxito em alterar as estruturas de segregação espacial vigentes na cidade. Esse cenário aponta para a necessidade de que o planejamento urbano em São Paulo mobilize um arcabouço teórico definido que possibilite avaliar empiricamente a capacidade das medidas por ele propostas em

de fato promover a transformação urbana e não apenas reforçar dinâmicas pré-existentes. Como argumentam (DESALVO; SU, 2018), o Modelo Monocêntrico ([SMM](#)), cujas bases teóricas foram descritas no Capítulo 2 apresenta-se como uma alternativa para abordar esse problema.

O presente capítulo insere-se nesse debate ao propor soluções metodológicas para a verificação empírica do [SMM](#). Ele apresenta uma metodologia de análise que permite contornar o pressuposto de um único distrito comercial central, simplificação que concentrou o debate na literatura empírica em torno do [SMM](#) sobre a questão de se as cidades são monocêntricas ou não. Como argumenta Duranton e Puga (2015, p. 15), essa discussão passa ao largo da principal contribuição teórica do modelo, que é a relação positiva entre acessibilidade, valor da terra e densidade.

O capítulo aborda essa contribuição teórica empiricamente, apresentando um estudo de caso sobre a relação entre adensamento construtivo e acessibilidade na cidade de São Paulo no decênio 2007-2017. Nele, é testada a hipótese de que um aumento na acessibilidade ao mercado de trabalho em um dado distrito da cidade nesse período esteve associado a um aumento na área construída de tipo residencial vertical, levando em consideração limitações regulatórias assim como a saturação por edificações pré-existentes. O modelo estatístico desenvolvido, uma regressão linear múltipla, reporta forte associação positiva entre acessibilidade e adensamento construtivo, com alta significância estatística ($p < 0.001$) e capacidade explicativa, alcançando um coeficiente de determinação de 57% — resultados que fornecem evidências para corroborar o [SMM](#). Com base nisso, o artigo conclui, em consonância com Bertaud (2022b), DeSalvo e Su (2018) e Gray (2022), sobre a necessidade do planejamento urbano incorporar o conhecimento sobre as dinâmicas socioeconômicas descritas pelo [SMM](#) para evitar efeitos indesejados e alcançar um desenvolvimento urbano integrado.

4.1.1 Objetivos

Este capítulo tem por objetivo analisar a relação entre acessibilidade ao mercado de trabalho e adensamento construtivo de tipo residencial vertical na cidade de São Paulo no período 2007-2017.

4.1.2 Organização

O capítulo está organizado da seguinte forma:

- A seção 4.2, [Materiais e Métodos](#) descreve os procedimentos realizados para gerar a base de dados que foi analisada neste capítulo. Essas informações são detalhadas no Anexo [Único](#);
- A seção 4.3, [Resultados](#) apresenta os resultados da análise exploratória de dados e dos modelos estatísticos desenvolvidos;
- A seção 4.4, [Considerações Finais](#) analisa os resultados apresentados na seção 4.3 à luz da questão da promoção da densidade pelo planejamento urbano;

4.2 MATERIAIS E MÉTODOS

4.2.1 Materiais

Os modelos estatísticos aqui desenvolvidos relacionam a acessibilidade ao mercado de trabalho nos distritos da cidade de São Paulo com o adensamento construtivo de tipo residencial vertical ocorrido neles no período 2007-2017. Para isso, foi preciso construir uma base de dados que identifique, para cada distrito, sua acessibilidade ao mercado de trabalho, assim como a área construída total, em metros quadrados, produzida por empreendimentos residenciais verticais em cada distrito no período.

Para calcular a acessibilidade ao mercado de trabalho, iniciou-se com a identificação da localização dos vínculos empregatícios cadastrados na [RAIS](#). Essa fonte de dados é descrita em maiores detalhes na seção [C.2.3](#) do Anexo [Único](#). A distribuição dos postos de trabalho na cidade nos anos de 2007 e 2017 foi obtida a partir da somatória da quantidade de vínculos ativos dos estabelecimentos situados em São Paulo, agrupada por distrito.

Em seguida, a Pesquisa Origem e Destino ([OD](#)) foi utilizada para analisar os padrões de deslocamento a trabalho da população de São Paulo. A seção [C.2.2](#) do Anexo [Único](#) descreve essa fonte de dados em maiores detalhes. Foram identificadas as viagens pendulares da pesquisa [OD](#) realizadas dentro dos limites do município de São Paulo, o que resultou em 48255 e

42956 viagens para 2007 e 2017, respectivamente. As viagens foram agrupadas de acordo com os pares de distritos de origem e destino, e a média de tempo de deslocamento para cada par foi calculada. O cruzamento de ambas as bases de dados geradas, isto é, a localização dos postos de trabalho por distrito e a média dos tempos de deslocamento a trabalho de um distrito a outro da cidade permitiu calcular, para cada um dos distritos, sua acessibilidade ao mercado de trabalho.

A variável gerada por esse método consiste no percentual dos empregos da cidade que podem ser acessados partindo de cada distrito. Ela foi calculada da seguinte forma: caso o tempo médio de deslocamento a trabalho de um distrito a outro seja inferior a 60 minutos (mediana dos tempos de deslocamento), a quantidade de postos de trabalho do distrito de destino é contabilizada como acessível pelo distrito de origem. Caso não haja nenhuma viagem na base de dados partindo desse distrito a algum outro, estabelece-se que os empregos do distrito de destino não são acessíveis. A acessibilidade ao mercado de trabalho é então obtida por meio da divisão do total de empregos acessíveis por um dado distrito pelo total de empregos da cidade, e é portanto limitada entre 0 e 1. A seção [D.7](#) detalha a metodologia de cálculo.

Essa metodologia permite identificar a acessibilidade ao mercado de trabalho sem a necessidade de pressupor um único distrito comercial central [CBD](#). Ela também prescinde de calcular os custos de deslocamento como a distância linear em relação ao [CBD](#), como na formulação original do modelo urbano padrão. Em seguida, a alteração na acessibilidade ao longo da década foi calculada a partir da subtração da taxa de acessibilidade em 2017 pela taxa em 2007 (que está assim limitada entre -1 e 1).

O adensamento construtivo de tipo residencial vertical na cidade foi calculado por meio da identificação dos imóveis cadastrados na base do [IPTU](#) para os anos de 2007 e 2017 cujo tipo de construção é vertical e cujo uso é residencial. Os imóveis produzidos no período foram identificados como aqueles que constam na base de 2017, mas que não estavam cadastrados na base anterior de 2007 e que foram construídos após esse ano. Eles foram geolocalizados por meio das camadas do cadastro fiscal disponibilizadas no portal GeoSampa, o que possibilitou identificar os distritos em que se situam. A área construída desses imóveis foi então agregada por dis-

trito, resultando na identificação da área total de tipo residencial vertical produzida por empreendimentos formais no período 2007-2017 na cidade de São Paulo, por distrito. A seção [D.1.3](#) do Anexo Único detalha esses procedimentos.

Em seguida, o aumento na densidade foi calculado por meio da divisão dessa área construída pela área total edificável do distrito. A área edificável foi obtida pela somatória das áreas dos polígonos das quadras urbanas por distrito, o que permitiu desconsiderar a área da estrutura viária assim como a área rural, que não é objeto de nosso estudo. Em seguida, com o uso de técnicas de geoprocessamento, foram removidas as áreas dos polígonos de parques e unidades de conservação, áreas de proteção ambiental, remanescentes de mata atlântica, terras indígenas, propriedades públicas, áreas de risco geológico e geotécnico, áreas contaminadas, imóveis tombados assim como zoneamentos restritivos e outras restrições urbanísticas, nomeadamente os bairros ambientais e faixas non aedificandi. Esses dados, ainda que não correspondam a todos os possíveis óbices à edificação representam um avanço significativo no cálculo da área efetivamente verticalizável.

Por fim, com o objetivo de identificar a saturação construtiva devido a edificações cujos custos de demolição inviabilizam a realização de empreendimentos imobiliários, removemos também, para cada distrito, a área total dos terrenos nos quais estão situados edifícios de 15 andares ou mais (número definido pela mediana do número de andares dos prédios da cidade). A metodologia para obtenção da área edificável por distrito é apresentada em mais detalhes na seção [D.5](#) do Anexo Único.

4.2.2 Métodos

Os dados foram então analisados em duas etapas. A primeira etapa, de caráter exploratório, teve por objetivo identificar padrões de concentração espacial na acessibilidade ao mercado de trabalho e no adensamento construtivo na cidade no período de estudo. Para isso, foi empregada a análise LISA com o método Moran Local (ANSELIN, 1995). Esse método não supervisionado foi utilizado no Capítulo 3. Ele permite identificar *clusters* espaciais a partir do cálculo da covariação entre a variável de interesse em determinado registro e sua rede de vizinhança, de acordo com a equação:

$$I_i = (X_i - \bar{X}) \sum_{j=1}^n w_{ij} (X_j - \bar{X}) \quad (4.1)$$

Neste estudo, foi construída uma rede de vizinhança binária, cujas relações de vizinhança foram definidas pela contiguidade entre os polígonos dos distritos (isto é, se eles fazem fronteira uns com os outros). A análise permite identificar os agrupamentos espaciais e caracterizá-los em 4 tipos: Alto-Alto (valores elevados cercados de valores elevados), Baixo-Baixo (valores baixos cercados de valores baixos), Alto-Baixo (valores altos cercados de valores baixos) e Baixo-Alto (o inverso, ou seja, valores baixos cercados de valores altos). Trata-se da mesma metodologia empregada na seção 3.2.5 do Capítulo 3, onde ela é apresentada de forma detalhada.

A segunda etapa, confirmatória, teve por objetivo testar a hipótese do estudo, nomeadamente a associação positiva entre acessibilidade ao mercado de trabalho e adensamento construtivo de tipo residencial vertical. Para isso, foi utilizada a Regressão Linear Múltipla com o uso da função de ligação identidade e erro quadrático médio (MSE) como função de custo, de acordo com a equação clássica:

$$\begin{aligned} y_i &= \alpha + \beta X_i + \epsilon_i \\ \epsilon_i &\sim N(0, \sigma) \end{aligned} \quad (4.2)$$

Neste estudo, a variável dependente da Regressão Linear Múltipla foi definida como o adensamento construtivo no período 2007-2017, enquanto as variáveis tratamento correspondem à acessibilidade ao mercado de trabalho em 2007 e sua variação nesses dez anos. A hipótese testada é a seguinte:

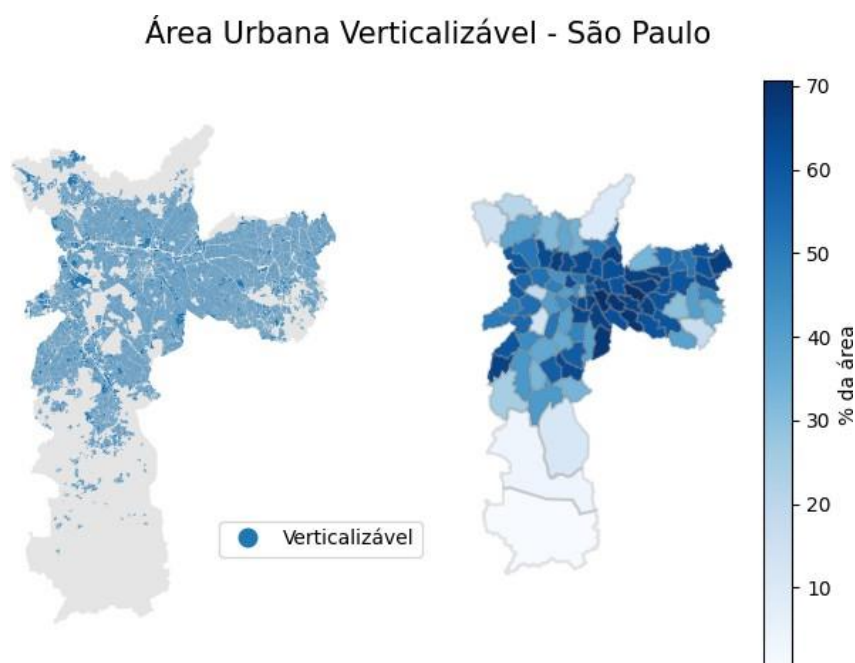
- H_0 : não há associação entre acessibilidade ao mercado de trabalho e adensamento construtivo de tipo residencial vertical;
- H_1 : há associação positiva entre acessibilidade ao mercado de trabalho e adensamento construtivo de tipo residencial vertical;

4.3 RESULTADOS

A Figura 4.1 representa área urbana verticalizável na cidade de São Paulo. O painel à esquerda apresenta, em azul, os polígonos das áreas que

podem ser edificadas. Verifica-se vazios nas zonas leste, norte e sul, nos limites da área urbanizada da cidade. As porções norte e sul correspondem, grosso modo, a áreas de proteção ambiental, enquanto na porção leste trata-se de áreas com grande concentração de loteamentos irregulares, que também estão presentes, em menor proporção, na zona sul. Há também áreas rurais nos extremos sul e leste, sendo esta última já bastante reduzida. Os vazios nas áreas mais centrais, na porção sudoeste da cidade, se devem a zoneamentos exclusivamente residenciais horizontais unifamiliares de baixa densidade (conhecidos na literatura como *single detached housing*).

Figura 4.1 – Área verticalizável em São Paulo.



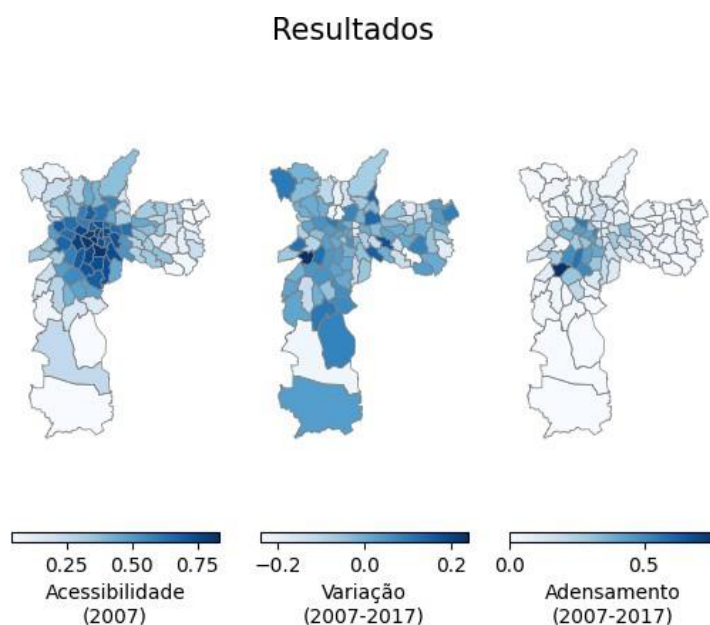
Fonte: Elaboração própria

O painel à direita, que plota o percentual da área total de cada distrito que é verticalizável, permite identificar com maior clareza a anomalia representada por esses vazios centrais. Os dois distritos com menor potencial de verticalização na área central, Alto de Pinheiros e Morumbi, estão entre as áreas mais valorizadas da cidade e provavelmente receberiam muitos empreendimentos imobiliários caso não houvesse esses impedimentos.

A Figura 4.2 plota os resultados das três variáveis cuja construção

foi descrita na seção 4.2. O painel à esquerda apresenta a acessibilidade ao mercado de trabalho por distrito no ano de 2007, o painel central, a variação nessa acessibilidade na década 2007-2017 e o painel à direita plota o adensamento construtivo de tipo residencial vertical ao longo desses dez anos, considerando a área efetivamente verticalizável de cada distrito.

Figura 4.2 – *Distribuição da acessibilidade ao mercado de trabalho em 2007, sua variação no período 2007-2017 e do adensamento construtivo no mesmo período em São Paulo, por distrito*

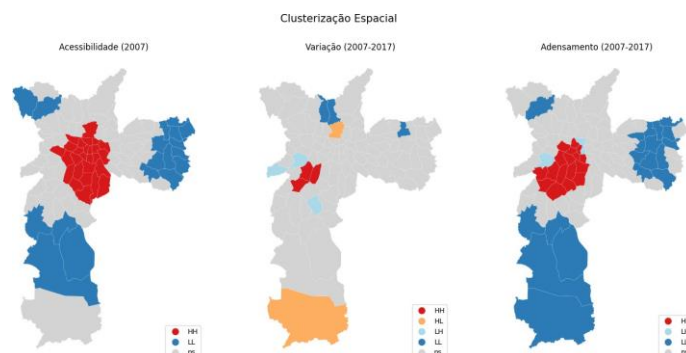


Fonte: Elaboração própria

Enquanto a acessibilidade ao mercado de trabalho em 2007 segue um padrão centro-periferia, a variação na acessibilidade entre 2007 e 2017 parece estar distribuída de forma mais aleatória pela cidade. Já o adensamento está mais concentrado, com alguma sobreposição em relação à acessibilidade ao mercado de trabalho em 2007. A Figura 4.3 plota a clusterização espacial dessas variáveis com o uso do método LISA (ANSELIN, 1995). Foram plotados apenas os clusters com significância estatística a $p \leq 0,5$.

O painel à esquerda mostra que de fato há um cluster Alto-Alto na acessibilidade ao mercado de trabalho em 2007 na região central da ci-

Figura 4.3 – *Clusterização Espacial da acessibilidade ao mercado de trabalho e adensamento construtivo em São Paulo*



Fonte: Elaboração própria

dade, incluindo alguns distritos das zonas Sul e Oeste. O padrão centro-periferia manifesta-se também na existência de clusters Baixo-Baixo nos extremos sul, leste e noroeste da cidade. O painel central assinala um cluster Alto-Alto na variação da acessibilidade ao mercado de trabalho no setor sudoeste da cidade, com alguns cluster Baixo-Alto à sua volta. Há no entanto menor concentração espacial. Já o painel à direita, que representa a clusterização do adensamento no período apresenta novamente um padrão centro-periferia mais claro. Cumpre assinalar que os clusters Alto-Alto dos três painéis situam-se em regiões próximas da cidade, havendo sobreposição entre eles.

Interessantemente, o cluster Alto-Alto para o adensamento construtivo inclui distritos que também estão presentes no cluster Alto-Alto para a variação na acessibilidade no período, mas que não estão presentes no cluster Alto-Alto da taxa basal de acessibilidade em 2007. Isso indica que ambas as variáveis (acessibilidade em 2007 e variação na acessibilidade no período 2007-2017) estão contribuindo para a concentração do adensamento construtivo.

As Tabelas 4.1, 4.2 e 4.3 e Quadros 4.1, 4.2 e 4.3 apresentam os resultados dos modelos estatísticos ajustados, que permitem testar essa hipótese. Todos os modelos consistem em regressões lineares múltiplas que têm como variáveis independentes a acessibilidade ao mercado de trabalho

em 2007 e variação na acessibilidade ao mercado de trabalho entre 2007 e 2017, ambas por distrito. A variável dependente varia, refletindo medidas progressivamente mais acuradas do adensamento construtivo no período. Todos os modelos são ajustados com base em 96 registros, que correspondem ao universo dos distritos da cidade de São Paulo.

Quadro 4.1 – Estatísticas – Modelo 1

Modelo 1	
Variável dependente	<i>área construída total</i>
R-squared	0.365
Log-lik.	-1382.8

Fonte: Elaboração própria.

Tabela 4.1 – Coeficientes – Modelo 1

Variável	Coeficiente	P-valor	2,5 percentil	97,5 percentil
intercepto	5069.4540	0.956	-1.78e+05	1.88e+05
Acessibilidade 2007	1.259e+06	0.000	8.82e+05	1.64e+06
Variação 2007-2017	1.799e+06	0.000	8.43e+05	2.75e+06

$n = 96$

Fonte: Elaboração própria.

O modelo 1, que tem como variável dependente a área construída bruta de construção de tipo residencial vertical acrescida no período 2007-2017 apresenta alta significância estatística para ambas as variáveis explicativas, mas não para o intercepto. Ele indica que, tudo o mais mantido igual, um distrito com 100% de acessibilidade ao mercado de trabalho em 2007 teria um aumento de aproximadamente 126 mil metros quadrados de área construída. Já uma variação de 100% nessa acessibilidade no período entre 2007 e 2017, para uma mesma acessibilidade inicial em 2007, resultaria em um aumento de aproximadamente 180 mil metros quadrados de área construída no distrito. Apesar da simplicidade da medida de adensamento, o modelo alcançou capacidade explicativa relativamente alta, com coeficiente de determinação de 36%.

O modelo 2 tem como variável dependente o adensamento de área construída residencial vertical, padronizado pela área efetivamente vertica-

Quadro 4.2 – Estatísticas – Modelo 2

Modelo 2	
Variável dependente	área construída total/ área verticalizável
R-squared	0.557
Log-lik.	98.639

Fonte: Elaboração própria.

Tabela 4.2 – Coeficientes – Modelo 2

Variável	Coeficiente	P-valor	2,5 percentil	97,5 percentil
intercepto	-0.0335	0.069	-0.070	0.003
Acessibilidade 2007	0.3768	0.000	0.302	0.452
Variação 2007-2017	0.4893	0.000	0.299	0.679

$n = 96$

Fonte: Elaboração própria.

lizável dos distritos. Assim como no modelo 1, o intercepto não alcançou significância estatística, e as variáveis independentes são altamente significantes, a $p < 0.001$. Esse modelo apresenta uma melhoria em relação ao modelo 1, tanto na verossimilhança quanto na capacidade explicativa, alcançando coeficiente de determinação de 55%. As variáveis independentes possuem tamanho de efeito substancial. Segundo o modelo, uma taxa basal de acessibilidade ao mercado de trabalho de 100% em 2007, tudo o mais mantido igual, acarretaria aumento de área construída equivalente a aproximadamente 38% da área verticalizável do distrito. Já uma taxa de variação de 100% na acessibilidade no período, para uma mesma taxa de acessibilidade inicial, resultaria em um aumento de área construída equivalente a 49%.

O modelo 3 é similar ao modelo 2, com a diferença de que agora estamos descontando, no cálculo da área edificável que define o adensamento construtivo, a área saturada por edificações de grande porte. Dessa vez, o intercepto apresenta significância estatística a, mas possui tamanho de efeito relativamente pequeno, variando de -0,1 a -0,01 no intervalo de confiança de 95%. Esse modelo tem performance similar ao modelo 2, com um coeficiente de determinação ligeiramente melhor, de 57%, e uma verossimilhança

Quadro 4.3 – Estatísticas – Modelo 3

Modelo 3	
Variável dependente	<i>área construída total/ área verticalizável não-saturada</i>
R-squared	0.570
Log-lik.	78.170

Fonte: Elaboração própria.

Tabela 4.3 – Coeficientes – Modelo 3

Variável	Coeficiente	P-valor	2,5 percentil	97,5 percentil
intercepto	-0.0559	0.015	-0.101	-0.011
Acessibilidade 2007	0.4826	0.000	0.390	0.575
Variação 2007-2017	0.6033	0.000	0.368	0.838

$n = 96$

Fonte: Elaboração própria.

um pouco menor, com redução de 20 unidades. As variáveis independentes permanecem altamente significantes, a $p < 0.001$, com tamanhos de efeito substanciais. Assim, segundo o modelo, uma acessibilidade ao mercado de trabalho de 100%, tudo o mais mantido igual, acarreta, para um dado distrito, em um aumento de área construída equivalente a 48% de sua área verticalizável. Já um aumento de 100% nessa taxa de acessibilidade no período 2007-2017, para uma mesma taxa basal em 2007, resulta em um aumento de área construída equivalente a 60% da área verticalizável.

4.4 CONSIDERAÇÕES FINAIS

Os três modelos apresentam alta capacidade explicativa, com variáveis independentes com alta significância estatística e tamanhos de efeitos substanciais. Eles apresentam evidências suficientes para rejeitar a hipótese nula e aceitar a hipótese alternativa de que há uma relação positiva entre a acessibilidade ao mercado de trabalho e o adensamento construtivo de tipo residencial vertical na cidade de São Paulo no período de 2007 a 2017. Dessa forma, o estudo contribui para a literatura empírica sobre o modelo urbano padrão ao apresentar evidências para confirmar sua principal con-

tribuição teórica (DURANTON; PUGA, 2015), assim como ao propor uma metodologia de verificação empírica que prescinde do pressuposto de um distrito comercial central (CBD) único.

A despeito do emprego dessa metodologia, os dados indicam que, em São Paulo, a relação positiva entre acessibilidade e adensamento construtivo resultou na formação de centralidades urbanas definidas, como previsto pelo SMM. Essas centralidades foram identificadas por meio da análise LISA realizada, cujos resultados assinalaram clusters espaciais Alto-Alto na área central da cidade tanto para o adensamento construtivo, quanto para a acessibilidade ao mercado de trabalho. Esses resultados fornecem subsídios para avaliar criticamente propostas do planejamento urbano que buscam gerar cidades policêntricas por meio de intervenções apenas a nível local, como a cidade de 15 minutos (MORENO et al., 2021).

Em consonância com Bertaud (2022b) e Gray (2022), conclui-se que os resultados do estudo assinalam a importância de que as políticas públicas urbanas não estejam alicerçadas exclusivamente no conhecimento e experiências pessoais dos planejadores urbanos sobre a cidade, mas sim sejam baseadas em evidências sobre as dinâmicas socioeconômicas mais amplas que afetam a estrutura urbana e que só podem ser verificadas em grande escala, tais como aquelas descritas pelo SMM e aqui verificadas empiricamente.

Como argumentam DeSalvo e Su (2018), é mister que as propostas que buscam combater o espraiamento por meio da promoção do adensamento urbano sejam dotadas de indicadores objetivos que permitam monitorar seus resultados e, principalmente, sejam elaboradas com base em teorias explicitadas, de modo a tornar claros os mecanismos causais que as fundamentam — o que não ocorre no caso da cidade de 15 minutos (KHAVARIAN-GARMSIR et al., 2023). Isto é importante pois intervenções urbanas baseadas nessas propostas, ao não considerarem as dinâmicas mais amplas aqui referenciadas, não apenas têm alta chance de falha, como também acarretam riscos de gerar efeitos indesejados, como a gentrificação e aumento da segregação (CASARIN; MACLEAVY; MANLEY, 2023; ABBIASOV et al., 2023).

Os resultados deste estudo mostram também que há um caminho por meio do qual é possível promover a criação de cidades policêntricas, com

base em evidências e de forma a tornar a cidade mais integrada, sem incorrer diretamente nos riscos de segregação e gentrificação aqui assinalados. Em todos os modelos estatísticos ajustados, a variação na acessibilidade ao mercado de trabalho ao longo da década 2007-2017 apresentou tamanhos de efeito em relação à densidade construtiva superiores àqueles da acessibilidade inicial em 2007. Esses resultados indicam que políticas públicas que promovam reduções nos tempos e custos de deslocamento podem levar à criação de novas centralidades, conforme previsto pela simulação realizada com base no [SMM](#) apresentada no Capítulo 2.

Há diversas medidas que podem ser realizadas para alcançar tais resultados. Elas se organizam em dois principais grupos:

- Medidas que buscam aumentar a acessibilidade das regiões da cidade aos postos de trabalho existentes;
- Medidas que buscam atrair os postos de trabalho para regiões com baixa acessibilidade;

O primeiro grupo de medidas envolve ações de ampliação da rede de transporte público de massa – como metrô, VLTs e corredores de ônibus. O segundo grupo de medidas tem por objetivo a descentralização dos postos de trabalho por meio, por exemplo, da criação de distritos com incentivos fiscais para localização de firmas voltadas a ramos de atividade específico, como distritos de inovação e tecnologia. Neste caso, será preciso também investir para garantir a infraestrutura urbana adequada para instalação desses serviços.

Por fim, a melhoria na performance dos modelos estatísticos conforme foram sendo utilizadas medidas progressivamente mais acuradas do adensamento construtivo, levando em consideração não apenas o acréscimo bruto de área construída, mas padronizando-o em relação à área efetivamente edificável, demonstra que esse é um cuidado metodológico importante em estudos que tenham a densidade urbana como objeto. A melhoria substancial de performance dos modelos ao serem considerados os impedimentos regulatórios à edificação indica, além disso, que estas regulações resultam na redução da área construída esperada.

De acordo com a lógica estabelecida pelo SMM, pode-se deduzir que a área construída seria maior, não fossem esses impedimentos. Conside-

rando sobretudo o zoneamento restritivo estritamente residencial, que está associado aos vazios de potencial construtivo na área central da cidade identificados na Figura 4.1, esses dados apontam para a necessidade de estudos posteriores que avaliem os efeitos negativos deste tipo de zoneamento no equilíbrio espacial urbano.

5 CONCLUSÃO

Em “As Ilusões do Plano Diretor” Flávio Villaça (2005), o falecido professor da Faculdade de Arquitetura e Urbanismo da Universidade de São Paulo analisa o processo de elaboração do Plano Diretor da cidade de São Paulo, promulgado em 2002¹, do qual participou ativamente. Na obra, Villaça explora os limites desse instrumento de planejamento urbanístico fundamentando sua análise na experiência adquirida como urbanista e servidor público da antiga Secretaria de Planejamento da Prefeitura de São Paulo.

Em uma das principais passagens da obra, por meio de uma descrição etnográfica das tensões sociais presentes na audiência pública do distrito de Pinheiros, o autor mostra como o discurso do planejamento urbano é mobilizado por dois diferentes setores da elite da cidade para defender seus interesses. De um lado, está a elite tradicional, detentora de terras, que busca garantir sua reserva de capital por meio da especulação imobiliária e da manutenção da exclusividade dos bairros onde reside. De outro, a máquina do capital imobiliário, constituída por incorporadoras e construtoras, ávida em obter terras nesses mesmos bairros para maximizar seu lucro, vendendo empreendimentos também exclusivos ao capitalizar sobre aquilo mesmo que valoriza os imóveis da elite tradicional: sua localização privilegiada na cidade.

No cerne do debate entre esse dois atores não estão os princípios urbanísticos norteadores do Plano Diretor, as diretrizes e a visão de cidade propostas pelos especialistas do Poder Público e da Academia, apesar de estes serem constantemente citados. O que está em jogo é algo mais específico e localizado: o zoneamento de alguns poucos bairros, que definirá o coeficiente de aproveitamento e a possibilidade de verticalização dos terrenos neles localizados. Disputam-se os imóveis localizados nas áreas de residência da elite, os chamados “bairros nobres”, situados no setor sudoeste da cidade (VILLAÇA, 2001). Neste cenário, o Plano Diretor parece deixar de ser protagonista da transformação urbana, para tornar-se o pano de fundo ideológico de um processo de capitalização da terra que dele se vale para avançar. E a periferia, ainda que chamada a atuar nas audiências públicas,

¹Trata-se do [PDE](#) anterior ao atual.

permanece fora de cena, alheia à disputa que se desenrola no processo de elaboração do plano.

Ao analisar os processos históricos e fatores socioeconômicos determinantes na construção do espaço urbano da cidade de São Paulo, Villaça (2001) argumenta que a cidade não pode ser concebida apenas como uma área urbanizada, pois, segundo ele, há uma estrutura urbana subjacente, com alta permanência histórica, que organiza o espaço intraurbano. Essa estrutura determina os territórios de moradia da elite e das classes oprimidas, refletindo e reforçando a desigualdade social da qual ela se originou. Villaça (2001) identifica no setor sudoeste da cidade o local de residência da elite, para o qual o interesse do capital imobiliário se volta.

Ao analisar os locais de moradia das diferentes classes sociais em São Paulo por meio da categorização dos dados de ocupação dos chefes de família coletados nos Censos de 2000 e 2010, Marques (2014) identifica uma maior concentração das classes altas em uma região similar à apontada por Villaça (MARQUES, 2014, páginas 699 e 701). Segundo o autor, os dados indicam que há em São Paulo uma estrutura geral de segregação com uma clara hierarquia entre os grupos. Refletindo sobre as alterações nesses padrões de residência ao longo da década, o autor afirma que os espaços de elite passaram por um processo de homogeneização, ao mesmo tempo em que os espaços das classes média e periféricos se tornaram mais diferenciados. Essas mudanças no entanto não são suficientes para vencer a exclusão social que vigora nos espaços da elite e que resulta no que o autor denomina uma dinâmica de *evitação social* por parte das classes superiores.

Os resultados alcançados neste estudo apresentam evidências de que, passados vinte anos da análise de Flávio Villaça (2005) e uma década da análise de Marques (2014), o novo Plano Diretor não obteve êxito em romper com a estrutura espacial de segregação social que subjaz ao espaço intraurbano paulistano. Pois, de acordo com os modelos estatísticos empregados, os incentivos construtivos fornecidos pelo PDE promulgado em 2014 resultaram em efeitos significativos na promoção do adensamento apenas nas áreas de classe média e, especialmente, na área nobre da cidade, sem gerar impactos relevantes nas periferias.

Os modelos indicam também que não foram os incentivos previstos pelo Plano que se revelaram mais efetivos, mas sim as restrições constru-

tivas por ele convalidadas – historicamente instituídas pela elite da cidade sobre seus bairros de moradia e por ela mobilizadas para preservar seus interesses especulativos e reforçar a segregação. Os dados aqui analisados assinalam, assim, uma possível via de intervenção que, sem a necessidade de dispêndio de recursos públicos, pode levar de forma muito mais efetiva ao alcance dos objetivos de adensamento e redução dos tempos de deslocamento ao trabalho estabelecidos pelo [PDE](#) para os Eixos de Estruturação da Transformação Urbana: a substituição dos zoneamentos exclusivamente residenciais unifamiliares nas áreas centrais da cidade por zonas mistas de média a alta densidade.

O impacto desta restrições sobre o equilíbrio urbano da cidade de São Paulo mostra-se assim como um importante objeto de estudos futuros. Mas, principalmente, esta dissertação assinala a importância de que as políticas públicas urbanas não estejam alicerçadas exclusivamente no conhecimento e experiências pessoais dos planejadores urbanos sobre a cidade, mas sim sejam baseadas em evidências sobre as dinâmicas socioeconômicas mais amplas que afetam a estrutura urbana e que só podem ser verificadas em grande escala. Só assim – por meio da criação de capacidades estatais e fortalecimento das estruturas de planejamento governamental – o Plano Diretor poderá deixar de ser considerado uma ilusão (Flávio Villaça, 2005).

REFERÊNCIAS BIBLIOGRÁFICAS

ABBIASOV, T. et al. The 15-minute city quantified using mobility data. *Nature Human Behaviour*, 2023. Working Paper No. W30752. Citado na pág. 33, 177.

ALEXANDER, E. R. **Density Measures: a review and analysis.** *Journal of Architectural and Planning Research*, Jstor, v. 10, n. 3, p. 181–202, 1993. doi:. Citado na pág. 229.

ALONSO, W. *Location and Land Use: Toward a General Theory of Land Rent*. [S.l.]: Harvard University Press, 1964. doi:. Citado na pág. 35, 37, 38.

ANDERSON, M. B. New urbanism. In: *International Encyclopedia of Geography: People, the Earth, Environment and Technology*. [S.l.]: Wiley, 2017. p. 1–3. Citado na pág. 29.

ANSELIN, L. **Local Indicators of Spatial Association - LISA.** *Geographical Analysis*, Wiley, v. 27, n. 2, p. 93–115, 1995. doi:. Citado na pág. 45, 65, 66, 68, 70, 111, 112, 169, 172.

ANSELIN, L. The moran scatterplot as an esda tool to assess local instability in spatial association. In: FISCHER, M.; SCHOLTEN, H.; UNWIN, D. (Ed.). *Spatial Analytical Perspectives on GIS in Environmental and Socio-Economic Sciences*. London: Taylor & Francis, 1996. p. 111–125. Citado na pág. 67, 70, 112.

BALBIM, R.; KRAUSE, L. *Eixos de Estruturação da Transformação Urbana: inovação e avaliação em São Paulo*. [S.l.]: IPEA, 2016. Citado na pág. 30, 50, 165, 234.

BECKMANN, M. J. Equilibrium models of residential land use. *Regional and Urban Economics*, v. 3, n. 4, p. 361–368, 1973. Citado na pág. 37.

BERTAUD, A. *The Last Utopia: the 15-minute city*. 2022. Urban Reform Institute. Disponível em: <<https://urbanreforminstitute.org/wp-content/uploads/2022/02/15mincity-bertaud.pdf>> (acessado em 10 de maio de 2024). Citado na pág. 33, 34.

BERTAUD, A. *Ordem Sem Design: como os mercados moldam as cidades*. [S.l.]: Bookman, Porto Alegre, 2022. Citado na pág. 35, 40, 120, 133, 166, 177, 256.

BLYTH, C. R. On simpson's paradox and the sure-thing principle. *Journal of the American Statistical Association*, v. 67, p. 338, 1972. Disponível em: <<https://www.jstor.org/stable/2284382>>. Citado na pág. 91.

- BOULOS, Guilherme. *Discurso à Câmara dos Deputados - Sessão 37.2023*. 2023. Disponível em: <<https://www.camara.leg.br/internet/SitaqWeb/TextoHTML.asp?etapa=5&nuSessao=37.2023&nuQuarto=2867823&nuOrador=1&nuInsercao=1&dtHorarioQuarto=15:44&sgFaseSessao=BC&Data=29/03/2023&txApelido=Guilherme%20Boulos,%20PSOL-SP&txFaseSessao=Breves%20Comunica%C3%A7%C3%B5es&txTipoSessao=Ordin%C3%A1ria%20-%20CD&dtHoraQuarto=15:44&txEtapa=>>>. Acesso em: 17 fev. 2024. Citado na pág. 30.
- BRUCKNER, J. K. The structure of urban equilibria – a unified treatment of the muth-mills model. In: BRUCKNER, J. K. (Ed.). *Handbook of Regional and Urban Economics*. [S.l.]: Elsevier, 1987. v. 2, cap. 20, p. 821–845. Citado na pág. 35.
- CALTHORPE, P. *The Next American Metropolis: Ecology, Community, and the American Dream*. New York: Princeton Architectural Press, 1993. Citado na pág. 29, 30, 234.
- CARPENTER, B. et al. Stan: A probabilistic programming language. *Journal of statistical software*, Columbia Univ., New York, NY (United States); Harvard Univ., Cambridge, MA (United States), v. 76, n. 1, 2017. Citado na pág. 197, 198.
- CASARIN, G.; MACLEAVY, J.; MANLEY, D. Rethinking urban utopiaism: the fallacy of social mix in the 15-minute city. *Urban Studies*, SAGE Publications, v. 60, n. 16, p. 3167–3186, 2023. Citado na pág. 32, 177.
- CHURCHMAN, A. **Disentangling the Concept of Density**. *Journal of Planning Literature*, Sage Journals, v. 13, n. 4, p. 389–411, 1999. doi:. Citado na pág. 215, 216.
- CINELLI, C.; FORNEY, A.; PEARL, J. A crash course in good and bad controls. *Sociological Methods and Research*, v. 0, n. 0, p. Pre–print, 2022. Disponível em: <<https://doi.org/10.1177/00491241221099552>>. Citado na pág. 137.
- CINELLI, C.; FORNEY, A.; PEARL, J. **A Crash Course in Good and Bad Controls**. *Sociological Methods & Research*, Wiley, v. 0, n. 0 (pre-print), p. 93–115, 2022. doi:. Citado na pág. 213.
- CLAYTON, A. *Bernoulli's Fallacy: Statistical Illogic and the Crisis of Modern Science*. New York: Columbia University Press, 2022. ISBN 9780231199940. Disponível em: <<https://cup.columbia.edu/book/bernoullis-fallacy/9780231199940>>. Citado na pág. 72.

CONTRIBUTORS, W. *Bairro nobre* — *Wikipédia, a enciclopédia livre*. 2024. Acessado em: outubro de 2024. Disponível em: <https://pt.wikipedia.org/wiki/Bairro_nobre>. Citado na pág. 59.

CONTRIBUTORS, W. *Location, Location, Location* — *Wikipedia, The Free Encyclopedia*. 2024. Accessed: October 12, 2024. Disponível em: <https://en.wikipedia.org/wiki/Location,_Location,_Location>. Citado na pág. 58.

COURGEAU, D. *Methodology and Epistemology of Multilevel Analysis: Approaches from Different Social Sciences*. Springer Science, 2003. Disponível em: <<https://link.springer.com/book/10.1007/978-1-4020-4675-9>>. Citado na pág. 94, 95, 97.

Câmara Municipal de São Paulo e Prefeitura Municipal de São Paulo. **Revisão Intermediária do Plano Diretor Estratégico do Município de São Paulo – PL 127/2023**. 2023. Disponível em: <<https://legislacao.prefeitura.sp.gov.br/leis/projeto-de-lei-executivo-127-de-22-de-marco-de-2023>>. Acesso em: 01 fev. 2024. Citado na pág. 203.

DESALVO, J. S.; SU, Q. The determinants of urban sprawl: theory and estimation. *International Journal of Urban Sciences*, Taylor & Francis, v. 23, n. 1, p. 88–104, 2018. Citado na pág. 31, 166, 177.

DURANTON, G.; PUGA, D. Urban land use. In: DURANTON, G.; HENDERSON, J. V.; STRANGE, W. C. (Ed.). *Handbook of Regional and Urban Economics*. [S.l.]: Elsevier, 2015. v. 5, cap. 8, p. 467–560. Citado na pág. 26, 42, 166, 177.

FERRARI, T.; MONTEIRO, A. M. V.; AMARAL, P. V. M. Estrutura familiar e padrões espaciais da escolha residencial. *Revista Brasileira de Estudos Regionais e Urbanos*, v. 13, n. 3, p. 405–421, 2020. Citado na pág. 34, 37.

Flávio Villaça. *As Ilusões do Plano Diretor*. [S.l.]: Publicação própria realizada pelo autor, 2005. Citado na pág. 180, 181, 182.

FREEMARK, Y. **Upzoning Chicago: Impacts of a Zoning Reform on Property Values and Housing Construction**. *Urban Affairs Review*, Sage Journals, v. 56, n. 3, p. 758–789, 2020. doi:. Citado na pág. 32.

FUJITA, M.; OGAWA, H. Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics*, v. 12, n. 2, p. 161–196, 1982. Citado na pág. 42.

FUNARI, A. P. *Eixos de Estruturação da Transformação Urbana: inovação e avaliação em São Paulo*. [S.l.]: IPEA, 2016. Citado na pág. 23, 31.

GAFISA. *Tonino Lamborghini Residences*. 2024. Acessado em: outubro de 2024. Disponível em: <<https://www.gafisa.com.br/imoveis/residencial/tonino-lamborghini>>. Citado na pág. 58.

GELMAN, A. et al. Bayesian workflow. *arXiv preprint arXiv:2011.01808*, 2020. Disponível em: <<https://arxiv.org/abs/2011.01808>>. Citado na pág. 25, 46, 47, 73.

GEOINFO/SMUL. **Camada - Distritos Municipais**. 2018. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/51f75eab-6d90-451e-9f80-c21c1ac72630>>. Acesso em: 27 mar. 2023. Citado na pág. 206.

GEOINFO/SMUL/PMSP. **Camada - Eixos de Estruturação da Transformação Urbana**. 2014. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/c7077dda-16b8-4b7d-ad32-24b1e6404162>>. Acesso em: 27 mar. 2023. Citado na pág. 207.

GEOINFO/SMUL/PMSP. **Camada - Macrozona PDE**. 2014. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/b1bfd556-f75d-4be1-93f3-24287474b126>>. Acesso em: 22 jan. 2024. Citado na pág. 207.

GEOINFO/SMUL/PMSP, CASE/SMUL/PMSP. **Camada - Áreas com Restrições Geotécnicas**. 2016. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/55000add-0eb3-4a53-9696-739dddb6790e>>. Acesso em: 10 jun. 2023. Citado na pág. 211.

GEOINFO/SMUL/PMSP, CASE/SMUL/PMSP. **Camada - Rampa Heliponto**. 2020. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/b1bfd556-f75d-4be1-93f3-24287474b126>>. Acesso em: 10 jun. 2023. Citado na pág. 211.

GEOINFO/SMUL/PMSP, CPA/SMA/SP. **Camada - Área de Proteção ao Manancial Alto Juquery**. 2015. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/30cf83cf-ffe0-411b-8136-072b253d97eb>>. Acesso em: 10 jun. 2023. Citado na pág. 210.

GEOINFO/SMUL/PMSP, CPA/SMA/SP. **Camada - Área de Proteção ao Manancial Guarapiranga**. 2015. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/>>

catalog.search#/metadata/7ebdcdfc-2a37-4495-838e-bded3bf85073>.

Acesso em: 10 jun. 2023. Citado na pág. 210.

GEOINFO/SMUL/PMSP, CPA/SMA/SP. **Camada - Área de Proteção ao Manancial Billings**. 2019. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/30cf83cf-ffe0-411b-8136-072b253d97eb>>.

Acesso em: 10 jun. 2015. Citado na pág. 210.

GEOINFO/SMUL/PMSP, CPA/SVMA/PMSP. **Camada - Parques e Unidades de Conservação**. 2023. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/bb465396-edc1-45b6-8438-6a9b56c1ddfe>>. Acesso em: 10 jun. 2023. Citado na pág. 210.

GEOINFO/SMUL/PMSP, CPA/SVMA/PSMP. **Camada - Reservas Particulares do Patrimônio Natural**. 2023. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/1e49a81a-177d-434d-b2aa-0f78ea4ead7>>.

Acesso em: 10 jun. 2023. Citado na pág. 210.

GEOINFO/SMUL/PMSP, DESAP/SMJ/PMSP. **Camada - Decretos de Interesse Social e de Utilidade Pública**. 2023. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/a6e17a6b-20d4-4d68-8e95-51b47d6db6c4>>.

Acesso em: 10 jun. 2023. Citado na pág. 208.

GEOINFO/SMUL/PMSP, DEUSO/SMUL/PMSP. **Camada - Zoneamento**. 2015. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/b59e58ea-4087-4369-ab47-5c58f7259c10>>. Acesso em: 27 mar. 2023. Citado na pág. 207.

GEOINFO/SMUL/PMSP, DPH/SC/PMSP. **Camada - Bairro Ambiental**. 2019. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/a6e17a6b-20d4-4d68-8e95-51b47d6db6c4>>. Acesso em: 10 jun. 2023. Citado na pág. 209.

GEOINFO/SMUL/PMSP, DPH/SC/PMSP. **Camada - Bens Tombados**. 2019. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/a6e17a6b-20d4-4d68-8e95-51b47d6db6c4>>. Acesso em: 10 jun. 2023. Citado na pág. 209.

GEOINFO/SMUL/PMSP, DPH/SC/PMSP. **Camada - Área Envoltória CONDEPHAAT**. 2019. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/ff73fa2d-b477-4492-bcd7-3f2a93245f01>>. Acesso em: 10 jun. 2023. Citado na pág. 209.

GEOINFO/SMUL/PMSP, DPH/SC/PMSP. **Camada - Área Envoltória CONPRES**. 2019. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/3fc14d45-85c0-4881-88ed-e5fc6321606d>>. Acesso em: 10 jun. 2023. Citado na pág. 209.

GEOINFO/SMUL/PMSP, DPH/SC/PMSP. **Camada - Área Envoltória IPHAN**. 2019. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/115368cd-693a-43bd-82d3-7c05e27598a0>>. Acesso em: 10 jun. 2023. Citado na pág. 209.

GEOINFO/SMUL/PMSP, IBGE. **Camada - Setores Censitários 2010**. 2010. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/008874e3-7c69-4f09-b5fb-432f26c57da7>>. Acesso em: 27 mar. 2023. Citado na pág. 206.

GEOINFO/SMUL/PMSP, Metrô/SP. **Camada - Zonas da Pesquisa Origem e Destino de 2017**. 2017. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/4ed47f83-653f-43ad-9883-4137dbbfaee1>>. Acesso em: 27 mar. 2023. Citado na pág. 206.

GEOINFO/SMUL/PMSP, SF/PMSP. **Camada - Quadras Fiscais**. 2018. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/4ca0d339-8510-4f4d-aac4-a5f6ee570c0b>>. Acesso em: 27 mar. 2023. Citado na pág. 206, 227.

GEOINFO/SMUL/PMSP, SMSU/PMSP. **Camada - Ocupações Subnormais em Áreas de Risco Geológico**. 2023. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/60e961a1-d2dd-42fa-befd-d374fa302f5b>>. Acesso em: 10 jun. 2023. Citado na pág. 211.

GEOINFO/SMUL/PMSP, SVMA/PMSP. **Camada - Áreas de Proteção Ambiental (APAs)**. 2018. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/6bd56f01-2164-45f5-ac0e-7ccd640aa475>>. Acesso em: 10 jun. 2023. Citado na pág. 210.

GEOINFO/SMUL/PMSP, SVMA/PMSP. **Camada - Áreas Contaminadas**. 2019. Disponível em: <<https://metadados.geosampa.prefeitura.sp.gov.br/geonetwork/srv/eng/catalog.search#/metadata/6bd56f01-2164-45f5-ac0e-7ccd640aa475>>. Acesso em: 10 jun. 2023. Citado na pág. 211.

GLAESER, E. *The 15-minute city is a dead end – cities must be places of opportunity for everyone*. 2021. London School of Economics Blog. Acessado em 10 de maio de 2024. Disponível em: <<https://blogs.lse.ac.uk/>>. Citado na pág. 34.

GOLDSTEIN, H. *Multilevel Statistical Models*. [S.l.]: Arnold Publishers, 1999. Citado na pág. 95.

Governo Federal do Brasil. **Estatuto da Cidade - Lei 10.257/2001**. 2001. Disponível em: <https://www.planalto.gov.br/ccivil_03/Leis/Leis_2001/110257.htm>. Acesso em: 01 mai. 2024. Citado na pág. 217.

GRAY, M. N. *Arbitrary Lines: How Zoning Broke the American City and How to Fix It*. Washington, DC: Island Press, 2022. Citado na pág. 166, 177.

GREENACRE, M. et al. Principal component analysis. *Nature Reviews Methods Primers*, v. 2, p. 100, 2022. Disponível em: <<https://doi.org/10.1038/s43586-022-00184-w>>. Citado na pág. 45, 65.

HARRIS, C. R. et al. Array programming with NumPy. *Nature*, Springer, v. 585, n. 7825, p. 357–362, 2020. Citado na pág. 198.

HOFFMAN, M. D.; GELMAN, A. *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo*. 2011. Citado na pág. 76, 197.

HOYLER, T. **Regulando a incorporação imobiliária em São Paulo: burocracia, instrumentos e negociações**. *Revista Brasileira de Estudos Urbanos e Regionais*, v. 21, n. 1, 2019. doi:. Citado na pág. 211, 242.

HSU, T. *He Wanted to Unclog Cities. Now He's 'Public Enemy No. 1.'* 2023. *The New York Times*. Accessed: 2024-05-10. Disponível em: <<https://www.nytimes.com/2023/03/28/technology/carlos-moreno-15-minute-cities-conspiracy-theories.html>>. Citado na pág. 31.

HUNTER, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, IEEE COMPUTER SOC, v. 9, n. 3, p. 90–95, 2007. Citado na pág. 198.

IMBENS, G. W.; LEMIEUX, T. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, Elsevier, v. 142, n. 2, p. 615–635, 2008. Citado na pág. 149.

IMÓVEIS, A. *A importância do mapa de localização no setor imobiliário*. 2020. Acessado em: 12 de outubro de 2024. Disponível em: <<https://www.aqua.com.br/blog/mapa-localizacao-no-setor-imobiliario>>. Citado na

pág. 59.

JORDAHL, K. et al. *geopandas/geopandas: v0.8.1*. Zenodo, jul. 2020. Disponível em: <<https://doi.org/10.5281/zenodo.3946761>>. Citado na página 198.

KHAVARIAN-GARMSIR, A. R. et al. **From Garden City to 15-Minute City: A Historical Perspective and Critical Assessment**. *Land*, MDPI, v. 12, n. 2, 2023. ISSN 2073-445X. doi:. Citado na página 23, 30, 31, 177.

KLUYVER, T. et al. Jupyter notebooks – a publishing format for reproducible computational workflows. In: LOIZIDES, F.; SCHMIDT, B. (Ed.). *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. [S.l.: s.n.], 2016. p. 87 – 90. doi:. Citado na página 199.

KUMAR, R. et al. Arviz a unified library for exploratory analysis of bayesian models in python. *Journal of Open Source Software*, The Open Journal, v. 4, n. 33, p. 1143, 2019. Disponível em: <<https://doi.org/10.21105/joss.01143>>. Citado na página 198.

LAMPORT, L.; PROJECT, T. L. *The L^AT_EX Project: A Document Preparation System*. [S.l.], 1994. Accessed: 2024-09-08. Disponível em: <<https://www.latex-project.org/>>. Citado na página 199.

LEWANDOWSKI, D.; KUROWICKA, D.; JOE, H. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, Elsevier, v. 100, p. 1989–2001, 2009. Citado na página 97.

LIOTTA, C.; VIGUIÉ, V.; LEPETIT, Q. Testing the monocentric standard urban model in a global sample of cities. *Regional Science and Urban Economics*, Elsevier, v. 97, p. 103832, 2022. Citado na página 41.

MAAS, C. J. M.; HOOX, J. J. **Sufficient sample sizes for multilevel modeling**. *Methodology*, Hogrefe & Huber Publishers, v. 1, n. 3, p. 86–92, 2005. Disponível em: <<https://www.joophox.net/publist/methodology05.pdf>>. Citado na página 72.

MANI, A.; RILEY, E. Social networks as levers of mobility. In: IVERSEN, V.; KRISHNA, A.; SEN, K. (Ed.). *Social Mobility in Developing Countries: Concepts, Methods and Determinants*. [S.l.]: Oxford University Press, 2021. cap. 17. Citado na página 33.

MARQUES, E. **Estrutura Social e Segregação em São Paulo: Transformações na Década de 2000**. *Dados - Revista de Ciências*

Sociais, Scielo, v. 57, n. 3, p. 675–710, 2014. ISSN 2073-445X. doi:. Citado na pág. 114, 162, 181.

MCELREATH, R. ***Statistical Rethinking***. [S.l.]: CRC Press, 2020. Citado na pág. 46, 72, 73, 86, 97, 99, 137, 145.

MCKINNEY, W. et al. Data structures for statistical computing in python. In: AUSTIN, TX. *Proceedings of the 9th Python in Science Conference*. [S.l.], 2010. v. 445, p. 51–56. Citado na pág. 198.

MILLS, E. S. **An Aggregative Model of Resource Allocation in a Metropolitan Area**. *Journal of Economic Theory*, Jstor, v. 57, n. 2, p. 197–210, 1967. doi:. Citado na pág. 35.

MORAN, P. A. P. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 10, p. 243–251, 1948. Citado na pág. 66, 67.

MORENO, C. et al. **Introducing the “15-Minute City”: Sustainability, Resilience and Place Identity in Future Post-Pandemic Cities**. *Smart Cities*, MDPI, v. 4, n. 1, p. 93–111, 2021. ISSN 2624-6511. doi:. Citado na pág. 30, 33, 177.

MTE. *RAIS - Quem deve declarar?* 2024. Disponível em: <http://www.rais.gov.br/sitio/quem_deve_declarar.jsf>. Acesso em: Acessado em: 20 de maio de 2024. Citado na pág. 214.

MTE. *RAIS - Quem deve ser relacionado?* 2024. Disponível em: <http://www.rais.gov.br/sitio/quem_deve_ser_relacionado.jsf>. Acesso em: Acessado em: 20 de maio de 2024. Citado na pág. 214.

MUTH, R. F. ***Cities and Housing: The Spatial Pattern of Urban Residential Land Use***. [S.l.]: University of Chicago Press, 1969. Citado na pág. 35.

OECD. **Causes and consequences of urban sprawl**. In: *Rethinking Urban Sprawl: Moving Towards Sustainable Cities*. [S.l.]: OECD Publishing, 2018. doi:. Citado na pág. 29.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Citado na pág. 198.

PINTO, V. C. ***Direito Urbanístico: Plano Diretor e Direito de Propriedade***. [S.l.]: Editora Revista dos Tribunais, 2012. Citado na pág. 217.

PMSP. *Lei nº 10.235, de 16 de dezembro de 1986*. 1986. <<https://www.prefeitura.sp.gov.br/cidade/secretarias/upload/arquivos/secretarias/financas/legislacao/Lei-10235-1986.pdf>>. Dispõe sobre o Imposto Predial e Territorial Urbano (IPTU) no Município de São Paulo. Citado na pág. 237.

Prefeitura Municipal de São Paulo. ***Consolidação das Leis Tributárias do Município de São Paulo - Decreto 61.810/2022***. 2014. Disponível em: <<https://legislacao.prefeitura.sp.gov.br/leis/decreto-61810-de-14-de-setembro-de-2022>>. Acesso em: 01 fev. 2024. Citado na pág. 200, 201, 238.

Prefeitura Municipal de São Paulo. ***Plano Diretor Estratégico do Município de São Paulo - Lei 16.050/2014***. 2014. Disponível em: <<https://gestaourbana.prefeitura.sp.gov.br/wp-content/uploads/2015/01/Plano-Diretor-Estrat%C3%A9gico-Lei-n%C2%BA-16.050-de-31-de-julho-de-2014-Estrat%C3%A9gias-ilustradas.pdf>>. Acesso em: 01 fev. 2024. Citado na pág. 50, 51, 52.

Prefeitura Municipal de São Paulo. ***Lei de Uso e Ocupação do Solo - Lei 16.402/2016***. 2016. Disponível em: <<https://gestaourbana.prefeitura.sp.gov.br/wp-content/uploads/2016/03/PL-272-15-com-raz%C3%B5es-de-veto1.pdf>>. Acesso em: 01 fev. 2024. Citado na pág. 217, 232, 234.

Prefeitura Municipal de São Paulo. ***Código de Obras e Edificações - Lei 16.642/2017***. 2017. Disponível em: <https://gestaourbana.prefeitura.sp.gov.br/wp-content/uploads/2018/04/codigo_de_obras_ilustrado.pdf>. Acesso em: 01 fev. 2024. Citado na pág. 217, 218, 231.

QUINTOANDAR. *Bairros Nobres: Guia de Cidades*. 2024. Acessado em: outubro de 2024. Disponível em: <<https://www.quintoandar.com.br/guias/cidades/bairros-nobres/#h-sao-paulo>>. Citado na pág. 59.

REY, S. J.; ANSELIN, L. PySAL: A Python Library of Spatial Analytical Methods. *The Review of Regional Studies*, v. 37, n. 1, p. 5–27, 2007. Citado na pág. 68, 198.

RODRÍGUEZ-POSE, A.; STORPER, M. **Housing, urban growth and inequalities: The limits to deregulation and upzoning in reducing economic and spatial inequality**. *Urban Studies*, Sage Journals, v. 57, n. 2, p. 223–248, 2020. ISSN 0965-8564. doi:. Citado na pág. 32, 33.

SF/PMSP. ***Cadastro Imobiliário Fiscal - 2007***. 2007. Disponível em: <https://geosampa.prefeitura.sp.gov.br/PaginasPublicas/downloadIfr.aspx?orig=DownloadCamadas&arq=12_Cadastro\IPTU_INTER\XLS_>

CSV\IPTU_2007&arqTipo=XLS_CSV>. Acesso em: 27 mai. 2023. Citado na pág. 204.

SF/PMSP. **Cadastro Imobiliário Fiscal - 2014**. 2014. Disponível em: <https://geosampa.prefeitura.sp.gov.br/PaginasPublicas/downloadIfr.aspx?orig=DownloadCamadas&arq=12_Cadastro\IPTU_INTER\XLS_CSV\IPTU_2014&arqTipo=XLS_CSV>. Acesso em: 27 mai. 2023. Citado na pág. 203.

SF/PMSP. **Cadastro Imobiliário Fiscal - 2017**. 2017. Disponível em: <https://geosampa.prefeitura.sp.gov.br/PaginasPublicas/downloadIfr.aspx?orig=DownloadCamadas&arq=12_Cadastro\IPTU_INTER\XLS_CSV\IPTU_2017&arqTipo=XLS_CSV>. Acesso em: 27 mai. 2023. Citado na pág. 204.

SF/PMSP. **Cadastro Imobiliário Fiscal - 2023**. 2023. Disponível em: <https://geosampa.prefeitura.sp.gov.br/PaginasPublicas/downloadIfr.aspx?orig=DownloadCamadas&arq=12_Cadastro\IPTU_INTER\XLS_CSV\IPTU_2023&arqTipo=XLS_CSV>. Acesso em: 27 mai. 2023. Citado na pág. 203.

SF/PMSP. *Manual do Usuário da Declaração Tributária de Conclusão*. [S.l.], 2024. Acessado em: 20 de maio de 2024. Disponível em: <https://www.prefeitura.sp.gov.br/cidade/upload/Manual-do-usuario-da-Declaracao-Tributaria-de-Conclusao_1389383273.pdf>. Citado na pág. 217.

SHERTZER, A.; TWINAM, T.; WALSH, R. P. Zoning and segregation in urban economic history. *Regional Science and Urban Economics*, v. 94, p. 103652, 2022. ISSN 0166-0462. Urban Economics and History. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0166046221000120>>. Citado na pág. 61.

SIMMEL, G. As grandes cidades e a vida do espírito. *Mana*, v. 11, n. 2, p. 577–591, 2005. Disponível em: <<https://www.scielo.br/j/mana/a/WfkbJzPmYNdfNWxpyKpcwWj/>>. Citado na pág. 216.

THÜNEN, J. H. von. *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. [S.l.: s.n.], 1826. Citado na pág. 35.

VEHTARI, A.; GELMAN, A.; GABRY, J. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, Springer Science and Business Media LLC, v. 27, n. 5, p. 1413–1432, ago. 2016. ISSN 1573-1375. Disponível em: <<http://dx.doi.org/10.1007/s11222-016-9696-4>>. Citado na pág. 125.

VICENTE-SAEZ, R.; MARTINEZ-FUENTES, C. **Open Science now: A systematic literature review for an integrated definition.**

Journal of Business Research, Science Direct, v. 88, p. 428–436, 2018.

doi:. Citado na pág. 196.

VILLAÇA, F. **Espaço Intraurbano no Brasil.** [S.l.]: Studio Nobel,

Fapesp, Lincoln Institute, 2001. Citado na pág. 45, 59, 61, 62, 63, 64, 90, 114, 133, 136, 162, 180,

181.

VIRTANEN, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, v. 17, p. 261–272, 2020. Citado na

pág. 198.

WASKOM, M. L. seaborn: statistical data visualization. *Journal of Open Source Software*, The Open Journal, v. 6, n. 60, p. 3021, 2021. Disponível

em: <<https://doi.org/10.21105/joss.03021>>. Citado na pág. 198.

ZAP, R. *Conheça os bairros nobres de São Paulo.* 2024. Acessado em: setembro de 2024. Disponível em: <<https://www.zapimoveis.com.br/blog/guia-de-cidades/quais-sao-os-bairros-nobres-de-sao-paulo/>>. Citado na pág. 59.

ANEXOS

ANEXO ÚNICO - ORIGEM E TRATAMENTO DOS DADOS

Resumo do anexo

Este anexo descreve como as bases de dados que fundamentaram o presente trabalho foram construídas. Ele se inicia com uma apresentação geral das tecnologias utilizadas, seguida da listagem das fontes de dados do estudo. Após isso, são detalhados os pipelines de extração e transformação de dados desenvolvidos, assim como os processos de feature engineering realizados para cada uma das variáveis que compõem as bases de dados finais. O anexo é concluído com uma descrição dessas bases.

Objetivo: Permitir a replicabilidade do estudo

A INTRODUÇÃO

Este anexo, de caráter essencialmente metodológico e procedimental, tem por objetivo possibilitar a replicabilidade dos resultados obtidos nesta dissertação. Para que o alcance desse objetivo seja fortalecido, a descrição textual aqui apresentada é acompanhada por uma série de outras medidas relacionadas ao paradigma de ciência aberta a que o trabalho de pesquisa aqui realizado está alinhado. ².

Dentre essas medidas, destacamos três principais. A primeira delas corresponde ao uso exclusivo de dados públicos abertos como fontes de dados da pesquisa. Isso permite que os dados sejam acessíveis sem custos ou necessidade de obtenção de autorização para todo pesquisador e leitor que desejar consultá-los.

A segunda medida consiste na automação de todo o processo de extração, transformação e carregamento dos dados (ETL), assim como da análise exploratória de dados e da modelagem estatística realizada. Essa automação em *código aberto*, acompanhada da descrição contida neste anexo, permite a reprodução dos resultados do estudo. Tudo isso foi realizado com o

²O conceito de *Open Science* (ciência aberta) foi definido por Vicente-Saez e Martinez-Fuentes (2018, p. 7) a partir de uma revisão sistemática da literatura como o conhecimento acessível e transparente que é compartilhado e desenvolvido por meio de redes colaborativas.

uso exclusivo de *softwares livres e abertos* em todas as etapas da pesquisa.

Por fim, a terceira medida consiste na publicação de todo o código aqui mencionado em um repositório aberto, com controle de versões e licenciamento livre ([AGPL v3.0](#)). Esse repositório pode ser acessado no link [<https://github.com/h-pgy/mestrado_uninove>](https://github.com/h-pgy/mestrado_uninove).

O anexo está organizado nas seguintes seções:

- **B Tecnologias utilizadas:** apresenta as tecnologias *open source* utilizadas.
- **C Fontes de dados:** descreve as fontes dos dados analisados no estudo.
- **D Engenharia de atributos:** especifica os procedimentos de *feature engineering* para obtenção das variáveis dos modelos estatísticos.
- **E Descrição das bases de dados:** descreve as versões finais das bases de dados produzidas.

B TECNOLOGIAS UTILIZADAS

Todas as etapas do trabalho de pesquisa foram realizadas com o uso da linguagem de programação livre e aberta *python*. Isso inclui não apenas os *scripts* de [ETL](#), mas também as análises exploratórias, as visualizações de dados e a modelagem estatística. Além disso, a própria escrita do texto final também foi realizada com *software livre*. O Quadro 1, lista as principais bibliotecas *python* utilizadas, com uma breve descrição da finalidade a que foram empregadas neste trabalho. Essa descrição é acompanhada da principal referência acadêmica (em geral, um *white paper* de apresentação da biblioteca), além do link para acesso à documentação técnica.

Como se pode notar pelo quadro Quadro 1, a modelagem estatística foi realizada com o uso da biblioteca CMDStanpy. Trata-se de uma plataforma para modelagem estatística bayesiana, que contém não apenas um amostrador **NUTs!** (HOFFMAN; GELMAN, 2011), como também consiste em uma interface em *python* para a linguagem de programação probabilística STAN (CARPENTER et al., 2017). Essa última permite expressar os modelos estatísticos de forma precisa e reproduzível. Os códigos que representam os modelos em STAN encontram-se na pasta “modelos” do

Quadro 1 – Stack de tecnologias.

Bibliotecas Python	Finalidade	Referência	Link Documentação
Pandas	Manipulação de dados tabulares	(MCKINNEY et al., 2010)	< https://pandas.pydata.org/ >
GeoPandas	Manipulação de dados geoespaciais	(JORDAHL et al., 2020)	< https://geopandas.org/en/stable/ >
Numpy	Computação numérica vetorializada	(HARRIS et al., 2020)	< https://numpy.org/ >
Scikit Learn	Padronização dos dados e análise fatorial	(PEDREGOSA et al., 2011)	< https://scikit-learn.org/stable/ >
Scipy	Testes estatísticos	(VIRTANEN et al., 2020)	< https://scipy.org/ >
LibPySal	Estatística espacial	(REY; ANSELIN, 2007)	< https://pysal.org/libpysal/ >
Matplotlib	Visualização de dados	(HUNTER, 2007)	< https://matplotlib.org/ >
Seaborn	Visualização de dados	(WASKOM, 2021)	< https://seaborn.pydata.org/ >
CMDStanpy	Modelagem estatística bayesiana	(CARPENTER et al., 2017)	< https://mc-stan.org/cmdstanpy/ >
ArViz	Diagnóstico e comparação de modelos bayesianos	(KUMAR et al., 2019)	< https://python.arviz.org/en/stable/ >

Fonte: Elaboração Própria

repositório no *github* que publica o código fonte desenvolvido para a dissertação. Os modelos estatísticos, assim como a metodologia de análise, foram detalhados ao longo da dissertação, nos capítulos que contém as análises empíricas, nomeadamente os capítulos [3 Efetividade do Plano Diretor](#) e [4 Retorno ao Modelo Monocêntrico](#).

Para publicação do código fonte, incluindo comentários explicativos,

utilizamos a ferramenta JupyterLab, que consiste em um ambiente de desenvolvimento e *runtime* para *jupyter notebooks* (KLUYVER et al., 2016). Esse formato é especialmente interessante para a publicação de códigos utilizados no trabalho científico pois possibilita unificar, em um único arquivo, o código fonte desenvolvido com a análise textual. No *jupyter notebooks*, a análise textual pode ser escrita em formato *Markdown*, inclusive com suporte para imagens – o que permite gerar uma documentação do código muito mais robusta, e tornando-o muito mais legível. Além disso, para escrita dessa dissertação, foi utilizada a ferramenta de escrita científica L^AT_EX (LAMPORT; PROJECT, 1994).

C FONTES DE DADOS

A totalidade dos dados utilizados neste trabalho provém de fontes públicas oficiais, livres e abertas. Essas fontes se subdividem em duas principais categorias:

- Dados cadastrais oficiais produzidos pela Prefeitura Municipal de São Paulo ([PMSP](#));
- Microdados provenientes de pesquisas e levantamentos realizados por órgãos públicos brasileiros (nomeadamente, [IBGE](#), [MTE](#) e [Metro-SP](#)).

Abaixo, realizaremos uma descrição geral dessas fontes de dados.

C.1 Dados produzidos pela Prefeitura de São Paulo

Os dados utilizados provenientes da Prefeitura de São Paulo subdividem-se em duas principais categorias:

- microdados não-identificados provenientes do Cadastro Imobiliário Fiscal;
- dados geoespaciais disponibilizados publicamente pela plataforma GeoSampa.

C.1.1 Cadastro Imobiliário Fiscal

O cadastro imobiliário fiscal consiste na base de dados que fundamenta o cálculo do Imposto Territorial e Predial Urbano (**IP****TU**). Trata-se de um dos principais impostos de competência municipal no Brasil e uma importante fonte de receita para o município de São Paulo. Os microdados dessa base foram tornados públicos em 2016 pela Prefeitura de São Paulo. A base foi publicada inicialmente de forma identificada (*i.e.*, com a veiculação dos nomes e CPFs dos proprietários³). Contudo, devido a processos fundamentados na Lei Geral de Proteção de Dados Pessoais (**LGPD**) que objetivavam tornar a base sigilosa novamente, a **PMSP** deliberou por remover as colunas referentes aos nomes e documentos – e, com isso, manter o restante dos dados públicos.

Em 2019, a Prefeitura publicou também os polígonos georreferenciados que compõem a base do **IP****TU** e permitem localizar os microdados no território do município. Esses polígonos correspondem a três diferentes camadas de informação:

- Os setores, que organizam o território do município em regiões. Esses polígonos englobam as quadras e lotes, no sentido geométrico;
- As quadras fiscais, que correspondem, *grosso modo* aos quarteirões da cidade e que, por sua vez, englobam os lotes;
- Os lotes, menor unidade de observação, e que correspondem aos limites dos terrenos dos imóveis cadastrados na base do **IP****TU** ⁴.

Esses três elementos de organização espacial (*setor*, *quadra* e *lote*) podem ser vinculados (*join*) aos microdados do **IP****TU** por meio do código que

³Tecnicamente, trata-se dos contribuintes do **IP****TU**, que podem ser tanto os proprietários, mas também os possuidores e detentores do domínio útil dos imóveis (Prefeitura Municipal de São Paulo, 2014a)

⁴Trata-se dos limites dos terrenos tal como reconhecidos pela autoridade municipal para fins de cobrança de impostos. Ainda que haja bastante equivalência, sobretudo no caso dos imóveis formalizados, esses limites não necessariamente correspondem aos limites definidos em cartório, nas matrículas dos imóveis. Inclusive, o cadastro no **IP****TU** independe do fato do imóvel possuir registro em cartório ou matrícula, bastando haver sobre ele qualquer forma de propriedade ou mesmo posse ou domínio útil (Prefeitura Municipal de São Paulo, 2014a).

identifica os imóveis, denominado “Número do Contribuinte”. Esse código tem o seguinte formato em expressão regular (*Regex*):

`\d{3}\.\d{3}\.\d{4}-\d{1}`

A expressão é equivalente à máscara “999.999.9999-9”. Os três primeiros dígitos correspondem ao código identificador dos setores, os três seguintes ao código identificador das quadras e os quatro finais (à exceção do dígito verificador) correspondem ao lote. Por isso, o “Número do Contribuinte” também é denominado “código Setor-Quadra-Lote”, ou **S-Q-L**, sigla que se consolidou no jargão dos servidores da Prefeitura e arquitetos e urbanistas da cidade⁵.

Não apenas os proprietários formais, com registro em matrícula, são considerados contribuintes do **IPTU**. Em termos jurídicos, esse imposto tem como fato gerador a propriedade, mas também a posse e o domínio útil de um imóvel em área urbana (Prefeitura Municipal de São Paulo, 2014a, art. 1º). Por isso, a cobrança do **IPTU** independe do imóvel ser ou não considerado regular pela Prefeitura. Dessa forma, essa base de dados contempla uma amostra em muito superior a apenas a cidade *formal*, isto é, o universo dos imóveis regulares com titulação de propriedade definida em matrícula. Mas, como será apresentado na subseção **D.1.1**, a representatividade da base é superior para os imóveis nessas condições.

A base cadastral do **IPTU** contém as informações que são disponibilizadas na folha de rosto da cartela do imposto que é recebida anualmente por todos os proprietários ou possuidores de imóveis da cidade de São Paulo. A Figura 1 representa essa folha de rosto.

O acesso à página ⁶ explicativa da cartela do IPTU no site oficial da Secretaria da Fazenda da Prefeitura de São Paulo (**SF**) permite consultar de forma interativa metadados para cada um dos campos da cartela, representada na Figura 1. As informações disponíveis podem ser organizadas em:

- Dados relativos ao contribuinte, como sua idade;
- Dados relativos ao terreno em que se situa o imóvel, como área do

⁵Sim, a sigla é fonte de confusão entre os arquitetos e servidores da área de TI. Aqui optou-se por grafá-la com hífen, “S-Q-L”, para diferenciá-la da sigla da linguagem de *queries* estruturada

⁶<<https://web1.sf.prefeitura.sp.gov.br/CartelaIPTU/>>

Figura 1 – Representação da Cartela do IPTU

Entenda cada campo da cartela do IPTU
 Passe o mouse sobre o campo ou clique no + para mais informações.

CONTRIBUINTE(S)

LOCAL DO IMÓVEL

MENSAGENS

USO **COB** **CIII** **SZU** **TT/EF** **TESTADA** **ÁREA OCUPADA** **ACC**

VENCIMENTOS

ÁREAS (M2) **VALOR M2 (R\$)** **FATORES** **BASE DE CÁLCULO (R\$)** **FATOR**

CONSTRUÍDA X CONSTRUÇÃO X OBSCOLESCÊNCIA = (A)

TERRENO INCORPORADO X TERRENO X PROFUNDIDADE X TIPO TERRENO X CONDOMÍNIO X FRAÇÃO IDEAL = (B)

EXCESSO DE ÁREA X TERRENO X PROFUNDIDADE X TIPO TERRENO X CONDOMÍNIO X FRAÇÃO IDEAL = (C)

BASE DE CÁLCULO (R\$) X ALIQUOTA PREDIAL X DESCONTO/ACRÉSCIMO PREDIAL = IMPOSTOS CALCULADOS PREDIAL R\$

(A)+(B) (R\$) - DESCONTO PREDIAL R\$ = IMPOSTOS CORRIGIDOS PREDIAL R\$

(C) (R\$) X TERRITORIAL = TERRITORIAL R\$

TERRITORIAL R\$ - DEDUÇÕES PREDIAL R\$ = IMPOSTO A PAGAR PREDIAL R\$

TERRITORIAL R\$ - DEDUÇÕES TERRITORIAL R\$ = IMPOSTO A PAGAR TERRITORIAL R\$

TOTAL DO IPTU (D) + (E) **CRÉDITO DA NOTA FISCAL ELETRÔNICA** = **TOTAL A PAGAR (R\$)**

Fonte: Prefeitura Municipal de São Paulo. Disponível em:
<https://web1.sf.prefeitura.sp.gov.br/CartelaIPTU/>. Consultado em 15 de fevereiro de 2024.

terreno, a quantidade de frentes e a testada⁷;

- Dados relativos à edificação, como a área construída, o tipo e padrão de construção, a idade estimada da edificação, entre outros.

A Prefeitura de São Paulo publicou os microdados referentes à base cadastral do IPTU para o período que vai de 1995 a 2023 (mais de um quarto de século, portanto), disponíveis em 29 arquivos .csv. Os arquivos podem ser baixados na seção de *downloads* da plataforma GeoSampa⁸.

⁷A quantidade de frentes se refere à quantidade de lados do polígono que constitui o terreno que são fronteiriços às ruas públicas. P. ex., esquinas possuem duas frentes, enquanto imóveis encravados, acessíveis apenas por servidão de passagem, não possuem frente. Já a testada corresponde à soma dos metros lineares dessas frentes.

⁸Mais sobre a plataforma no item C.1.2

Cada um desses arquivos corresponde à base de dados do IPTU naquele determinado ano.

Mais precisamente, os dados são extraídos da base em 1º de janeiro desse ano. Sua publicação dessa forma, com uma extração para cada ano, deve-se a uma exigência de transparência em relação à base de cálculo que fundamentou a cobrança do IPTU para cada um dos imóveis da cidade naquele ano fiscal. A publicação da base anualizada é uma excelente oportunidade para entender o desenvolvimento imobiliário da cidade, pois o arquivo referente a cada ano representa o estado do estoque construtivo da cidade naquele momento.

A série histórica completa dessa base de dados possui aproximadamente 70 milhões de registros, e os arquivos correspondentes aos anos posteriores a 2010 possuem mais de 3 milhões de linhas cada um. Isto pois verifica-se um aumento progressivo da quantidade de registros ao longo dos anos, que reflete as novas unidades imobiliárias, principalmente apartamentos, produzidos na cidade no período. Como foi argumentado no capítulo 3, é a identificação da localização destas novas unidades por meio da base do IPTU que nos permite avaliar a efetividade do planejamento urbano em direcionar o adensamento construtivo em São Paulo, por meio da identificação das áreas da cidade onde esse adensamento ocorreu. Além disso, a somatória da área construída dessas novas unidades, agrupada por distrito, permite identificar as regiões da cidade que concentraram o investimento imobiliário – o que foi analisado no capítulo 4

O capítulo 3, que tem por objetivo analisar a efetividade do PDE em direcionar o adensamento na cidade, utiliza as bases correspondentes aos anos de 2014 (SF/PMSP, 2014) e de 2023 (SF/PMSP, 2023). Os dados de 2014 registram a situação base, anterior à publicação do PDE. Isto pois a publicação do PDE ocorreu em 31 de julho de 2014, enquanto a base do IPTU de 2014 corresponde à situação em 1º de janeiro desse ano, aproximadamente sete meses antes, portanto. Já os dados de 2023 correspondem aos dados mais recentes disponíveis no momento de escrita da dissertação. Trata-se, além disso, dos últimos dados coletados anteriores à entrada em vigor da revisão do Plano Diretor aprovada em junho de 2023 (Câmara Municipal de São Paulo e Prefeitura Municipal de São Paulo, 2023)⁹.

⁹Como os dados correspondem à situação da base de dados em 1º de janeiro de 2023,

O capítulo 4, que consiste em uma verificação empírica da relação positiva entre acessibilidade ao mercado de trabalho e adensamento baseada em um modelo proveniente da economia urbana conhecido pela sigla [SMM](#), utiliza as bases de dados do [IPTU](#) para os de 2007 (SF/PMSP, 2007) e 2017 (SF/PMSP, 2017). Como foi apresentado nesse capítulo, esse recorte temporal – que infelizmente não coincide com o recorte do capítulo 3 – se deve à disponibilidade de dados da pesquisa [OD](#) do [Metro-SP](#), que é essencial para analisar a acessibilidade ao mercado de trabalho na cidade.

Cada uma das variáveis provenientes da base do [IPTU](#) que foram utilizadas neste trabalho será descrita em maiores detalhes na seção ???. Por ora, cumpre ressaltar que ambas as variáveis dependentes do estudo foram construídas a partir dessa base de dados. São elas a quantidade de novas unidades de tipo de construção vertical produzidas de 2013 a 2023 na cidade de São Paulo (analisada no capítulo 3), assim como a quantidade de metros quadrados de área construída por distrito produzidas nesse período (analisada no capítulo 4).

C.1.2 GeoSampa

A plataforma GeoSampa consiste em um repositório de dados abertos geoespaciais dotado de interface de usuário com mapa interativo mantido pela Secretaria Municipal de Urbanismo e Licenciamento da Prefeitura de São Paulo ([SMUL/PMSP](#)). Esse repositório está aberto à população e pode ser acessado neste link. A Figura 2 representa a interface de usuário da plataforma.

Desenvolvido com a ferramenta open-source GeoServer, o GeoSampa disponibiliza seus dados no formato de camadas espacializadas (também chamadas *layers*). Todas as camadas disponibilizadas provém de fontes públicas e abertas. A grande maioria delas foi desenvolvida a partir dos mapas oficiais e/ou dos registros administrativos georreferenciados produzidos pela Prefeitura de São Paulo. O restante das camadas provém de fontes públicas oficiais de outros entes da Federação, como o [IBGE](#).

A publicação dos dados no GeoSampa requer a autorização da Secreta-

eles não registram qualquer efeito da revisão, que foi aprovada apenas em junho desse ano

Figura 2 – Plataforma GeoSampa

Fonte: Prefeitura Municipal de São Paulo. Disponível em:
<<https://geosampa.prefeitura.sp.gov.br>>. Consultado em 15 de fevereiro de 2024.

ria/Unidade responsável pelos mesmos e passa pela análise técnica e crítica dos especialistas em geoprocessamento que compõem a equipe da Coordenadoria de Análise e Produção da Informação (**GEOINFO/SMUL/PMSP**). Esse controle tem por objetivo garantir a qualidade dos dados, sobretudo do ponto de vista espacial e é realizada também para os dados provenientes de outros entes da federação. Uma das tarefas realizadas é, por exemplo, padronizar a projeção das camadas para a projeção oficial da Plataforma (**SIRGAS 2000 UTM 23-Sul**).

Os quadros 2, 5, 4 e 6 apresentam as camadas do GeoSampa utilizadas nesta pesquisa que têm como fonte dos dados a própria Prefeitura de São Paulo. Já o quadro 3 apresenta as camadas que possuem como fonte de dados outros entes da federação. Em cada um deles, fornecemos uma breve descrição dos dados que constam na camada, de sua finalidade no estudo, assim como sua referência bibliográfica. A URL para acesso aos metadados oficiais da camada está disponível nas referências.

Os quadros 2 e 3 listam as camadas cujos polígonos foram utilizados para georreferenciar as bases de dados do estudo. O quadro 2 apresenta as camadas que foram utilizadas para georreferenciar os microdados do **IPTU**, assim como para georreferenciar os dados coletados a nível de dis-

trito municipal. Por exemplo, o total de área construída de empreendimentos imobiliários verticais na cidade agrupada por distrito que é analisado no capítulo 4, ou os distritos que compõem os *clusters* espaciais que correspondem à área nobre e às áreas periféricas da cidade, identificados pela análise LISA realizada no capítulo 3. As camadas disponíveis no GeoSampa apresentadas no quadro 3 foram empregadas para georreferenciar os dados do Censo IBGE 2010 (utilizados no capítulo 3) e da Pesquisa OD (utilizada nos capítulos 3 e 4).

Quadro 2 – Georreferenciamento – Prefeitura de São Paulo

Identificação da camada	Descrição	Finalidade	Referência
Quadra	Quadras fiscais - correspondem ao “segundo nível de agregação” dos dados do IPTU, conforme apresentado no item C.1.1	Georreferenciar as novas unidades e a área construída produzida na cidade.	(GEOINFO/SMUL/PMSP, SF/PMSP, 2018)
Distrito	96 distritos oficiais do município de São Paulo	Georreferenciar os dados a nível de distrito	(GEOINFO/SMUL, 2018)

Fonte: Elaboração Própria

Quadro 3 – Georreferenciamento – Outros Órgãos Públicos

Identificação da camada	Descrição	Finalidade	Referência
Setor Censitário	Polígonos dos setores censitários do Censo IBGE 2010 C.1.1	Georreferenciar os dados censitários.	(GEOINFO/SMUL/PMSP, IBGE, 2010)
Zonas OD 2017	Polígonos das zonas de ponderação da pesquisa OD 2017	Georreferenciar dados da pesquisa OD 2017	(GEOINFO/SMUL/PMSP, Metrô/SP, 2017)

Fonte: Elaboração Própria

O quadro 4 lista as camadas que foram empregadas para a construção da variável de tratamento analisada no capítulo 3. Essa variável identifica

se determinada quadra da cidade foi definida pelo PDE como estando situada em uma “Zona de Eixo de Estruturação da Transformação Urbana” (ZEU). Como foi apresentado no capítulo 3, trata-se das áreas próximas ao transporte público coletivo da cidade para as quais o Plano Diretor concede incentivos construtivos, que receberam a denominação EETU. Foi incluída também a camada que identifica o polígono da Macrozona de Proteção e Recuperação Ambiental. Trata-se de região da cidade com concentração de áreas verdes que é destinada pelo PDE para preservação ambiental. Os incentivos construtivos estabelecidos para os EETU possuem efeito reduzidos nesta macrozona ¹⁰.

Quadro 4 – *Legislação Urbana – Variáveis Tratamento*

Identificação da camada	Descrição	Finalidade	Referência
Zoneamento 2016	Zoneamento das quadras da cidade	Identificar o incentivo construtivo na quadra.	(GEOINFO/SMUL/PMSP, DEUSO/SMUL/PMSP, 2015)
EETU	Polígonos dos Eixos de Estruturação tal como definidos pelo P.D.E.	Visualizar a distribuição espacial dos incentivos construtivos	(GEOINFO/SMUL/PMSP, 2014a)
MPQA	Polígono da Macrozona de Proteção e Recuperação Ambiental tal como definida pelo PDE	Identificar a região onde os incentivos construtivos são reduzidos	(GEOINFO/SMUL/PMSP, 2014b)

Fonte: Elaboração Própria

Os quadros 5 e 6 listam camadas que correspondem a diferentes restrições à execução de empreendimentos imobiliários. Essas restrições consistem, essencialmente, em restrições ambientais e tombamentos (em outras palavras, à preservação ambiental e cultural). Elas foram incluídas nos modelos estatísticos desenvolvidos nesta dissertação, pois, na medida em que implicam em limites à edificação, são covariantes que afetam negativamente as variáveis dependentes do estudo ¹¹.

¹⁰O Coeficiente de Aproveitamento Básico dos eixos na Macrozona de Preservação Ambiental é 2, e não 4 como na Macrozona de Estruturação e Qualificação Urbana

¹¹Qual seja, a quantidade de novas unidades residenciais verticais, analisada no capí-

O quadro 5 lista as camadas relacionadas ao tombamento e à preservação da paisagem urbana. Essas camadas contêm os dados georreferenciados de imóveis tombados e suas respectivas áreas envoltórias, conforme definidas pelos três principais órgãos de preservação da paisagem urbana hoje atuantes no território da cidade de São Paulo: o **IPHAN** (esfera federal), **CONDEPHAAT** (esfera estadual) e **CONPRESP** (esfera municipal). Além disso, o quadro contempla também a camada referente aos polígonos dos “Bairros Jardim”. Trata-se de bairros planejados, ocupados predominantemente pela elite da cidade de São Paulo, que tiveram seu traçado urbano protegido por tombamentos específicos.

O quadro 6 lista as camadas do GeoSampa que dizem respeito a diferentes instrumentos de preservação ambiental, como Unidade de Conservação (**UC**), Área de Proteção Ambiental (**APA**) e áreas de manancial. Esses instrumentos representam diferentes tipos de restrição à atividade construtiva. Dentre essas restrições, podemos citar, a título de exemplo, a preservação dos exemplares arbóreos presentes nos terrenos, a obrigatoriedade de manter certa proporção de área dos terrenos permeável (para aumentar a drenagem e reduzir enchentes) ou ainda proibições de ocupação das margens dos reservatórios de água da cidade.

Por fim, o quadro 7 apresenta camadas relacionadas a outras restrições construtivas, como restrições geotécnicas referentes ao risco de solapamento do solo ou então limites de altura definidos pelas rampas de pouso e decolagem de helipontos.

Foram consideradas também outras camadas que representam restrições construtivas e normativas, mas que em segunda análise não se mostraram adequadas, seja por falta de precisão de dados, seja pela impossibilidade de identificar regras definidas em relação aos efeitos restritivos sobre a aprovação de alvarás de controle de atividade edilícia. Um dado exemplificativo desta situação é a camada referente a Decreto de Interesse Social (**DIS**) e Decreto de Utilidade Pública (**DUP**) (GEOINFO/SMUL/PMSP, DESAP/SMJ/PMSP, 2023).

Pois, a rigor, ainda que a **DUP** indique a possibilidade de desapropriação do imóvel, da qual decorre a falta de interesse em edificar nele – e

tulo 3, ou de metros quadrados de área construída produzidos na cidade por empreendimentos residenciais verticais, referente ao capítulo 4

Quadro 5 – Restrições relacionadas a Tombamentos

Identificação da camada	Descrição	Finalidade	Referência
Bens Tombados	Polígonos dos imóveis (lotes) tombados	Identificar restrições a empreendimentos imobiliários relacionadas a tombamento.	(GEOINFO/SMUL/PMSP, DPH/SC/PMSP, 2019b)
Área Envolvente - CONDEPHAAT	<i>Buffers</i> das áreas envoltórias de bens tombados pelo CONDEPHAAT	Identificar restrições a empreendimentos imobiliários relacionadas à área envoltória de bens tombados.	(GEOINFO/SMUL/PMSP, DPH/SC/PMSP, 2019c)
Área Envolvente - CONPRESP	<i>Buffers</i> das áreas envoltórias de bens tombados definidos pelo CONPRESP	Identificar restrições a empreendimentos imobiliários relacionadas à área envoltória de bens tombados.	(GEOINFO/SMUL/PMSP, DPH/SC/PMSP, 2019d)
Área Envolvente - IPHAN	<i>Buffers</i> das áreas envoltórias de bens tombados pelo IPHAN	Identificar restrições a empreendimentos imobiliários relacionadas à área envoltória de bens tombados.	(GEOINFO/SMUL/PMSP, DPH/SC/PMSP, 2019e)
Bairros Jardim	Polígonos dos "Bairros Jardim" cujo urbanismo foi tombado.	Identificar restrições a empreendimentos imobiliários relacionadas à preservação da paisagem urbana.	(GEOINFO/SMUL/PMSP, DPH/SC/PMSP, 2019a)

Fonte: Elaboração Própria

por vezes mesmo a vedação para tanto –, ela pode não se efetivar em uma desapropriação propriamente dita. Ou, mesmo quando essa desapropriação océ efetivada, ela pode ocorrer décadas após a emissão do decreto. Neste ínterim, até que seja feita a notificação formal do proprietário, da qual decorre a proibição construtiva, o terreno pode ter sido edificado. Já o [DIS](#), ainda que também possa implicar em desapropriação, pode se tornar na realidade um incentivo à empreendimentos imobiliários, na medida em que

Quadro 6 – Restrições Ambientais

Identificação da camada	Descrição	Finalidade	Referência
Manancial - Alto Juquery	Área de manancial do Alto Juquery	Identificar restrições a empreendimentos imobiliários relacionadas à preservação das áreas de manancial.	(GEOINFO/SMUL/PMSP, CPA/SMA/SP, 2015a)
Manancial - Billings	Área de manancial Billings	Identificar restrições a empreendimentos imobiliários relacionadas à preservação das áreas de manancial.	(GEOINFO/SMUL/PMSP, CPA/SMA/SP, 2019)
Manancial - Guarapiranga	Área de manancial Guarapiranga	Identificar restrições a empreendimentos imobiliários relacionadas à preservação das áreas de manancial.	(GEOINFO/SMUL/PMSP, CPA/SMA/SP, 2015b)
APA	Polígonos das Áreas de Preservação Ambiental [APA] do município	Identificar restrições a empreendimentos imobiliários relacionadas a instrumentos de preservação ambiental.	(GEOINFO/SMUL/PMSP, SVMA/PMSP, 2018)
RPPN	Polígonos das Reservas Particulares do Patrimônio Natural	Identificar restrições a empreendimentos imobiliários relacionadas a instrumentos de preservação ambiental.	(GEOINFO/SMUL/PMSP, CPA/SVMA/PSMP, 2023)
UC	Polígonos das Unidades de Conservação no município	Identificar restrições a empreendimentos imobiliários relacionadas a instrumentos de preservação ambiental.	(GEOINFO/SMUL/PMSP, CPA/SVMA/PMSP, 2023)

Fonte: Elaboração Própria

Quadro 7 – Restrições Construtivas Diversas

Identificação da camada	Descrição	Finalidade	Referência
Heliponto	Polígonos das rampas de pouso e decolagem de helicópteros	Identificar restrições construtivas (de altura).	(GEOINFO/SMUL/PMSP, CASE/SMUL/PMSP, 2020)
Áreas Contaminadas	Polígonos das áreas contaminadas do município	Identificar restrições construtivas.	(GEOINFO/SMUL/PMSP, SVMA/PMSP, 2019)
Restrição Geotécnica	Polígonos das áreas de restrição geotécnica do município (risco de solapamento do solo)	Identificar restrições construtivas.	(GEOINFO/SMUL/PMSP, CASE/SMUL/PMSP, 2016)
Risco Geológico	Polígonos das áreas de risco geológico (risco de deslizamento)	Identificar restrições construtivas.	(GEOINFO/SMUL/PMSP, SMSU/PMSP, 2023)

Fonte: Elaboração Própria

a finalidade deste instrumento é o provimento de habitação de interesse social. Além disso, os dados publicados no GeoSampa são bastante imprecisos: essa camada apresenta o georreferenciamento de todas as plantas que são anexas aos decretos, sem identificar os imóveis que de fato serão objeto de desapropriação e sobrepondo-se também a imóveis que não são objeto da mesma.

Por isso, optamos por não incluir esses dados no modelo. Assim como outros que avaliamos que podem estar relacionados à restrições construtivas, como a presença de gasodutos ou torres de transmissão de energia, mas que não impedem necessariamente a aprovação dos empreendimentos. Pois a aprovação dependerá das características do projeto e – vale ressaltar – muitas vezes também da discricionariedade de avaliação do técnico da Prefeitura que o analisar. Hoyler (2019) descreve o processo de aprovação dos alvarás de execução na Prefeitura de São Paulo, apresentando inclusive alguns dos problemas advindos dessa discricionariedade, que leva, além de outras questões, a uma falta de padronização nas análises.

Com isso, é finalizada a descrição das fontes de dados que têm como

origem a [PMSP](#).

C.2 Dados produzidos por outros órgãos públicos

Os dados produzidos por outros órgãos públicos utilizados neste estudo têm como origem o [IBGE](#), o [Metro-SP](#) e o [MTE](#). Trata-se de pesquisas ou levantamentos produzidos por esses órgãos.

C.2.1 Censo IBGE

O Censo Demográfico é uma pesquisa censitária (ou seja, que compreende a totalidade da população) realizada pelo Instituto Brasileiro de Geografia e Estatística ([IBGE](#)) com periodicidade decenal. Durante a coleta de dados, os recenseadores dirigem-se a todos os domicílios do país e entrevistam seus moradores, registrando informações sobre suas famílias, assim como sobre o domicílio em que residem e seu entorno. Em nosso estudo, utilizamos os microdados desta pesquisa.

Os microdados do Censo fornecem um conjunto bastante detalhado de informações, agregadas a nível de *setor censitário*. Trata-se da menor unidade territorial utilizada pelo [IBGE](#). Os polígonos dos setores são construídos de modo a conter aproximadamente 300 domicílios, o que permite organizar e gerenciar o trabalho dos recenseadores. O [IBGE](#) procura respeitar também limites naturais, como rios (que dificultariam o deslocamento do recenseador), assim como limites administrativos. Nas grandes cidades, como é o caso de São Paulo, os setores são definidos de modo a estarem contidos em apenas um distrito da cidade, o que permite a agregação dos dados de acordo com essa organização territorial mais ampla.

Os microdados são agregados ao nível de setor censitário para evitar a possibilidade de identificação dos domicílios e das pessoas a que os dados correspondem. Os dados fornecidos pelo [IBGE](#) utilizados neste estudo correspondem às seguintes variáveis:

- Quantidade de pessoas brancas residentes no setor
- Quantidade de pessoas residentes no setor
- Quantidade de chefes de família alfabetizados residentes no setor

- Média da renda nominal do chefe de família, considerando apenas os chefes de família que possuem renda
- Quantidade de domicílios particulares permanentes situados no setor

Como descrito na seção [D](#), esses dados foram agregados por distrito e utilizados para construir variáveis que permitem identificar o nível de privi-
légio socioeconômico do distrito, conforme análise apresentada no capítulo [3](#).

A última versão do Censo foi realizada no ano de 2023, 3 anos após o término do período de 10 anos entre um censo e outro. Esses três anos de atraso devem-se às medidas de restrição sanitárias impostas pela pandemia de coronavírus, mas também a decisões político-administrativas que resultaram em insuficiência orçamentária para realização do censo no ano de 2022. No entanto, apesar da versão mais recente do censo ter sido realizada em 2023, em nosso estudo utilizamos os dados referentes à versão de 2010.

O uso dos dados de 2010 não se deve apenas ao fato de que os microdados referentes ao ano de 2023 não foram, no momento de escrita desta dissertação, ainda disponibilizados pelo [IBGE](#). Mas, sobretudo, pois a inclusão dos dados do Censo em nossa pesquisa tem por objetivo identificar características socioeconômicas dos distritos da cidade na situação base anterior à promulgação do [PDE](#)¹². Como o [PDE](#) foi publicado em 2014, os dados censitários mais recentes para descrição da situação base são os de 2010.

C.2.2 Pesquisa Origem e Destino

A pesquisa Pesquisa Origem e Destino ([OD](#)) é realizada pelo Metrô de São Paulo com periodicidade decenal. Ela tem por principal objetivo subsidiar o planejamento da expansão da rede de metrô e trens urbanos a partir da compreensão dos padrões de deslocamentos dos cidadãos da região

¹²Essa medida é necessária para evitar a circularidade na estrutura causal do modelo estatístico. Por exemplo: a verticalização nos [EETU](#) pode acarretar na mudança do padrão socioeconômicos dos moradores da região, de forma que caso variáveis socioeconômicas para o ano de 2023 sejam incluídas no estudo, elas correm o risco de se confundirem com o efeito de nossa variável tratamento. O artigo (CINELLI; FORNEY; PEARL, 2022b) faz uma excelente discussão sobre os riscos de inclusão de variáveis de controle sem a devida consideração da estrutura causal que se busca analisar.

metropolitana de São Paulo. Os dados utilizados neste estudo provém da etapa domiciliar da pesquisa e são referentes aos anos de 2007 e 2017 (data da última pesquisa publicada), que correspondem à aproximadamente 400 mil viagens. Nesta etapa, os pesquisadores do [Metro-SP](#) entrevistam de forma aleatória cidadãos do município de São Paulo em suas casas, coletando informações sobre as viagens, dentro e fora do município, realizadas rotineiramente por essas pessoas.

Além dos pontos de origem e destino dessas viagens (que fornecem o nome da pesquisa) e a duração das mesmas em minutos, os pesquisadores também pedem aos entrevistados para classificar a motivação da viagem em categorias. Essas categorias permitem identificar viagens que correspondem ao deslocamento pendular, isto é, da residência ao local de trabalho e do local de trabalho de volta à residência. Esses dados serão utilizados neste estudo para calcular a acessibilidade ao mercado de trabalho partindo de cada um dos distritos da cidade.

C.2.3 RAIS

A RAIS, mantida pelo Ministério do Trabalho, é um cadastro de preenchimento obrigatório por todos os empregadores do território nacional, sejam eles pessoas físicas, empresas, empreendimentos rurais ou mesmo órgãos da administração direta (MTE, 2024a). Nela, devem ser relacionados todos os colaboradores do estabelecimento, independentemente da relação trabalhista (MTE, 2024b).

Os microdados da base listam, para cada estabelecimento, a quantidade de vínculos empregatícios naquele ano. Além disso, para cidades de grande porte, o distrito do estabelecimento é fornecido. Informação que consta para 97% dos registros de São Paulo nos anos de 2007 e 2017, de modo que é possível identificar a distribuição de postos de trabalho na cidade para esses anos. Ainda que a base compreenda todos os tipos de relação de trabalho, inclusive as informais, vale ressaltar que seu conteúdo é muito mais representativo dos empregos formais, pois para estes há maiores incentivos em manter-se em dia com as obrigações trabalhistas. Ao mesmo tempo, mesmo que não haja impedimentos para o registro dos dados de postos de trabalho informais, cumpre reconhecer que essas empresas po-

dem temer a identificação e consequências jurídicas posteriores por terem se cadastrado na base.

Os microdados anonimizados são fornecidos pelo [MTE](#) por meio de um servidor [FTP](#), no qual constam arquivos para extrações anuais da base. Os dados para acesso a esse servidor podem ser encontrados neste link. Neste estudo, estamos utilizando os dados para os anos de 2007 e 2017. Esta série temporal deve-se à necessidade de cruzamento com os dados da pesquisa [OD](#), cujas duas últimas versões disponibilizadas ao público correspondem a esses anos, conforme afirmado anteriormente ([C.2.2](#)).

D ENGENHARIA DE ATRIBUTOS

Nesta seção, são detalhados os procedimentos de tratamento de dados realizados a partir das bases originais, apresentadas na seção anterior ([E](#)). Os tratamentos de dados são organizados nos seguintes temas principais:

- Variação da densidade construtiva – Subseção [D.1](#)
- Incentivos construtivos – Subseção [D.2](#)
- Fatores associados ao potencial construtivo – Subseção ??
- Restrições à edificação – Subseção [D.4](#)
- Áreas edificáveis – Subseção [D.5](#)
- Acessibilidade ao mercado de trabalho – Subseção [D.7](#)

Como afirmado na seção ([B](#)), todos os procedimentos de tratamento de dados e de engenharia de atributos foram desenvolvidos em *scripts* automatizados reproduzíveis, que estão disponíveis no repositório deste estudo no *github*.

D.1 Densidade construtiva

Churchman (1999) realiza um levantamento bibliográfico sobre o conceito de densidade na literatura sobre planejamento urbano e urbanismo. O autor demonstra que esse conceito, à primeira vista simples, é concebido

de diversas formas. Dentre elas, podemos citar a densidade populacional, o tamanho médio dos lotes, a taxa de moradores por domicílio ou por metro quadrado de área útil domiciliar, entre outros. Mas há também medidas mais subjetivas, como a densidade percebida pelos cidadãos ou a sensação de aglomeração ¹³, cujos efeitos psicossociais foram apontados pelo clássico de Simmel em 1903 (SIMMEL, 2005). Por isso, conforme argumenta Churchman (1999), é necessário fornecer uma definição precisa de qual o entendimento do conceito de densidade que está sendo mobilizado.

Neste estudo, são empregadas duas operacionalizações complementares do conceito de densidade, aqui entendido sobretudo enquanto densidade construtiva:

- A quantidade de unidades produzidas
- A quantidade de metros quadrados de área construída

Em ambos os casos, essas variáveis foram obtidas a partir da base de dados do IPTU, descrita na subseção C.1.1. Isso é possível pois a Prefeitura de São Paulo publicou essa base de dados de forma anualizada, de modo que a comparação das bases de dados referente a um ano anterior com um ano posterior permite identificar os imóveis produzidos no período.

D.1.1 Identificação da produção imobiliária

Como apresentado na subseção C.1.1, a base do IPTU contém um cadastro censitário de todos os imóveis urbanos do município, com maior precisão para os imóveis formais. Essa maior precisão deve-se tanto a incentivos econômicos quanto regulatórios que ensejam a atualização cadastral desses imóveis.

Do ponto de vista regulatório, a maior precisão está relacionada ao fato de que o processo de licenciamento edilício alimenta a base de dados do IPTU. O licenciamento edilício é uma das principais formas pelas quais os municípios brasileiros exercem sua competência constitucional de controle de uso do solo e por meio da qual efetivam os instrumentos de planejamento urbano, fazendo com que o direito urbanístico, fundamentado

¹³Crowding, no original

no interesse coletivo dos habitantes da cidade, prevaleça sobre os interesses individuais dos proprietários de imóveis (PINTO, 2012). Esse controle se dá por duas principais vias: uma via fiscalizatória, que pune infrações à ordem urbana, mas também por uma via autorizatória, que busca avaliar e aprovar previamente intervenções no espaço urbano. Essa última via consiste na principal atividade do licenciamento edilício.

Assim, a realização de construções, reformas ou loteamentos na cidade requer a solicitação prévia de aprovação do projeto pela Prefeitura, mais especificamente pelo corpo técnico da Secretaria Municipal de Urbanismo e Licenciamento ([SMUL](#)). Essa aprovação é registrada por meio da emissão de um documento denominado “alvará”, que pode ser “alvará de execução de reforma”, “alvará de execução de construção nova”, entre outros. Esse processo é regido por extensa legislação de âmbito municipal¹⁴, da qual destacam-se a Lei de Uso, Parcelamento e Ocupação do Solo (Prefeitura Municipal de São Paulo, 2016) e o Código de Obras (Prefeitura Municipal de São Paulo, 2017).

Ao término da obra ou empreendimento, o responsável pela mesma deve notificar a Prefeitura, solicitando um Certificado de Conclusão de Obra, que é também chamado popularmente, para os imóveis residenciais, de “Habite-se” (Prefeitura Municipal de São Paulo, 2016). É a emissão desse documento que consolida o processo de licenciamento, e torna o imóvel, empreendimento ou loteamento efetivamente *regular* aos olhos da Prefeitura de São Paulo. Como condição para emissão do Certificado de Conclusão, é obrigatório que o responsável emita anteriormente uma Declaração Tributária de Conclusão de Obra ([DTCO](#)).

Essa declaração tem por objetivo informar a Secretaria Municipal da Fazenda ([SF](#)) sobre os serviços de construção civil realizados no território do município e ensejar a cobrança do [ISS](#) sobre eles. Como demonstra o manual publicado por [SF](#) (SF/PMSP, 2024), para formalizar a declaração é necessário informar o número identificador do alvará de execução aprovado anteriormente pela [SMUL](#). O sistema de declaração utiliza essa informação para integrar os dados do alvará à sua base de dados – que posteriormente

¹⁴No Brasil, o controle de uso do solo urbano é competência constitucional municipal, assim a legislação será definida por cada município. Há, no entanto, alguma padronização, decorrente do conhecimento acumulado em urbanismo mas também de leis federais, em especial o Estatuto da Cidade (Governo Federal do Brasil, 2001).

alimentará também a base do IPTU.

Edificar no território da cidade sem a expressa autorização da Prefeitura resulta em multas, embargos ou até mesmo na determinação, pelo poder público, da demolição do imóvel (Prefeitura Municipal de São Paulo, 2017, cap. VI, seção I). Dessa forma, empreendimentos que ensejam a construção de edifícios de maior porte, assim como obras e reformas em imóveis localizados em áreas mais valorizadas da cidade, nas quais há maior visibilidade e presença do poder público, em geral solicitam os alvarás, devido aos maiores riscos envolvidos.

Além disso, a emissão do “Habite-se” é condição necessária para a entrega de chaves dos empreendimentos imobiliários verticais¹⁵ pela construção aos novos proprietários, sem a qual a mesma não consegue finalizar a venda do imóvel. Assim, além da questão regulatória, há também um forte incentivo econômico para a formalização desse tipo de empreendimento, que resulta em uma maior acuidade e atualização da base do IPTU para unidades em edifícios verticais.

Esse incentivo econômico ocorre também devido à necessidade de formalização, junto à Prefeitura, dos novos proprietários dos apartamentos e lotes produzidos, sem a qual a cobrança de IPTU recairia sobre a construtora ou sobre a incorporadora. Considerando tratar-se de quantidade significativa de imóveis, essa cobrança pode rapidamente se tornar um prejuízo elevado, sobretudo nos empreendimentos verticais, nos quais a somatória da área construída dos apartamentos e escritórios do empreendimento costuma ser bem alta.

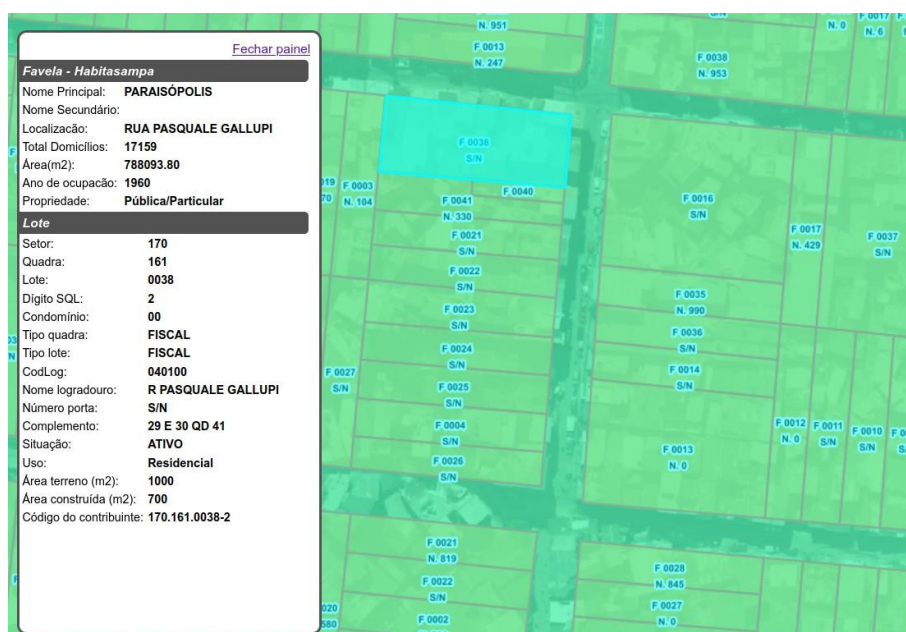
Por essas razões, a base de dados do IPTU é altamente representativa dos imóveis formais da cidade, chegando a quase o universo dos imóveis produzidos por empreendimentos imobiliários verticais. No entanto – cumpre ressaltar – essa base de dados *não se restringe a esses imóveis*. A chamada “cidade informal” ou “irregular” também se faz presente na base do IPTU, sendo alimentada na mesma a partir de uma série de levantamentos e procedimentos fiscalizatórios realizados pela Prefeitura. Nos últimos anos, esses levantamentos tornaram-se muito mais abrangentes a partir da adoção de novas tecnologias.

Anteriormente, a inclusão na base cadastral do IPTU requeria uma

¹⁵Isto é, os “prédios”

fiscalização *in loco* realizada por um auditor fiscal do município ou fiscal de posturas, ou ainda o cadastro do imóvel por parte da equipe de regularização fundiária da Secretaria de Habitação (SEHAB). No entanto, nos últimos 15 anos, a Prefeitura tem usado novas tecnologias, como ortofotos aéreas, imagens de satélite, LIDAR e, mais recentemente, *drones* para atualizar a base de forma remota. Esse investimento na fiscalização – que, vale ressaltar, é motivado também por seus efeitos arrecadatários – resultou em uma ampliação significativa do cadastro de imóveis irregulares e em áreas periféricas.

Figura 3 – Lotes cadastrados na favela de Paraisópolis.

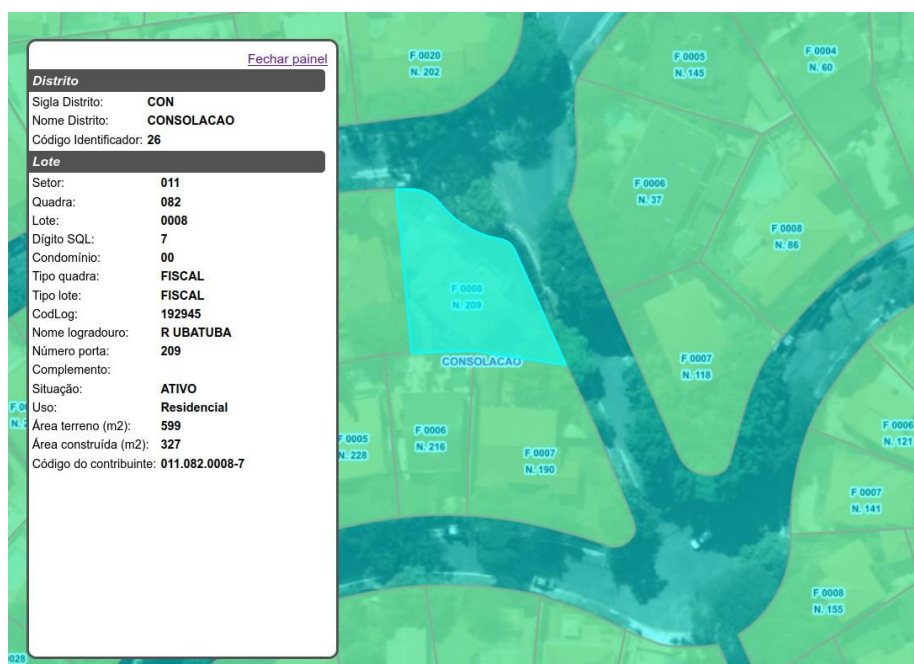


Fonte: GeoSampa/Prefeitura de São Paulo

Como um exemplo, podemos citar o caso da favela de Paraisópolis, que possui lotes cadastrados na base de dados do IPTU, como mostra a Figura 3, que consiste em captura de tela extraída da plataforma GeoSampa. Por suas dimensões, pelo padrão construtivo e pelo valor do metro quadrado do terreno, é possível que boa parte desses imóveis sejam isentos da cobrança de IPTU. No entanto, essa isenção é definida pelo valor venal do imóvel, cujo cálculo é obtido a partir da consulta dessa base de dados, de modo que mesmo os imóveis isentos da cobrança do imposto estão

Como se pode notar na Figura 3, o contorno dos lotes não segue exatamente o contorno das edificações existentes – o que mostra tanto uma menor acuidade dos dados para essas áreas irregulares, como também reflete em parte sua situação fundiária, na qual os terrenos são ocupados em condomínio, em muitos casos por famílias nucleares de descendentes do proprietário original ou com outros vínculos de parentesco e compadrio. A Figura 4, por sua vez, representa um lote cadastrado no Pacaembu, bairro exclusivamente residencial localizado na área nobre da cidade, com alto valor de metro quadrado. A despeito da geometria irregular dos lotes, que seguem o traçado urbano curvilíneo do bairro, o cadastro da Prefeitura apresenta muito maior precisão em relação à ocupação fática dos terrenos, que são ocupados por residências unifamiliares.

Figura 4 – *Lotes cadastrados no Pacaembu.*



Fonte: GeoSampa/Prefeitura de São Paulo

Dessa forma, ainda que não se possa afirmar que a base de dados do IPTU está *restrita* à cidade formal, cumpre reconhecer que ela possui *menor precisão* em relação aos imóveis informais, tanto devido à ausência de incentivos para a atualização da base por parte de seus proprietários, como

também devido à própria natureza da ocupação dos terrenos. Feita essa ressalva, cumpre assinalar também que, devido aos incentivos regulatórios e econômicos mencionados acima, essa base de dados se mostra uma fonte bastante segura para análises que se voltem sobre a densidade construtiva produzida pela produção imobiliária formal, com especial destaque para os empreendimentos verticais, sejam eles residenciais ou comerciais.

A produção imobiliária na cidade pode ser identificada na base de dados do IPTU a partir da comparação das alterações da base de um ano a outro¹⁶. Como apresentado na subseção C.1.1, os imóveis da cidade são identificados na base de dados a partir de seu Número de Contribuinte, também denominado de código Setor-Quadra-Lote (S-Q-L). Esse código possui uma lógica hierarquizada e espacializada, que permite localizar o imóvel nas quadras da cidade.

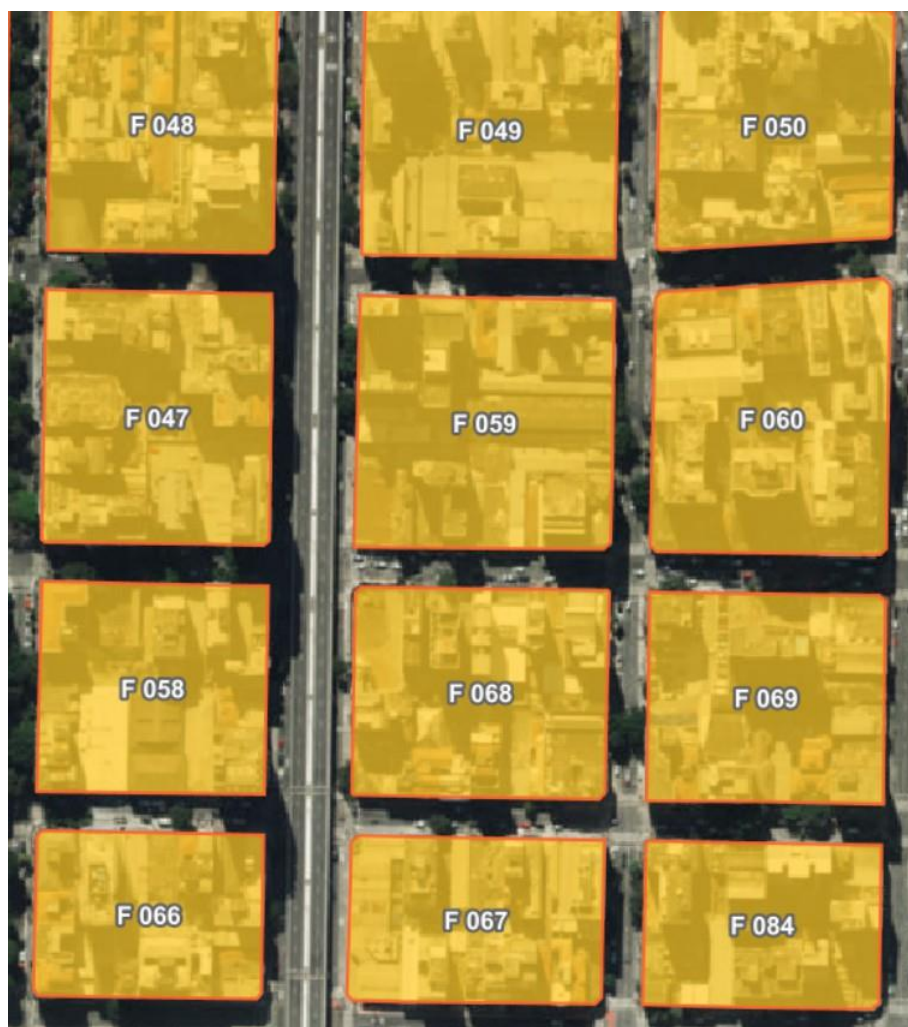
As quadras fiscais correspondem, *grosso modo*, às quadras viárias da cidade (popularmente chamadas de “quarteirões”), excluídas as áreas de calçada. Elas tendem a se manter estáveis ao longo do tempo, pois alterá-las requer a alteração na malha viária da cidade. A Figura 5 contém um exemplo de quadras fiscais – note-se como seus limites são definidos pelo traçado das ruas e calçadas.

Dentre os fatores que implicam na alteração dos quarteirões, o principal seria a aprovação de novos loteamentos na cidade. No entanto, as glebas loteáveis no território do município já se esgotaram há algumas décadas, de modo que praticamente não há mais aprovações de novos arruamentos¹⁷. Outros fatores seriam a regularização fundiária de loteamentos irregulares cujos imóveis ainda não foram cadastrados para efeitos de cobrança de IPTU e a realização de obras viárias que impliquem em desapropriações e na alteração do traçado urbano existente. Trata-se, em ambos os casos, de eventos relativamente raros.

O segundo caso, referente às obras viárias, pode implicar apenas no redesenho do polígono que representa espacialmente a quadra que foi entrecortada pela nova via, mas mantendo-se ainda a identificação original na base de dados. Essa quadra passará assim a ser representada por mais

¹⁶Como descrito na subseção C.1.1, os dados são disponibilizados de forma anualizada, e são sempre referentes à situação da base em 1º de janeiro daquele ano.

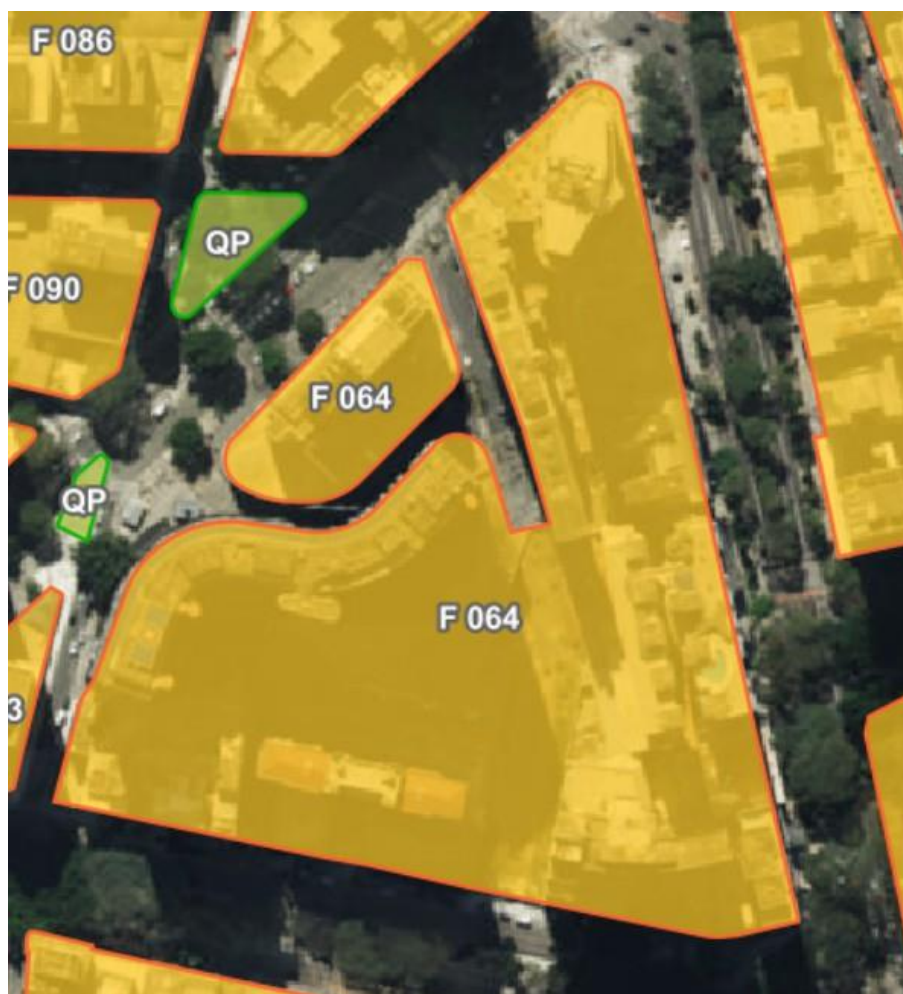
¹⁷Termo técnico para os loteamentos que implicam na construção de estrutura viária e não apenas no desdobro de um lote existente.

Figura 5 – *Correspondência entre quadras fiscais e quarteirões.*

Fonte: GeoSampa/Prefeitura de São Paulo

de um polígono, no entanto sem que se perca a historicidade do registro. Isso pode ser notado, por exemplo, na quadra de número 064 do setor 006, onde se situa o edifício Copan. Como mostra a Figura 6, essa quadra foi subdividida em dois polígonos pela via de acesso ao Copan – no entanto, ambos os polígonos permanecem identificados pela mesma numeração.

Dessa forma, os registros das quadras se mantém bastante estáveis ao longo da série temporal disponível na base do IPTU. Com efeito, menos de 2% das quadras fiscais presentes na base de dados para o ano de 2023 não estavam presentes em 2014, quase uma década antes. Essa permanência dos

Figura 6 – *Quadra com dois polígonos - COPAN.*

Fonte: GeoSampa/Prefeitura de São Paulo

registros das quadras contrasta com os registros dos lotes, que se alteram ao longo do tempo devido aos procedimentos de unificação e desdobro de lotes relacionados aos empreendimentos imobiliários de maior porte, sobretudo os verticais.

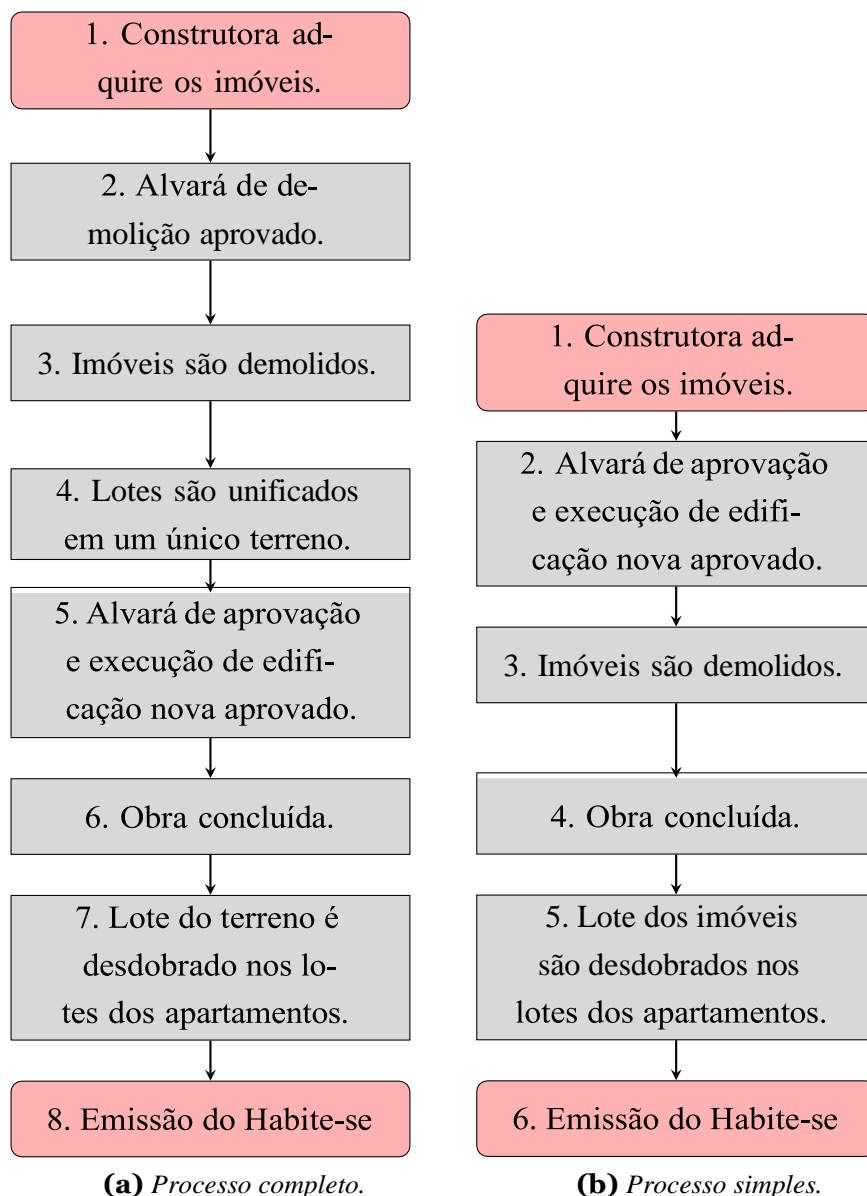


Figura 7 – Unificação e desdobro de lotes em empreendimento vertical.

O diagrama na Figura 7 descreve esse processo. Existem duas principais formas pelas quais ocorre o processo de unificação e desdobro de lotes quando associado a empreendimentos verticais, sejam eles residenciais ou comerciais. A primeira forma, mais comum entre os empreendimentos de grande porte ou alto valor agregado, é representada pela Figura 7a. Nesses casos, os imóveis – em geral várias casas em uma mesma quadra – são adquiridos por uma incorporadora. A incorporadora então solicita um alvará de demolição para esses imóveis e unifica o terreno resultante em um único

lote. Não raramente, esses terrenos são mantidos como reserva de capital especulativo por alguns anos. Nesse período é comum que sejam alugados a empresas que os exploram como estacionamentos. Uma caminhada por bairros em valorização da cidade permite identificar facilmente esses terrenos. A Figura 8 representa um desses casos.

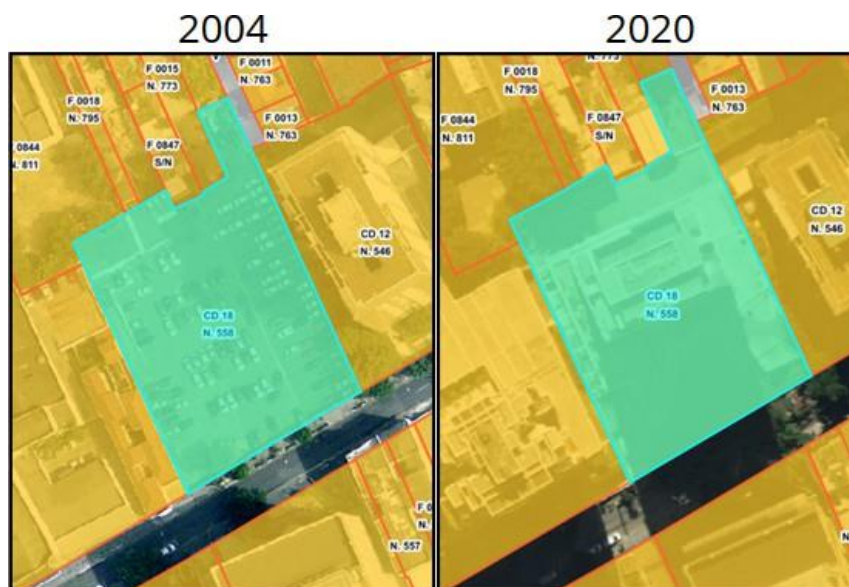
Quando o valor da terra aumenta suficientemente para que a incorporadora aufera os lucros especulativos esperados, o terreno – agora um único lote – é vendido para uma construtora. A construtora protocola então uma solicitação de Alvará de Aprovação e Execução de obra nova¹⁸. Com esse alvará em mãos, é edificado o empreendimento. Após o término da obra, o lote unificado do terreno é então desdobrado nos lotes referentes às unidades produzidas, isto é, os apartamentos, escritórios e/ou pontos comerciais. O desdobro é realizado logo antes da solicitação do Habite-se, que permitirá a entrega do empreendimento aos compradores dessas unidades individualizadas.

O outro processo, que costuma ser mais rápido, é descrito pela Figura 7b. Nesses casos, não há a etapa intermediária de incorporação, na qual os lotes dos imóveis residenciais são demolidos e unificados em um terreno único. Os imóveis são adquiridos diretamente pela construtora, que os demole e os edifica logo em seguida. Esse processo ocorre também quando o empreendimento é edificado no terreno referente a um único imóvel, não havendo portanto necessidade de unificação de lotes. Isso ocorre em bairros que estão passando por mudança no uso do solo predominante industrial para residencial, nos quais há imóveis de grande porte disponíveis, em geral, galpões. De todo modo, o processo é finalizado com o desdobro do lote inicial nos lotes referentes às unidades produzidas, que comporão o condomínio.

Em ambos os casos, nota-se que os registros dos lotes são transformados ao longo do processo de incorporação e edificação. Essas alterações serão identificáveis na base de dados do IPTU por meio da comparação do ano diretamente anterior ao início do empreendimento com o ano logo se-

¹⁸É possível solicitar apenas um Alvará de Aprovação, por meio do qual registra-se a aprovação do projeto em si. Mas, para que se inicie de fato a obra, é necessário solicitar o Alvará de Execução em até dois anos. Essa não é, no entanto, a prática das construtoras, que costumam solicitar ambos os alvarás em um único documento, denominado Alvará de Aprovação e Execução

Figura 8 – Estacionamento transformado em condomínio vertical - r. Frei Caneca (2004-2020).



Fonte: GeoSampa/Prefeitura de São Paulo

guinte à sua finalização. No entanto, na grande maioria dos casos não será possível definir com exatidão quais os lotes originais que foram demolidos para dar lugar aos empreendimentos. Nem mesmo é possível identificar o terreno incorporado, quando houve essa etapa. Isso pois não há relações de chave primária e secundária entre as unidades produzidas e os lotes que lhes deram origem, assim como não há lógica segura de análise dos dados na base do IPTU que permita vincular esses registros diretamente¹⁹. Além disso, os polígonos dos lotes disponibilizados no portal GeoSampa não permitem realizar a análise geográfica a partir da localização do empreendimento, pois é fornecida apenas a situação atual dos lotes.

Entretanto, como a informação da quadra mantém-se estável, é possível comparar, para cada quadra, a quantidade de imóveis nela contida de um ano para outro da série temporal. Isso pois, como mostra a Figura

¹⁹ Isso seria possível, por exemplo, pela manutenção, nos códigos S-Q-L dos apartamentos produzidos, do código do imóvel que lhes deu origem, sendo apenas acrescentados novos dígitos que referenciaríamos o número do apartamento no condomínio. Não é este o caso, no entanto: cada apartamento receberá um número de contribuinte próprio

7, a incorporação e edificação de empreendimentos imobiliários resulta no cadastro de novos lotes na quadra, referentes às unidades produzidas. A comparação entre os lotes presentes na quadra de um ano a outro permite identificar a produção imobiliária, tanto em termos da quantidade de unidades produzida, como também em termos da área construída. Nesse último caso, para obter a informação basta calcular a somatória da área construída dos lotes da quadra e compará-la com a mesma somatória para o ano anterior. Esse método traz ainda a vantagem de permitir georreferenciar a produção imobiliária por meio do cruzamento do código da quadra na base do IPTU com os polígonos das quadras fiscais disponibilizados no GeoSampa (GEOINFO/SMUL/PMSP, SF/PMSP, 2018).

D.1.2 Quantidade de unidades produzidas

O cálculo da quantidade de unidades produzidas por quadra é realizado a partir da base do IPTU de acordo com os seguintes passos:

1. Filtrar os *dataframes* do IPTU para ambos os anos de modo a conter apenas os imóveis com a tipificação buscada (p. ex., imóveis residenciais verticais);
2. Construir um conjunto com os códigos S-Q-L dos imóveis constantes no *dataframe* do ano anterior;
3. Filtrar o *dataframe* do ano posterior de modo a conter apenas os imóveis cujos códigos S-Q-L não estão contidos no conjunto gerado no passo 2;
4. Filtrar os imóveis do *dataframe* gerado no passo 3 de modo a conter apenas os imóveis cuja data de construção é posterior ao ano inicial do estudo;
5. Agrupar o *dataframe* gerado no passo 4 pelo código da quadra (primeiros 6 dígitos do código S-Q-L);
6. Contar a quantidade de registros únicos para a coluna de Número de Contribuinte (ID de cada imóvel) por quadra. O resultado deve ser conforme o Quadro 8.

Quadro 8 – *Unidades por Quadra*

Código da quadra (str)	Quantidade de novas unidades (int)
006.064	322
006.065	251
...	...

Fonte: Elaboração Própria

O filtro na etapa 1 definirá o tipo de unidade que será contabilizado. Ele pode ser realizado também pelo uso do imóvel, independentemente do tipo de construção. Por exemplo, apenas imóveis residenciais, considerando conjuntamente verticais e horizontais. No capítulo 3, que tem por objetivo analisar o adensamento construtivo de tipo vertical, foram selecionadas todas as unidades referentes a imóveis verticais, de uso comercial, residencial ou misto, excluindo-se usos industriais, edifícios de estacionamento, vagas de garagem e estações de radiobase²⁰.

A etapa de número 6 obtém um *dataframe* que possui o código da quadra como índice e uma coluna que contém a quantidade de unidades produzidas por quadra no período entre os anos das bases do IPTU selecionadas, para o tipo de unidade definida na etapa 1 (p. ex., apenas imóveis comerciais).

Os passos 2 e 3 são necessários pois não é possível calcular a quantidade de imóveis produzidos diretamente, a partir da diferença da quantidade de imóveis agrupados por quadra para cada ano. Em outras palavras, subtraindo da quantidade de imóveis do ano posterior a quantidade de imóveis do ano anterior. Como vimos na subseção D.1.1, ainda que seja raro, é possível que os registros das quadras se alterem ao longo do tempo, sobretudo devido à criação de novas quadras por meio de processos de parcelamento do solo. Essas novas quadras, se criadas após o ano inicial do estudo, não estarão presentes nele, de modo que terão registros nulos na coluna do *dataframe* para esse ano.

Uma possível solução para esse problema seria preencher esses valores

²⁰Termo técnico para antenas de rádio e celular.

nulos com zero para prosseguir com o cálculo da diferença entre as colunas. Essa solução é, à primeira vista, adequada à realidade, pois glebas que foram loteadas em geral não possuíam imóveis nelas. No entanto, ela pode gerar uma distorção para loteamentos irregulares que foram regularizados. Nesses casos os imóveis aparecerão no estudo como tendo sido construídos no período, mas na realidade o que ocorreu é que eles foram *cadastrados* na base nesse período.

A metodologia aqui apresentada busca contornar esse desafio. Isso é realizado por meio da identificação dos imóveis cadastrados na base do ano posterior que não estavam presentes no ano anterior²¹. Uma vez identificados os imóveis cadastrados no período, seleciona-se apenas os imóveis cujo ano de construção é igual ou maior ao ano inicial do estudo (etapa 4). Assim, garante-se que não sejam contabilizados como produzidos no período do estudo imóveis que por alguma não estavam cadastrados na base do IPTU no ano inicial, mas que foram construídos em momento anterior ao período do estudo.

O resultado do passo 6 corresponderá à quantidade de unidades correspondentes à tipificação selecionada no passo 1 que foram produzidas na cidade no período. Por fim, pode ser realizado o *merge* dessa base de dados com a camada de quadras fiscais disponível no GeoSampa (ver 2). Esse merge é realizado com base no código da quadra e permite georreferenciar a produção imobiliária identificada no território da cidade.

D.1.3 Quantidade de metros quadrados produzidos

Como mostra o trabalho de Alexander (1993), a quantidade de unidades não é a única medida de densidade possível. Considere-se a construção de um grande edifício de uso especial, como um Centro Educacional Unificado (CEU) ou a sede de um órgão público ou empresa de grande porte. Eles serão cadastrados na base do IPTU sob um único código S-Q-L. Assim, de acordo com a metodologia apresentada na seção D.1.2, esses grandes edifícios contarão como apenas uma nova unidade, a despeito de representarem

²¹No capítulo 3 foi realizado um passo a mais, que teve por objetivo selecionar também os imóveis que estavam presentes no ano anterior mas que tiveram mudança de uso para os usos que são objeto do estudo – medida que busca identificar unidades produzidas por meio de *retrofit*

uma intensificação significativa uso do solo urbano.

A densidade construtiva, medida em metros quadrados de área construída, permite contornar esse problema a partir de uma outra abordagem sobre o adensamento que identifica com maior precisão o investimento de capital imobiliário. Ela permite diferenciar uma pequena casa de classe média baixa de uma grande mansão, assim como permite distinguir dois edifícios com a mesma quantidade de unidades, mas com áreas privativas por unidade totalmente diferentes. Por exemplo, considere-se que o primeiro edifício é um pequeno prédio de *kitnetes*, enquanto o segundo é um edifício de alto padrão com apartamentos do tipo um por andar. No método apresentado na subseção [D.1.2](#) esses edifícios seriam idênticos – mas é evidente que há maior investimento do mercado imobiliário no edifício de alto padrão. Esse edifício terá uma área construída total muito maior do que o prédio de *kitnetes*.

O cálculo da densidade construtiva em termos de metros quadrados construídos é muito similar ao apresentado na subseção [D.1.2](#), contendo etapas idênticas. Não obstante, para garantir a clareza e a reprodutibilidade, o método é descrito de forma completa:

1. Filtrar os *dataframes* do [IPTU](#) para ambos os anos de modo a conter apenas os imóveis com a tipificação buscada (p. ex., imóveis residenciais verticais);
2. Construir um conjunto com os códigos [S-Q-L](#) dos imóveis constantes no *dataframe* do ano anterior;
3. Filtrar o *dataframe* do ano posterior de modo a conter apenas os imóveis cujos códigos [S-Q-L](#) não estão contidos no conjunto gerado no passo 2;
4. Filtrar os imóveis do *dataframe* gerado no passo 3 de modo a conter apenas os imóveis cuja data de construção é posterior ao ano inicial do estudo;
5. Agrupar o *dataframe* gerado no passo 4 pelo código da quadra (primeiros 6 dígitos do código [S-Q-L](#));

6. Somar a área construída dos imóveis por quadra. O resultado deve ser conforme o Quadro 9.

Quadro 9 – Área construída por Quadra

Código da quadra (str)	Área construída (float)
006.064	1322
006.065	4251
...	...

Fonte: Elaboração Própria

A metodologia acima descrita permite identificar onde ocorreu a produção imobiliária na cidade, de forma mais sensível à intensidade de investimento do capital imobiliário do que a contagem de unidades produzidas apresentada na seção D.1.2. Nessa dissertação, essa metodologia é utilizada no Capítulo 4, no qual é analisada a distribuição espacial da área construída de tipo residencial vertical produzida na cidade de São Paulo no período 2007-2017. Assim como no caso da metodologia apresentada na subseção D.1.2, os dados foram georreferenciados. Inicialmente a nível de quadra, mas em seguida a nível de distrito, a partir da identificação do distrito onde estão situadas as quadras. Isso permitiu o cruzamento com outras bases de dados que não possuem uma granularidade espacial tão detalhada, como é o caso da Pesquisa Origem e Destino (ver subseção C.2.2) e da RAIS (subseção C.2.3).

D.2 Incentivos Construtivos

O Código de Obras do município de São Paulo estabelece em seu artigo 2º que a análise dos pedidos de autorização para atividade edilícia²² deve ser realizada de forma a observar as normas do Plano Diretor Estratégico (Prefeitura Municipal de São Paulo, 2017). Uma das principais medidas do PDE que está diretamente relacionada ao adensamento construtivo da

²²No contexto dessa dissertação, a atividade edilícia pode ser compreendida como a construção e a reforma de imóveis. Mas ela envolve outras atividades, como a alteração de uso e o parcelamento do solo.

cidade é a definição do Coeficiente de Aproveitamento (**CA**) máximo em diferentes áreas da cidade.

Trata-se de um indicador calculado a partir da proporção entre a área construída do imóvel e a área do terreno no qual ele se situa, bastante similar ao **FAR** (do inglês, *Floor Area Ratio*) utilizado nos Estados Unidos. Ele é calculado de acordo com a seguinte equação: $\frac{ac}{at}$, onde *ac* representa a área construída e *at* representa a área de terreno, ambos em metros quadrados. Há no entanto um detalhe relacionado à medição da área construída. Isso pois a **LPUOS** estabelece em seu artigo 62 que as áreas ocupadas por determinados usos não são computáveis nessa fórmula (Prefeitura Municipal de São Paulo, 2016).

No caso dos imóveis residenciais horizontais isso tem poucos efeitos práticos, pois a maior parte da área construída nesses imóveis é utilizada de fato para habitação e é computável. Além disso, caso sejam respeitados os recuos e afastamentos estabelecidos pela lei, dificilmente imóveis com essa tipificação (isto é, casas unifamiliares de até 2 andares) ultrapassarão o **CA** máximo²³. No entanto isso faz toda a diferença para edifícios, sejam eles residenciais verticais ou comerciais. Pois a **LPUOS** estabelece como não computáveis as áreas de circulação de pessoas e veículos, assim como outras áreas técnicas, como as casas de máquinas de elevadores ou as guaritas de entrada. Dessa forma, a área computável total dos edifícios tende a se aproximar da somatória da área útil das unidades (apartamentos ou escritórios) que o compõem.

O Quadro 3 do **PDE**, que pode ser acessado neste link, estabelece o Coeficiente de Aproveitamento básico 1 para todas as zonas da cidade, com exceção das **ZEPAM**, nas quais o **CA** básico é reduzido para garantir sua função de preservação ambiental. Isso significa que todos os imóveis da cidade, exceto os localizados em **ZEPAM**, podem edificar o equivalente à área de seu terreno em área computável, independentemente da zona em que estão situados.

No entanto, como há outras regras definidas na legislação cujos efeitos acabam resultando na impossibilidade de se edificar acima do limite do

²³Com a exceção notável de imóvel de alto padrão com grandes terrenos, como aqueles encontrados na região dos Jardins em São Paulo. Nesses casos, é comum que os arquitetos desenvolvam projetos que se valem das áreas não computáveis para conseguir uma área construída maior.

CA básico, mesmo ele sendo de valor apenas 1. Dentre essas regras, a principal é a que limita o gabarito (isto é, a altura máxima) da edificação, que em algumas zonas é de no máximo 15 metros. Mas outras regras, como aquelas que estabelecem recuos laterais e de fundos, além dos afastamentos frontais, impedem que o **CA** básico seja atingido na prática. Essa situação ocorre sobretudo em áreas na cidade cujos loteamentos originais possuíam lotes de pequena área de terreno (p. ex., 150 metros quadrados) ou então com pequeno perímetro de frente para a rua. Nesses casos, mesmo que o zoneamento permita o adensamento, é necessário adquirir mais de um lote e unificá-los, conforme descrito na Figura 7a, reservando parte da área do terreno unificado para os recuos e afastamentos determinados em lei.

Caso o projeto ultrapasse o **CA** básico, é facultado ao empreendedor, mediante pagamento em pecúnia ao Fundo Municipal de Desenvolvimento Urbano (**FUNDURB**), adquirir mais potencial construtivo. Trata-se do instrumento denominado Outorga Onerosa do Direito de Construir (**OODC**), cujos valores por metro quadrado estão definidos no Quadro 14 do **PDE** (que pode ser acessado nesse link). O potencial construtivo adquirido por meio da **OODC** não pode ultrapassar o **CA** máximo da zona em que está situado o empreendimento (também definido no Quadro 3 do **PDE**).

Como a consulta ao quadro 3 do **PDE** permite verificar, há diversas zonas cujo **CA** máximo é idêntico ao **CA** mínimo, de modo que não é possível adquirir potencial construtivo por meio de **OODC**. Nestes casos, destacam-se as **ZER** e as **ZEPAM**. Mas há também zonas em que o **CA** máximo é bastante elevado, chegando a 4, de modo que é possível adquirir por meio de **OODC** quatro vezes mais potencial construtivo do que o originalmente conferido. Trata-se das **ZEU** e algumas das **ZEIS**. Nesses casos, há também o estabelecimento de um **CA** mínimo obrigatório de 0,5 vezes a área do terreno – ou seja, novos empreendimentos deverão ocupar, com área construída computável, no mínimo essa proporção. Por fim, há também os casos intermediários, como os das **ZM** e **ZC**, com **CA** máximo 2.

Conclui-se assim que o **CA** é uma das principais formas pelas quais o planejamento urbano municipal, por meio do **PDE** e da **LPUOS**, definem os parâmetros de adensamento construtivo da cidade. Mas há ainda uma série de instrumentos constantes na legislação que estão relacionadas ao cômputo da área construída, dentre os quais destacam-se a fachada ativa

e a doação de áreas de calçada. Esses instrumentos buscam incentivar a fruição pública dos terreos dos edifícios, por meio do estabelecimento de pontos comerciais nos mesmos (fachada ativa) e da criação de áreas públicas (doação de calçadas). Além disso, para incentivar a criação de Habitação de Interesse Social, há a previsão de que a área de apartamentos que se enquadram nesse conceito não seja considerada na área computável do CA, até um certo limite²⁴.

Todas esses instrumentos se tornam incentivos construtivos na medida em que atuam por meio da redução da área computável. Em um primeiro momento, concluiria-se que eles têm objetivo fazer com que a realização de projetos de edifícios mais democráticos seja mais vantajosa financeiramente aos construtores por meio da redução dos custos com aquisição de OODC. Contudo, na prática, quando associados à ausência de gabarito máximo da edificação em algumas zonas, eles permitem, caso utilizados em conjunto e de forma estratégica pelo autor do projeto, a edificação de um CA bastante superior a 4.

Dentre as zonas da cidade mais beneficiadas por esse tipo de incentivos está a Zona de Estruturação Urbana (ZEU), que são definidas na LPUOS como porções do território nas quais se objetiva promover o densidades demográfica e construtiva altas (Prefeitura Municipal de São Paulo, 2016, art. 7º). Essas zonas buscam implementar os Eixos de Estruturação da Transformação Urbana (EETU), áreas próximas aos terminais de transporte público da cidade para as quais o PDE busca direcionar a produção imobiliária. Trata-se de uma medida inspirada no Desenvolvimento Orientado ao Transporte Público(do inglês, *Transit Oriented Development*) (TOD), baseada no conceito, inicialmente proposto por Calthorpe (1993) de que o adensamento próximo às estações de metrô, trem e terminais de ônibus promoveria uma redução da dependência do automóvel e do espraiamento das cidades (BALBIM; KRAUSE, 2016).

Neste estudo, identificamos as quadras da cidade que estão localizadas em ZEU da seguinte forma:

- O zoneamento de cada quadra foi obtido a partir do cruzamento geográfico entre as camadas de quadra fiscal e zoneamento (ver Quadros

²⁴Essa medida tem sido alvo de polêmicas recentes devido à falta de fiscalização em relação à destinação dessas unidades.

2 e 4):

- Foi calculada a intersecção de cada quadra com os polígonos das zonas;
 - Para as quadras que se intersectam com apenas uma zona, a zona intersectada foi definida como a zona da quadra;
 - Para as quadras que se intersectam com mais de um polígono de zona²⁵, a zona da quadra foi definida como a zona do polígono da camada de zonas com a maior área intersectada.
- As quadras cujo zoneamento é igual a “ZEU” foram identificadas como situadas em Zonas de Estruturação Urbana ZEU

O passo 2 resultou na não-consideração como ZEU das quadras situadas em zonas do tipo “ZEUa” e “ZEUp”. Trata-se, no primeiro caso, de ZEUs localizadas em área de proteção ambiental, nas quais há limites para o adensamento. O segundo caso consiste em ZEU previstas, pois estão localizadas no entorno de estação de metrô, trem ou terminal de ônibus em planejamento, cujos incentivos construtivos só são válidos após a inauguração das mesmas e mediante publicação de decreto. Esses dois tipos de ZEU não foram considerados na análise realizada no Capítulo 3 pois nestes casos os incentivos construtivos – isto é, o *tratamento* cujos efeitos o capítulo tem por objetivo avaliar – ou são bastante reduzidos (caso das ZEUa) ou não se fazem presentes (caso da ZEUp).

D.3 Fatores associados à verticalização

Não obstante a existência de incentivos construtivos, o potencial de verticalização de uma dada quadra da cidade é determinado também por outros fatores relativos à situação de urbanização da mesma. Por exemplo: quadras com lotes muito pequenos ou que não possuem terrenos livres são

²⁵Trata-se de algo raro, pois a LPUOS buscou definir apenas uma zona por quadra viária, de modo a resolver uma dificuldade do zoneamento anterior, que permitia que uma mesma quadra tivesse mais de uma zona, abrindo espaço para discricionariedade e questionamentos por parte das construtoras. No entanto, como as quadras podem ser alteradas e como há pequenas diferenças de desenho geográfico entre os polígonos das zonas e das quadras fiscais, há alguns poucos casos de quadras intersectadas por mais de uma zona

mais difíceis de incorporar do que quadras que possuem grandes lotes de galpão ou terrenos baldios. Além disso, há que se considerar que a própria área das quadras varia, de forma que, todos os outros fatores mantidos iguais, quadras maiores terão mais área edificável. Neste estudo, foram consideradas as seguintes variáveis relacionadas aos imóveis:

- O custo de incorporação, medido pelo valor do metro quadrado do terreno;
- A depreciação das edificações, medida pela idade da construção;
- A idade do proprietário;
- A quantidade de pavimentos do imóvel;
- A área edificada, medida em metros quadrados;
- A área de terrenos disponível, também medida em metros quadrados;

Todas essas variáveis foram obtidas diretamente dos dados do IPTU (ver subseção C.1.1). Elas são utilizadas como controles estatísticos na análise desenvolvida no Capítulo 3. Como o capítulo tem como unidade de análise as quadras, os dados dos imóveis foram agrupados por quadra. Foram feitos os seguintes cálculos, em relação aos imóveis situados em cada uma das quadras da cidade:

- Valor médio do metro quadrado de terreno;
- Idade média dos imóveis, calculada a partir do ano de construção;
- Média da idade dos proprietários, calculada a partir de sua data de nascimento;
- A média da quantidade de pavimentos dos imóveis;

Além disso, foram calculadas as seguintes proporções de áreas em relação à área total da quadra:

- A proporção da área total edificada de imóveis de tipo horizontal;
- A proporção da área total edificada de imóveis de tipo vertical;

- A proporção da área total dos terrenos ocupados por galpões e usos similares;
- A proporção da área do maior terreno situado na quadra não-verticalizado;

A área da quadra em si também foi acrescentada, pela razão mencionada acima (que quadras maiores, tudo o mais mantido igual, possuem maior área verticalizável). A seguir, são pontuadas algumas observações sobre essas variáveis.

Em relação ao valor médio do metro quadrado de terreno, cumpre ressaltar que essa variável foi escolhida em detrimento do valor venal por duas principais razões:

- *O valor venal corresponde ao valor de mercado:* Trata-se da base de cálculo de um imposto (o IPTU) e não de uma estimativa do valor de mercado do imóvel, havendo em muitos casos grande disparidade entre o valor venal e o valor de venda. Isso pois, apesar de nas últimas décadas a SF ter passado a empregar técnicas mais modernas, orientadas a dados, para atualização da Planta Genérica de Valores, a fórmula de cálculo do valor venal foi estabelecida em 1986 (PMSP, 1986). Além do valor do metro quadrado do terreno, que é baseado na Planta Genérica de Valores, essa fórmula contém uma série de fatores de desconto pré-estabelecidos em lei, relacionados, por exemplo, ao tamanho da frente do imóvel em relação à rua, que não necessariamente seguem a lógica do mercado;
- *O custo da edificação existente é marginal em relação ao valor do solo criado:* um componente importante no cálculo do valor venal é o valor da edificação em si (pois o imposto predial incide sobre a mesma). No entanto, em empreendimentos verticais, o custo da edificação existente é marginal em relação ao potencial de retorno do solo criado (isto é, os novos andares construídos). Ainda que haja um prêmio relativo à qualidade da edificação e padrão do empreendimento, o valor de venda das unidades será determinado sobretudo pelo valor médio do metro quadrado na localidade em que elas se situam. Como mencionado, na base do IPTU o valor do metro quadrado de terreno é obtido a partir da Planta Genérica de Valores, que

nas últimas décadas tem sido atualizada por meio de estimativas a partir dos valores de transações dos imóveis no território da cidade, sendo portanto um bom indicador do potencial de venda do metro quadrado de solo criado.

Em relação à medida da depreciação dos imóveis a partir do ano de construção dos mesmos pode-se objetivar que a realização de reformas pode alterar esse quadro. Com efeito, dois imóveis construídos no mesmo ano podem ter fatores de depreciação muito distintos caso apenas um deles tenha sido reformado. No entanto, vale ressaltar que reformas significativas ensejam a alteração dessa variável na base de dados do IPTU, de modo que passará a constar na base o ano de conclusão da reforma, e não da edificação original do imóvel (Prefeitura Municipal de São Paulo, 2014a). Dessa forma, a variável gerada para o estudo – a média da idade dos imóveis da quadra – reflete também ciclos de valorização que porventura tenham ocorrido e que resultaram em investimentos nos imóveis existentes por meio de reformas.

Em relação à média da idade dos proprietários, essa variável foi incluída no estudo pois o processo de urbanização da cidade não está desvinculado de processos demográficos mais amplos. Assim, loteamentos realizados há algumas décadas – boa parte deles nos anos 1970-1980–, que foram inicialmente adquiridos por jovens adultos com famílias recém-formadas, encontram-se hoje ocupados por uma população idosa, cujos filhos já deixaram o lar. Essa população com idade mais avançada possui maior propensão a vender seus imóveis em busca de soluções de moradia mais adequadas às suas necessidades atuais.

A média da quantidade de pavimentos dos imóveis, assim como a proporção de área construída de tipo vertical buscam mensurar o grau de saturação da quadra. Com efeito, quadras já muito verticalizadas possuem menor probabilidade de verticalização devido aos altos custos de aquisição e demolição dos imóveis verticais existentes, além da dificuldade de incorporação (pois são muitos proprietários para negociar). A proporção da área total edificada horizontal busca também identificar dificuldades de incorporação advindas de loteamentos com lotes muito pequenos, que em geral tem maior proporção de área edificada em relação à área de terreno, como é o caso de imóveis residenciais geminados. Note-se no entanto que as áreas

de ocupação irregular em geral possuem apenas edificações horizontais e têm alta densidade construtiva – de modo que a alta densidade construtiva horizontal pode indicar áreas pouco valorizadas ou com problemas fundiários.

As demais variáveis, nomeadamente a proporção da área da quadra ocupada por lotes de galpões e a proporção que representa a área do maior terreno não verticalizado buscam identificar facilidades de incorporação e menores custos de demolição. Essa última variável – a proporção da área do maior terreno não verticalizado – foi calculada da seguinte forma:

- Para cada quadra, foram selecionados apenas os imóveis cujo tipo de construção não é vertical;
- Dentre esses imóveis, foi identificado aquele com o maior terreno;
- A proporção da área de terreno desse imóvel foi calculada em relação à área da quadra

Essa variável permite identificar, por exemplo, quadras que ainda possuem casas antigas, com grandes quintais, que têm um menor custo de demolição por metro quadrado de terreno em relação à mesma área quando ocupada de forma mais densa, por casas geminadas, por exemplo. Além de em muitos casos representar maior facilidade de incorporação, por ser necessário negociar com menor quantidade de proprietários, sendo possível inclusive edificar o empreendimento apenas no terreno dessa casa.

D.4 Restrições à edificação

No estudo dos fatores que levam ao adensamento e à verticalização na cidade, há no entanto que se considerar a miríade de dispositivos jurídicos, advindos sobretudo da legislação ambiental e de preservação cultural, que resultam em restrições à edificação. Essas restrições podem ter como resultado que um imóvel com alto potencial de verticalização não tenha esse potencial efetivado. Com efeito, mesmo um imóvel situado em uma área com alto interesse do mercado imobiliário, em uma zona com potencial construtivo elevado, localizado em uma região com alto valor de metro quadrado de terreno e que possui facilidade de incorporação e baixos custos de

demolição pode não ser adensado devido a esses impedimentos. Isso pois, caso ele se enquadre em alguma das restrições à edificação vigentes, pode ser que seja impossível ter um projeto aprovado pela Prefeitura, ou ainda que os projetos aprováveis sejam inviáveis do ponto de vista econômico, de modo que o imóvel sequer seja considerado pelo mercado imobiliário.

As fontes de dados utilizadas para obtenção das informações relativas a essas restrições está listada nos Quadros 5, relativo à preservação cultural, 6, relativo à preservação ambiental e 7, relativo a outras restrições. As restrições de ordem cultural são:

- O tombamento do imóvel por suas características históricas, arquitetônicas e/ou relevância cultural, de modo que essas devem ser preservadas;
- O imóvel estar situado em área envoltória de bem tombado – situação que implica restrições à verticalização e/ou a manutenção de características arquitetônicas de modo a não descaracterizar a localidade;
- Outros tombamentos, como é o caso do tombamento do traçado urbano dos lotes nos denominados “Bairros Ambientais”

Em relação aos tombamentos, foram selecionados apenas os imóveis de fato tombados, e não em estudo ou ainda em abertura de processo de tombamento. Não foi realizado, no entanto, o filtro para garantir que os imóveis tombados tenham tido a decisão emitida antes do período do estudo. Isto se deu pelo entendimento de que os imóveis que por ventura estavam em estudo no período de 2014 a 2023 e que não foram de fato tombados ou foram edificados ou não foram considerados significativos, de modo que eventual restrição à edificação não teve efeitos. Assim como pode-se presumir que os imóveis tombados durante o período mantiveram suas características originais e, portanto, não foram edificados. Outras formas de restrições, como a presença de sítio arqueológico ou de acervo significativo, não foram consideradas. Isso pois ainda que esses casos por vezes requeiram a contratação de pessoal técnico especializado para estudo e preservação dos artefatos encontrados, eles não necessariamente impedem a edificação (vide o caso de diversas estações de metrô).

As restrições de cunho ambiental são:

- A localização em área de manancial;
- A localização em [APA](#);
- A presença de uma [RPPN](#) no imóvel;
- A localização em [UC](#);

Em relação às áreas de proteção aos mananciais, foi realizada a análise para garantir que tenham sido selecionadas apenas as áreas nas quais há restrições à ocupação, que correspondem a 65% da área total dos polígonos de áreas de proteção aos mananciais no GeoSampa. Nomeadamente, trata-se das áreas classificadas como:

- Área de Proteção Permanente (Faixa 50 metros);
- Área de Restrição de Ocupação;
- Sub-área de Conservação Ambiental;
- Sub-área de Ocupação de Baixa;
- Sub-área de Baixa Densidade;
- Sub-área Envoltória da Represa;
- Baixa Densidade 1;
- Baixa Densidade 2;
- Baixa Densidade 3.

Note-se que não foram consideradas a localização do imóvel em parques municipais, estaduais ou federais ou mesmo em áreas verdes livres, como praças, pois estamos trabalhando apenas com as Quadras Fiscais, ou seja, aquelas que têm como proprietários ou possuidores os particulares. Parques, praças e áreas verdes livres são áreas públicas e estão devidamente classificados assim na base do [IPTU](#). A principal razão para isso é que não é realizada a cobrança desse imposto sobre esse tipo de área, denominadas "de uso comum do povo".

As demais restrições urbanísticas consideradas no estudo são:

- A localização em rampa de heliponto;
- A localização em área contaminada;
- A localização em área com restrição geotécnica, casos nos quais as características do solo não permitem a verticalização sob risco de desabamento;
- A localização em área de risco geológico;

Em relação às áreas contaminadas, foram selecionadas apenas aquelas que não tenham sido reabilitadas – filtro que corresponde a 65%²⁶ dos polígonos disponíveis na camada. Isso pois as áreas que foram reabilitadas anteriormente ao período de estudo já poderiam ter sido edificadas, assim como as áreas habilitadas durante o período de estudo indicam que houve um esforço por parte de construtoras em descontaminar a área para edificação. Quanto às rampas de heliponto, trata-se das rampas de pouso e decolagem de helicópteros na cidade, que estão associadas a restrições de altura máxima dos imóveis impostas pelo COMAER.

Por fim, tanto as áreas de restrição geotécnica quanto as áreas de risco geológico implicam em risco de desabamento e portanto não comportam a verticalização. Nas áreas de restrição geotécnica o solo não é adequado para as fundações de edifícios verticais. Já as áreas de risco geológico são propensas a deslizamentos de terra e solapamento do solo.

Cumprе ressaltar que todos os fatores elencados nessa seção são analisados pelos arquitetos da Prefeitura e implicam no indeferimento dos projetos ou então na emissão de pareceres negativos solicitando a realização de modificações para adequar o projeto (por exemplo para preservar a fachada de um imóvel tombado). Hoyler (2019) argumenta como a criação de um cadastro digital no qual boa parte dessas informações podem ser consultadas pelos técnicos da Prefeitura, denominado **BDT**, gerou maior padronização e previsibilidade nas análises de projetos, fazendo com que as restrições aqui listadas tenham maior efetividade.

Por fim, como vimos na subseção **D.1.1**, o próprio zoneamento pode implicar em restrições a verticalização, em especial as **ZER**, zonas nas quais

²⁶O mesmo valor para as áreas de proteção aos mananciais – trata-se de mera coincidência.

só se permite a edificação de residências unifamiliares, com altura máxima de 10 metros em relação ao piso térreo. Ainda que essa altura comporte até 3 andares, muitas vezes isso não é possível de ser alcançado, devido à imposição de **CA** máximo 1 assim como à análise de adequação do tipo de uso do imóvel ao zoneamento (residências em condomínio não são permitidas nessas zonas).

D.5 Áreas edificáveis

A identificação das restrições à edificação permite calcular as áreas edificáveis por distrito, por meio da subtração das mesmas em relação à área total do distrito. Esse dado é utilizado no Capítulo 4. Nesse caso, como o nível de agregação do capítulo é o distrito – e não mais as quadras como na subseção D.4 – foram consideradas como restritas também as áreas públicas. Ao todo, foram considerados como não-edificáveis os polígonos presentes no GeoSampa referentes aos seguintes fatores:

- Quadras não-fiscais: isso é, quadras de titularidade pública ou de uso comum do povo, em área rural²⁷ além de outros casos, como os canteiros centrais e servidões de passagem;
- Imóveis tombados;
- Bairros ambientais²⁸;
- Áreas contaminadas;
- Áreas de risco geológico;
- Áreas de restrição geotécnica;
- Remanescentes de mata atlântica;
- Terras indígenas;
- Parques e Unidades de Conservação;

²⁷Nesse caso a exclusão não diz respeito diretamente a um impedimento à verticalização, mas sim pois o escopo do estudo é a área urbana.

²⁸Tipo específico de tombamento do traçado urbano que impede a unificação de lotes

- Zoneamento do tipo ZER

Note-se que muitos dos polígonos dessas camadas possuem sobreposição entre si. O caso mais evidente são as áreas de remanescentes de Mata Atlântica e os limites das Unidades de Conservação da cidade. Mas isso ocorre para outras camadas também, como é o caso do zoneamento do tipo ZER e os tombamentos do tipo Bairro Ambiental.

Por isso, não é possível empregar uma metodologia mais simplificada, que consistiria em encontrar a área total edificável por distrito por meio da subtração, da área total de cada distrito, da somatória das áreas desses polígonos que representam restrições à edificação e que estão situados no distrito. Pois nesse caso estaria-se duplicando ou mesmo triplicando a contagem das áreas não-edificáveis, gerando distorções bastante elevadas, sobretudo em distritos mais ao norte e ao sul da cidade que possuem parte significativa de suas áreas ocupadas por matas protegidas por unidades de conservação. Mas também haveria distorções em distritos nas áreas nobres da cidade, onde estão localizadas os Bairros Ambientais, que possuem, quase sem exceção, o zoneamento do tipo ZER.

Dessa forma, foi realizada a seguinte metodologia:

- Filtrou-se os polígonos das quadras para conter apenas as quadras fiscais;
- Os polígonos filtrados foram dissolvidos em um único multipolígono;
- A partir do multipolígono gerado no passo 2, foram sendo sucessivamente “recortados” os polígonos referentes às restrições;
- É calculada então a intersecção do multipolígono resultante do passo 3 com a camada de distritos, de modo a identificar a porção desse multipolígono relativa ao território de cada distrito;
- A área de cada um dos multipolígonos referentes aos territórios dos distritos é calculada e a geometria é descartada, resultando em um *dataframe* com o seguinte formato:

Quadro 10 – Área edificável por distrito

Nome do distrito (str)	Área edificável m2 (float)
Sé	1322002
Pinheiros	4251012
...	...

Fonte: Elaboração Própria

O passo 3 foi realizado utilizando a operação *difference* do método *overlay* da biblioteca *geopandas*²⁹. Basicamente esse método tem por efeito “recortar” de um polígono a área de sua intersecção com outro polígono. Assim, ao ser realizada de forma sucessiva, essa operação tem por efeito evitar a contagem duplicada das áreas não edificáveis. Por exemplo, após ter sido removida da área das quadras fiscais a área referente às Unidades de Conservação, quando for realizado o mesmo procedimento em relação aos remanescentes de Mata Atlântica, não haverá nada a ser obtido nas áreas que correspondiam às UC, mesmo que nelas haja Mata Atlântica, pois a área referente às UC já foi removida anteriormente.

Pode-se utilizar como analogia o processo de feitura de biscoitos com o uso de moldes, representados pela Figura 9. A área das quadras fiscais pode ser concebida como a massa de biscoito original, aberta sobre a bancada, da qual sucessivamente vão sendo retirados pedaços com os moldes. Nessa analogia, os moldes correspondem a cada uma das áreas não edificáveis referidas acima. Caso um molde seja posicionado sobre a região onde um outro molde foi aplicado anteriormente, não será obtida nenhuma massa, pois essa já foi removida anteriormente.

O resultado final – a massa que restará sobre a bancada após aplicação dos moldes relativos às áreas não edificáveis – corresponderá à área total edificável da cidade. A massa restante será então “recortada” uma última vez, utilizando-se os moldes relativos aos territórios de cada um dos distritos das cidades. Por fim, será calculada a área de cada pedaço de massa que restou dentro da área do molde relativo a cada distrito. A área desses pedaços de massa representará a área edificável em cada distrito.

²⁹a documentação pode ser consultada no seguinte link

Figura 9 – Analogia – moldes de biscoito.

Fonte: Elaboração Própria com auxílio do ChatGPT

Além disso, também foi calculada para cada distrito a área já saturada com edificações. Para isso, utilizou-se da base do IPTU, identificando a área de terreno de todos os imóveis que possuem mais de 15 pavimentos. Essa quantidade de pavimentos foi escolhida pois representa um edifício de porte significativo, cujos custos de demolição e dificuldade de incorporação implicariam em um impedimento a projetos de verticalização. Trata-se do valor referente a um acima da mediana da quantidade de pavimentos dos edifícios em que se situam os imóveis do município³⁰.

Neste caso, novamente é preciso empregar uma metodologia específica

³⁰Note-se que cada apartamento situado em um edifício com n pavimentos contribuirá com n para essa distribuição, de modo que esse valor não pode ser lido como uma *proxy* para a “altura mediana” dos edifícios na cidade.

para evitar que o mesmo terreno seja contabilizado mais de uma vez. Isso ocorre sobretudo em edifícios residenciais verticais ou edifícios de escritório, nos quais cada unidade do condomínio está registrada no IPTU sob um Número de Contribuinte distinto. Para isso, foi construído um ID do edifício a partir do código S-Q-L dos apartamentos nele situados, que foi truncado para os seis primeiros dígitos, representando o setor e a quadra, portanto. Esse código identificador da quadra foi concatenado com o código do logradouro em que se situa o imóvel³¹ e com o número da rua do campo de endereço, de acordo com o seguinte algoritmo (onde o operador + representa a concatenação de *strings*):

$$\text{id_edifício} = \text{codigo_quadra} + \text{'.'} \text{codlog} + \text{'.'} \text{numero_rua}$$

Esse identificador resultou em uma menor variação da área do terreno para os registros do IPTU referentes a um mesmo valor de *id* (isso é, as unidades do prédio segundo esse método) do que outro identificador que foi testado, que consistia no uso do código da quadra concatenado ao código de condomínio também disponível na base do IPTU.

A metodologia de cálculo da área saturada de edificações consistiu assim nos seguintes passos:

- Identificação dos edifícios por meio da construção do *id* do edifício;
- Exclusão dos registros duplicados, mantendo-se apenas o primeiro imóvel para cada edifício, que passará a representá-lo;
- Identificação dos distritos dos lotes que representam os edifícios por meio da base georreferenciada de quadras fiscais;
- Agrupamento dos edifícios obtidos no passo 2 com base no distrito em que os lotes que os representam eles se situam;
- Somatória da área dos terrenos dos lotes representativos dos edifícios por distrito, resultando no seguinte *dataframe* que representa a área edificada saturada (maior que 15 pavimentos) por distrito:

³¹Trata-se de um *id* para as ruas da cidade mantido pela Prefeitura de São Paulo.

Quadro 11 – Área saturada por distrito

Nome do distrito (str)	Área edificada > 15 pavimentos (float)
Sé	22002
Pinheiros	41012
...	...

Fonte: Elaboração Própria

A área edificada saturada foi então subtraída da área edificável obtida segundo a metodologia apresentada anteriormente, obtendo-se a área efetivamente edificável por distrito, considerando não apenas as restrições à edificação como também terrenos nos quais é inviável financeiramente adensar ainda mais por já estarem saturados.

D.6 Privilégio social

Além dos dados associados ao potencial construtivo e que definem restrições à edificação, foram construídas variáveis que buscam mensurar o nível de privilégio social de cada distrito da cidade. Esses dados são analisados no Capítulo 3, com o objetivo de testar a hipótese de que os incentivos construtivos fornecidos pelo planejamento urbano possuem maior efetividade nos bairros onde reside a elite branca de alta renda e alta escolaridade, que possuem maior valor de mercado e para os quais estão voltados os interesses do capital imobiliário. E não surtem efeitos significativos na periferia, na qual há pouco interesse do mercado imobiliário formal.

Para construir essas variáveis, foram utilizados os microdados do Censo IBGE 2010 apresentados na Subseção C.2.1. Foram selecionadas variáveis que representam a renda, a escolaridade e a composição racial dos setores censitários, nomeadamente:

- A quantidade de pessoas de cada uma das raça/cor³² residentes no setor;

³²Nomeadamente: negras, pardas, amarelas e indígenas. Trata-se das categorias utilizadas pelo próprio IBGE.

- A quantidade de responsáveis pelo domicílio alfabetizados residentes no setor;
- A média da renda nominal dos responsáveis por domicílio residentes no setor que possuem renda;
- A quantidade de domicílios no setor com renda *per capita* superior a 10 salários mínimos;
- O total de domicílios particulares permanentes no setor;

Os dados foram agrupados por distrito. Neste caso, não foi necessário realizar operações geográficas com os polígonos dos setores censitários, pois os microdados fornecidos pelo [IBGE](#) já contém uma coluna que identifica o distrito da cidade em que o setor está localizado. Foi necessário apenas realizar a padronização dos nomes dos distritos, de modo a permitir posteriormente realizar com sucesso o *join* com os polígonos dos distritos da cidade disponíveis no Portal GeoSampa. Assim, para cada distrito foram calculadas as seguintes variáveis:

- A proporção de pessoas brancas residentes no distrito, de acordo com os seguintes passos:
 - Somatória da quantidade de pessoas brancas residentes no distrito;
 - Somatória da quantidade de pessoas de todas as raça/cor residentes no distrito – obtendo assim o total de pessoas residentes;
 - Divisão da quantidade de pessoas brancas no distrito pelo total de pessoas residentes no distrito;
- A proporção dos responsáveis por domicílio alfabetizados no distrito, a partir dos seguintes passos:
 - Somatória do total de responsáveis por domicílio alfabetizados no distrito;
 - Somatória do total de domicílios particulares permanentes no distrito;

- Divisão do total de responsáveis por domicílio alfabetizados no distrito pelo total de domicílios particulares permanentes;
- A proporção de domicílios com renda *per capita* superior a 10 salários mínimos, conforme as etapas:
 - Somatória do total de domicílios com renda *per capita* superior a 10 salários mínimos no distrito;
 - Somatória do total de domicílios particulares permanentes no distrito;
 - Divisão do total de domicílios com renda *per capita* superior a 10 salários mínimos no distrito pelo total de domicílios particulares permanentes;
- A média da renda nominal média dos responsáveis por domicílios dos setores contidos no distrito, a partir do agrupamento dos setores por distrito e cálculo da média por distrito;

A proporção de pessoas brancas tem por objetivo identificar distritos que possuem uma maior concentração de pessoas dessa raça/cor, que, devido ao processo histórico de colonização brasileiro e ao racismo estrutural que permanece mesmo após a abolição da escravidão, encontra-se hoje em posições mais privilegiadas na sociedade. A proporção de responsáveis por domicílio alfabetizados busca identificar diferenças nos níveis de escolaridade dos trabalhadores residentes nos distritos.

Há duas variáveis associadas à renda: a média da renda nominal média dos responsáveis por domicílio dos setores contidos no distrito e a proporção de domicílios com renda *per capita* superior a 10 salários mínimos no distrito. A primeira variável consiste em uma *proxie* para a renda média dos responsáveis por domicílio no distrito – infelizmente, devido ao risco de identificação, o IBGE não divulga diretamente a renda domiciliar. Já a segunda variável tem por objetivo ser mais sensível a distritos altamente desiguais, nos quais há uma população significativa de baixa renda que tende a reduzir a renda média do distrito, mas que possuem áreas de enclave de alta renda. Um exemplo clássico disso é o distrito da Vila Andrade, que contém condomínios de luxo com unidades de alta metragem ao lado da favela

de Paraisópolis, que é uma das áreas com a maior densidade habitacional da cidade.

A distribuição espacial dessas variáveis no território da cidade foi analisada no Capítulo 3, em conjunto com a média do tempo de deslocamento a trabalho. Isso pois o estudo tem por objetivo identificar a correlação entre o privilégio social e o acesso à cidade. A média do tempo de deslocamento a trabalho foi calculada utilizando os microdados dados da pesquisa OD para o ano de 2017³³, apresentada na subseção C.2.2 e que contém dados de viagens pendulares dos cidadãos da região metropolitana de São Paulo. Para isso, foram realizadas as seguintes etapas:

- Filtro dos registros da pesquisa OD para conter apenas as viagens originadas no município de São Paulo e que tem como destino o local de trabalho, isto é, aquelas cujo motivo no destino está entre as seguintes categorias:
 - Trabalho Indústria;
 - Trabalho Comércio;
 - Trabalho Serviços;
 - Procurar Emprego;
- Georreferenciamento da zona de origem das viagens com base na camada de Zonas Origem e Destino disponível no GeoSampa (ver quadra 2);
- Identificação do distrito das zonas de origem das viagens por meio do cruzamento espacial com a camada de distritos³⁴;
- Agrupamento dos dados por distrito;
- Cálculo da média do tempo de deslocamento por distrito. O *dataframe* resultante tem o seguinte formato:

³³Trata-se do último ano disponível da pesquisa, e é também o ano mais próximo do ano-base da análise realizada no Capítulo 3, que tem como base o ano de 2014.

³⁴Como não há correspondência exata entre os polígonos das zonas OD e dos distritos, esse cruzamento foi realizado por meio da identificação, para cada Zona, do distrito cujo centróide está mais próximo dela. Note-se que como foram selecionadas apenas as viagens originadas no município de São Paulo, todas as zonas da base de dados estão contidas no território da cidade.

Quadro 12 – Média do tempo de desloc. a trabalho por distrito (minutos)

Nome do distrito (str)	Média tempo desloc. trabalho (float)
Sé	30.1
Pinheiros	35.4
Itaim Paulista	72.3
...	...

Fonte: Elaboração Própria

Esses mesmos dados serão analisados de acordo com uma metodologia similar para calcular a acessibilidade ao mercado de trabalho por distrito, que será apresentada na subseção [D.7](#).

D.7 Acessibilidade ao Mercado de Trabalho

Por fim, o último conjunto de dados produzido corresponde à mensuração da acessibilidade ao mercado de trabalho por distrito da cidade. Para isso, foram analisados os dados de viagens pendulares disponíveis na pesquisa [OD](#) referidos na seção anterior ([D.6](#)). Assim como os dados de vínculos empregatícios na cidade de São Paulo disponibilizados nos microdados da [RAIS](#), base de dados descrita na subseção [C.2.3](#). Esses dados são analisados no Capítulo 4, e correspondem aos anos de 2007 e 2017. Como mencionado na subseção [C.2.2](#), esses anos correspondem aos dois anos mais recentes de publicação da pesquisa [OD](#)³⁵.

Para calcular a acessibilidade ao mercado de trabalho, em primeiro lugar foi preciso identificar o tempo médio de deslocamento das viagens pendulares originadas em cada um dos distritos da cidade e que têm como destino cada um dos demais distritos. As viagens pendulares são aquelas que têm como origem a residência e destino o local de trabalho (ida ao trabalho) ou como origem o local de trabalho e destino a residência (retorno à casa).

Na pesquisa [OD](#), o motivo declarado da viagem categorizado como “9

³⁵Os dados da [RAIS](#) são publicados anualmente. Eles foram selecionados para 2007 e 2017 para corresponder à pesquisa [OD](#).

- Residência” na origem indica que se está partindo da residência, enquanto o mesmo motivo no destino indica que se está retornando a ela. A mesma lógica vale para os motivos associados ao trabalho: a viagem de ida ao trabalho é quando o motivo associado ao trabalho está no destino, e a de volta quando ele se encontra na origem. Dessa forma, as viagens pendulares na pesquisa OD são aquelas que se enquadram em um dos dois casos:

- Motivo na origem como residência e motivo no destino como trabalho; ou
- Motivo na origem como trabalho e motivo no destino como residência

Foram considerados os seguintes motivos como relacionados a viagens relacionadas ao trabalho³⁶:

- 1 - Trabalho/Indústria;
- 2 - Trabalho/Comércio;
- 3 - Trabalho/Serviços;

Uma vez identificadas as viagens pendulares, elas foram filtradas para selecionar apenas as viagens que têm como origem e destino zonas situadas na cidade de São Paulo. Isso é necessário para garantir que se trata de deslocamentos que ocorrem dentro do território do município, e que portanto têm como origem e destino um par de distritos da cidade (ou um único distrito para os casos em que o distrito do local de trabalho é o mesmo de residência). Foram identificadas 48255 viagens pendulares em São Paulo para o ano de 2007 e 42956 para o ano de 2017 ³⁷.

Como não há disponibilidade do *shapefile* contendo os polígonos das Zonas OD para o ano de 2007, para identificar os distritos das viagens foi necessário utilizar os arquivos de correspondência entre as zonas disponibilizados pelo Metro-SP. Trata-se de planilhas em formato .xlsx que contém

³⁶Note-se que nesse caso não se está considerando as viagens motivadas pela procura de emprego, como na metodologia apresentada na subseção D.6, pois os dados da RAIS indicam vínculos de trabalho existentes.

³⁷A maior quantidade de viagens pendulares dentro da cidade no ano anterior indica que houve uma mudança na distribuição dos postos de trabalho, com mais postos situados na região metropolitana.

o de-para entre as zonas de um ano a outro da pesquisa, para permitir a comparabilidade. Essas planilhas contém também a identificação textual do distrito em que se encontra a zona – dado que foi utilizado para identificar o distrito de todas as zonas de origem e destino das viagens pendulares para ambos os anos da pesquisa OD analisados³⁸. Isso permitiu identificar os distritos das zonas de origem e de destino, resultando em um *dataframe* com o seguinte formato:

Quadro 13 – Viagens pendulares - distritos de origem e destino

Id viagem (int)	Distrito origem (str)	Distrito destino (str)	Tempo desloc. min. (int)
132	Pinheiros	Moema	35
133	Itaim Paulista	Consolação	72
...

Fonte: Elaboração Própria

As viagens foram então agrupadas pelos pares de distritos de origem e destino, utilizando-se o método *groupby* da biblioteca *pandas* e parame-
trizando como chave ambas as colunas. O tempo de deslocamento médio para cada par de distritos foi então calculado, resultando em um *dataframe* com o seguinte formato:

Quadro 14 – Média tempo desloc. - dist. origem e destino

Distrito origem (str)	Distrito destino (str)	Média tempo desloc. min. (float)
Pinheiros	Moema	27.5
Itaim Paulista	Consolação	68.2
...

Fonte: Elaboração Própria

³⁸Isso permitiu também identificar uma inconsistência os dados de 2007, pois algumas viagens originadas em ou destinadas a zonas que estão situadas fora do município de São Paulo foram cadastradas na base como tendo como origem e/ou o município (trata-se de uma coluna distinta). Essas viagens foram excluídas do estudo.

O processo acima foi realizado para ambos os anos do estudo, nomeadamente 2007 e 2017. Em seguida, procedeu-se ao georreferenciamento, por distrito, dos postos de trabalho da cidade. Isso foi realizado utilizando-se os microdados da RAIS, também para os anos de 2007 e 2017, de modo a garantir a compatibilidade com os dados de deslocamento fornecidos pela Pesquisa OD. Foi utilizada a tabela RAIS Estabelecimentos, que contém informação sobre os estabelecimento que declaram vínculos empregatícios ao Ministério do Trabalho. Os dados incluem o total de postos de trabalho do estabelecimento, o município no qual se localiza e – especificamente para a cidade de São Paulo, devido às suas dimensões – o detalhamento do distrito em que está situado.

Os estabelecimentos foram filtrados para conter apenas aqueles com indicativo de que estavam em atividade naquele ano³⁹ e que estão situados no município de São Paulo. Note-se que é preciso realizar esse filtro, pois, no ano de 2017, por exemplo, 27% dos estabelecimentos não possuíam indicação de atividade. Em seguida, com o uso dos arquivos de *layout* e metadados fornecidos pelo MTE, obtivemos o de-para entre os códigos numéricos contidos na coluna de distrito dos microdados da RAIS e os nomes dos distritos de fato. Assim como no caso da OD, a grafia dos nomes foi padronizada para tornar-se equivalente à grafia da camada de distritos disponibilizada no GeoSampa.

Os dados dos estabelecimentos foram então agrupados por distrito, e procedeu-se à somatória da quantidade de vínculos ativos⁴⁰ por distrito, resultando em um *dataframe* com esse formato:

Quadro 15 – *Quantidade de vínculos empregatícios por distrito*

Distrito (str)	Qtd. vínculos (int)
Alto de Pinheiros	16058
Anhanguera	2769
...	...

Fonte: Elaboração Própria

³⁹Coluna “Ind Atividade Ano” igual a 1.

⁴⁰Que é igual à quantidade de vínculos estatutários somada à quantidade de vínculos CLT.

Os *dataframes* representados pelos Quadros 14 e 15 foram então utilizados para calcular o percentual dos postos de trabalho da cidade que podem ser acessados a partir de uma viagem pendular de até 60 minutos a partir de cada distrito da cidade. A metodologia requer a definição desse limite temporal pois uma vez que não há qualquer acidente geográfico intransponível, todos os empregos da cidade são potencialmente acessíveis de qualquer distrito, mas em alguns casos mediante tempos de deslocamento extremamente longos, que são inviáveis ou então resultam em um elevado custo de oportunidade, quando não mesmo de saúde, pelo indivíduo que realiza essas viagens pendulares.

O limite temporal escolhido – 60 minutos – foi definido pois está próximo da mediana do tempo médio de deslocamento para todos os pares de distritos no ano de 2007⁴¹, que é de 62.5 minutos. Ele foi arredondado para a primeira hora cheia para facilitar a interpretação. Mas também para se aproximar mais das decisões que os indivíduos de fato realizam quando estão resolvendo o *trade-off* entre a qualidade e acessibilidade financeira da moradia e a proximidade da mesma ao local de trabalho que foi descrito no Capítulo 2: evitar ter que realizar deslocamentos com “mais de uma hora de duração” é muito mais provável do que “com mais de 62 minutos e meio”. Além disso, trata-se também do limite temporal utilizado por Bertaud (2022b) em sua análise. Por fim, não foi escolhido o limite de 15 minutos, como no caso da Cidade de 15 Minutos analisado no Capítulo 2, pois, considerando-se os dados de tempo médio de deslocamento a trabalho descritos na subseção D.6, esse valor é ainda uma utopia para a realidade de São Paulo, mesmo dentre os mais privilegiados.

O percentual de postos de trabalho acessíveis em até 60 minutos partindo de cada um dos distritos foi calculado da seguinte forma:

- Filtro da base de dados que possui a média do tempo de deslocamento entre os distritos, representada pelo Quadro 14, selecionando apenas as linhas que contém o distrito de interesse como origem;
- Para cada uma das linhas do passo anterior, que contém o tempo médio de deslocamento a um distrito de destino:

⁴¹Para haver comparabilidade, é necessário que o mesmo limite temporal seja utilizado em ambos os anos, por isso selecionamos apenas a mediana do primeiro ano.

- Caso o tempo médio de viagem pendular seja inferior ou igual a 60 minutos, o total de postos de trabalho localizados no distrito de destino, conforme consta no *dataframe* representado no Quadro 15, será adicionado ao total de postos de trabalho acessíveis pelo distrito de interesse;
- Caso o tempo médio de viagem seja superior a 60 minutos, o total de postos de trabalho do distrito de destino não será contabilizado;
- O total de postos de trabalho acessível pelo distrito de interesse é então dividido pelo total de postos de trabalho existentes na cidade (isto é, a somatória do total de postos de trabalho situados em todos os distritos), obtendo-se o percentual de postos de trabalho acessível pelo distrito de interesse em até 60 minutos;

O resultado final, após o percentual ser calculado para cada um dos distritos da cidade, é um *dataframe* com o seguinte formato:

Quadro 16 – *Percentual de Postos de Trabalho Acessíveis em até 60 minutos*

Distrito (str)	Percent. postos de trabalho (float)
Alto de Pinheiros	0.94
Itaim Paulista	0.45
...	...

Fonte: Elaboração Própria

Note-se que, caso não haja viagens registradas na base de dados da pesquisa OD entre o distrito de interesse e algum outro distrito de destino da cidade, os postos de trabalho do distrito de destino serão desconsiderados na somatória dos postos de trabalho acessíveis a partir do distrito de interesse. Essa decisão metodológica está baseada no fato de que altos tempos de deslocamento do distrito de origem ao distrito de destino e/ou a falta de oferta de empregos no distrito de destino inibem os cidadãos a realizar essas viagens pendulares de forma rotineira. Ao desconsiderar o total de postos de trabalho do distrito de destino, a metodologia empregada é consistente com esse fato, pois, ainda que não se possua registros

do tempo médio de deslocamento entre ambos os distritos, sabe-se que não há interesse em realizar a viagem.

Por fim, como esses dados foram gerados para ambos os anos (2007 e 2017), é possível identificar as mudanças nesse percentual ao longo do período. Para isso, o valor relativo ao ano de 2007 é subtraído do percentual de 2017. Caso o resultado dessa operação seja positivo, sabe-se que houve um aumento na acessibilidade ao mercado de trabalho para aquele distrito no período. Caso seja negativo, a acessibilidade diminuiu.

Essa metodologia permite capturar mudanças promovidas pela alteração na distribuição espacial dos postos de trabalho, por exemplo devido à criação de novos empregos em uma região comercial que se valorizou, ou mesmo devido à extinção de postos de trabalho em alguma região pela saída de uma fábrica da cidade. Ela é sensível também a um aumento na acessibilidade gerado por uma alteração nas condições de transporte. Assim, a inauguração de uma estação de metrô, que levou a menores tempos de deslocamento entre um distrito e o centro da cidade, pode acarretar no aumento da acessibilidade do mesmo ao mercado de trabalho. Ou o contrário: uma piora no trânsito em um determinado trajeto pode resultar na diminuição da acessibilidade desse distrito.

A lista de etapas abaixo resume as etapas da metodologia empregada, que foi replicada para ambos os anos, 2007 e 2017:

- Identificação das viagens pendulares (que têm como origem a residência e destino o trabalho ou vice-versa);
- Seleção apenas das viagens pendulares que têm como origem e destino zonas dentro do município de São Paulo;
- Identificação dos distritos das zonas de origem e destino das viagens pendulares;
- Agrupamento das viagens pendulares pelos pares de distritos de origem e destino;
- Cálculo da média de tempo de deslocamento para cada par;
- Seleção dos estabelecimentos em atividade em São Paulo;

- Agrupamento dos estabelecimentos por distrito;
- Somatória da quantidade de vínculos ativos por distrito;
- Identificação dos postos de trabalho acessíveis em viagens pendulares com duração igual ou inferior a 60 minutos, partindo de um distrito, a todos os demais, com base nos dados gerados nos passos 5 (tempo médio de deslocamento) e 8 (distribuição dos postos de trabalho);
- Somatória do total de postos de trabalho acessíveis para cada distrito;
- Divisão da somatória do total de postos de trabalho acessíveis pelo total geral de postos de trabalho da cidade, obtendo o percentual de postos de trabalho acessíveis partindo do distrito em viagens de até 60 minutos.

Isso conclui os processos de engenharia de dados realizados nesta dissertação.

E DESCRIÇÃO DAS BASES DE DADOS

Ao final do processo de engenharia de dados, foram gerados dois bancos de dados distintos:

- Uma base de dados agregado a nível de quadra que mensura o adensamento construtivo na cidade no período 2014-2023 a partir da quantidade de unidades produzidas – detalhado na subseção [E.1](#);
- Uma base de dados agregado a nível de distrito, que mensura o adensamento construtivo na cidade no período 2007-2017 a partir do total de metros quadrados de área construída produzidos – detalhado na subseção [E.2](#);

Nesta seção, que encerra o presente anexo, ambos as bases de dados serão descritos, com a apresentação do dicionário de variáveis.

E.1 Base de dados - adensamento medido por unidades produzidas

Essa base de dados é analisado no Capítulo 3. Nela, cada registro representa uma quadra fiscal da cidade de São Paulo, contendo o universo das quadras fiscais contidas na base do IPTU para os anos de 2014 e 2023, com exceção de um percentual baixo de quadras ($< 0,1\%$) que não pôde ser georreferenciado pois seu polígono não foi encontrado na camada disponibilizada no GeoSampa⁴². Ao todo, são 45662 registros e 31 variáveis.

Além da quantidade de unidades produzidas em empreendimentos verticais em cada uma dessas quadras (descrita na subseção D.1.2), a base de dados contém diversas variáveis que estão associadas ao potencial de verticalização da quadra (subseção D.3) ou consistem em impedimentos para que essa verticalização ocorra (subseção D.4). Ela contém também a identificação das quadras que estão situadas em zona de Eixo de Estruturação da Transformação Urbana (ZEU), conforme descrito na subseção D.2. Além disso, a base de dados contém o resultado de uma análise de *clusters* espaciais realizada no Capítulo 3 utilizando-se as variáveis descritas na subseção D.6 que identifica as quadras situadas nas áreas nobre e periférica da cidade.

Os quadros a seguir – três ao todo – descrevem as variáveis do base de dados. O primeiro deles, Quadro 17 apresenta as variáveis índice assim como *dummies* relacionadas à localização espacial da quadra e ao zoneamento da mesma.

⁴²As razões para isso estão associadas a parcelamentos do solo e obras viárias, que alteram os polígonos das quadras, conforme descrito na subseção D.1

Quadro 17 – *BD Unidades – índices e dummies*

Variável	Tipo	Descrição
quadra	integer	Índice da quadra
index_multinivel	factor	Índice do cluster. 1 representa as áreas periféricas, 2 a área nobre e 3 o restante da cidade.
ZEU	booleana	Quadra encontra-se em ZEU
ZEU_area_nobre	booleana	Quadra encontra-se em ZEU e também na área nobre (index_multinivel=2).
ZEU_periferia	booleana	Quadra encontra-se em ZEU e também em área periférica (index_multinivel=1).
ZER	booleana	Quadra encontra-se em ZER .
area_nobre	booleana	Quadra encontra-se na área nobre (index_multinivel=2).
area_periferica	booleana	Quadra encontra-se em área periférica (index_multinivel=1)
macrozona_ambiental	booleana	Quadra encontra-se na Macro-zona de Proteção Ambiental definida no PDE
proxima_eixo	booleana	Quadra é vizinha a uma quadra em ZEU , mas não está em ZEU ela mesma.

Fonte: Elaboração Própria

O Quadro 18 descreve as variáveis associadas às restrições à edificação apresentadas na subseção D.4.

Quadro 18 – *BD Unidades – restrições*

Variável	Tipo	Descrição
area_envoltoria_tomb	booleana	Quadra encontra-se em área envoltória de bem tombado (CONPRES , CONDEPHAAT e/ou IPHAN)
tombo	booleana	Há um imóvel tombado na quadra.
bairro_ambiental	booleana	Quadra encontra-se em Bairro Ambiental.
area_contaminada	booleana	Há uma área contaminada na quadra.
area_protecao_amb	booleana	Quadra encontra-se em APA .
protecao_mananciais	booleana	Quadra encontra-se em Área de Proteção aos Mananciais.
rppn	booleana	Há RPPN na quadra.
unidade_de_conservacao	booleana	Quadra encontra-se em UC .
restricao_geotecnica	booleana	Quadra encontra-se em área com restrição geotécnica.
risco_geologico	booleana	Quadra encontra-se em área de risco geológico.
restricao_geotecnica	booleana	Há área com restrição geotécnica na quadra.
rampa_heliponto	booleana	Quadra encontra-se em rampa de pouso e decolagem de heliponto.

Fonte: Elaboração Própria

O Quadro 19 descreve as variáveis associadas ao potencial de verticalização da quadra descritas na subseção D.3.

Quadro 19 – *BD Unidades – potencial construtivo*

Variável	Tipo	Descrição
valor_medio_m2_terreno	float	Valor médio do metro quadrado do terreno na quadra.
idade_imoveis_2014	float	Média da idade dos imóveis na quadra em 2014.
idade_contrib_2014	float	Média da idade dos contribuintes dos imóveis na quadra em 2014.
media_qtd_pavimentos	float	Média da quantidade de pavimentos dos imóveis na quadra em 2014.
prop_maior_terreno	float	Proporção, em relação à área total da quadra, da área do maior terreno de imóvel não-vertical.
prop_area_constru_horiz	float	Proporção da área construída total de edificações horizontais em relação à área total da quadra.
prop_area_constru_vert	float	Proporção da área construída total de edificações verticais em relação à área total da quadra.
prop_area_galpao	float	Proporção da área construída total de galpões e terrenos baldios em relação à área total da quadra.
qd_area	float	Área total da quadra em metros quadrados.
qt_unidades	int	Total de unidades produzidas por empreendimentos verticais no período 2014-2023.

Fonte: Elaboração Própria

As tabelas 1 e 3 apresentam a descrição das variáveis explicativas da base de dados, agrupadas, como anteriormente, entre as variáveis que repre-

sentam restrições construtivas e aquelas que estão associadas ao potencial de edificação. São apresentados a contagem de registros, a média, os valores mínimo, máximo e o primeiro, segundo e terceiro quartil.

Descrição da Base de Dados - Restrições à edificação por quadra da Cidade de São Paulo.								
index	count	mean	std	min	25%	50%	75%	max
area_envoltoria_tombamento	45662.000	0.045	0.208	0.000	0.000	0.000	0.000	1.000
tombo	45662.000	0.023	0.149	0.000	0.000	0.000	0.000	1.000
bairro_ambiental	45662.000	0.015	0.123	0.000	0.000	0.000	0.000	1.000
area_contaminada	45662.000	0.011	0.103	0.000	0.000	0.000	0.000	1.000
area_protecao_ambiental	45662.000	0.011	0.105	0.000	0.000	0.000	0.000	1.000
protecao_aos_mananciais	45662.000	0.014	0.120	0.000	0.000	0.000	0.000	1.000
restricao_geotecnica	45662.000	0.001	0.038	0.000	0.000	0.000	0.000	1.000
risco_geologico	45662.000	0.001	0.022	0.000	0.000	0.000	0.000	1.000
rppn	45662.000	0.000	0.007	0.000	0.000	0.000	0.000	1.000
unidade_de_conservacao	45662.000	0.000	0.013	0.000	0.000	0.000	0.000	1.000
rampa_heliponto	45662.000	0.011	0.103	0.000	0.000	0.000	0.000	1.000

Tabela 1 – *Descrição da Base de Dados - Restrições à edificação por quadra da Cidade de São Paulo.*

Descrição da Base de Dados - Potencial de verticalização por quadra da Cidade de São Paulo.								
index	count	mean	std	min	25%	50%	75%	max
valor_medio_m2_terreno	45662.000	0.000	1.000	-0.794	-0.535	-0.316	0.150	21.880
idade_media_imoveis_em_2014	45662.000	0.000	1.000	-2.352	-0.496	-0.036	0.370	7.766
idade_media_contribuintes_em_2014	45662.000	-0.000	1.000	-2.971	-0.548	0.169	0.781	1.391
media_qtd_pavimentos	45662.000	-0.000	1.000	-0.993	-0.309	-0.196	-0.040	18.925
prop_maior_terreno_nao_vertical	45662.000	-0.000	1.000	-0.750	-0.487	-0.382	-0.014	4.504
prop_area_total_edificada_horizontal	45662.000	-0.000	1.000	-1.773	-0.673	0.072	0.640	14.552
prop_area_total_edificada_vertical	45662.000	0.000	1.000	-0.307	-0.307	-0.307	-0.267	22.677
prop_area_total_terrenos_e_galpoes_etc	45662.000	0.000	1.000	-0.427	-0.427	-0.427	-0.045	7.208
qd_area	45662.000	0.000	1.000	-0.303	-0.196	-0.127	-0.016	107.257

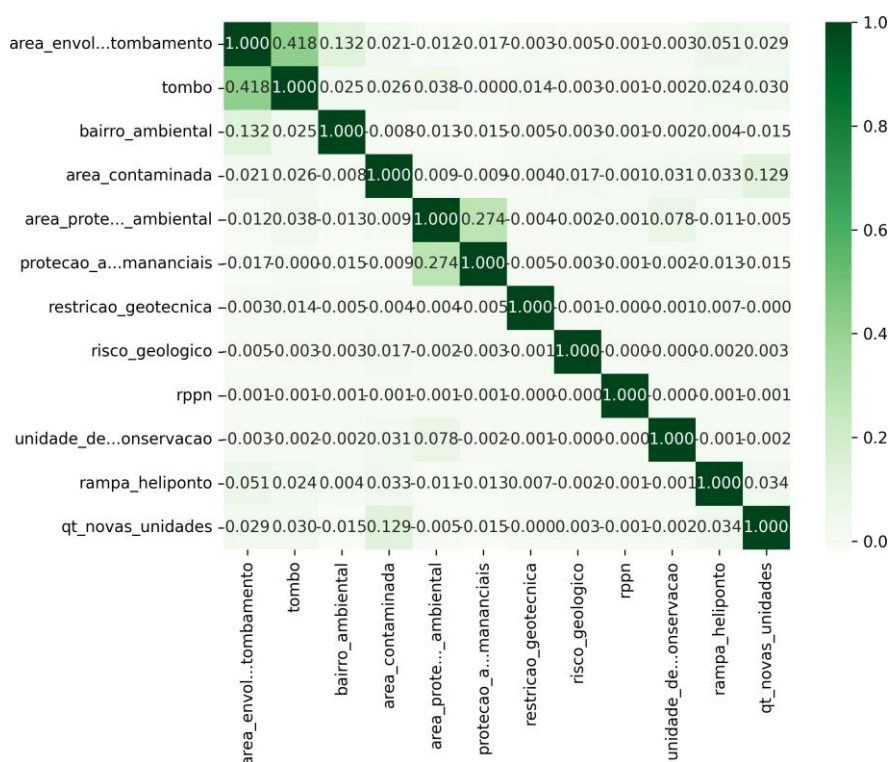
Tabela 2 – *Descrição da Base de Dados - Potencial de verticalização por quadra da Cidade de São Paulo.*

Descrição da Base de Dados - Dependente, tratamento e índices espaciais.								
index	count	mean	std	min	25%	50%	75%	max
qt_novas_unidades	45662.000	6.946	54.850	0.000	0.000	0.000	0.000	2576.000
ZEU	45662.000	0.088	0.283	0.000	0.000	0.000	0.000	1.000
ZEU_area_nobre	45662.000	0.026	0.158	0.000	0.000	0.000	0.000	1.000
ZEU_periferia	45662.000	0.011	0.105	0.000	0.000	0.000	0.000	1.000
ZER	45662.000	0.053	0.224	0.000	0.000	0.000	0.000	1.000
zer_para_zcor	45662.000	0.007	0.085	0.000	0.000	0.000	0.000	1.000

Tabela 3 – *Descrição da Base de Dados - Potencial de verticalização por quadra da Cidade de São Paulo.*

As Figuras 10 e 11 apresentam as matrizes de correlação de Pearson anotadas para cada um dos grupos de variáveis (restrições construtivas e potencial de verticalização). Em ambas as matrizes foi incluída também a variável *y* do estudo, isto é, a quantidade de unidades produzidas por quadra. Por fim, a Figura 12 apresenta a matriz de correlação entre as variáveis de tratamento e os índices espaciais com a variável dependente.

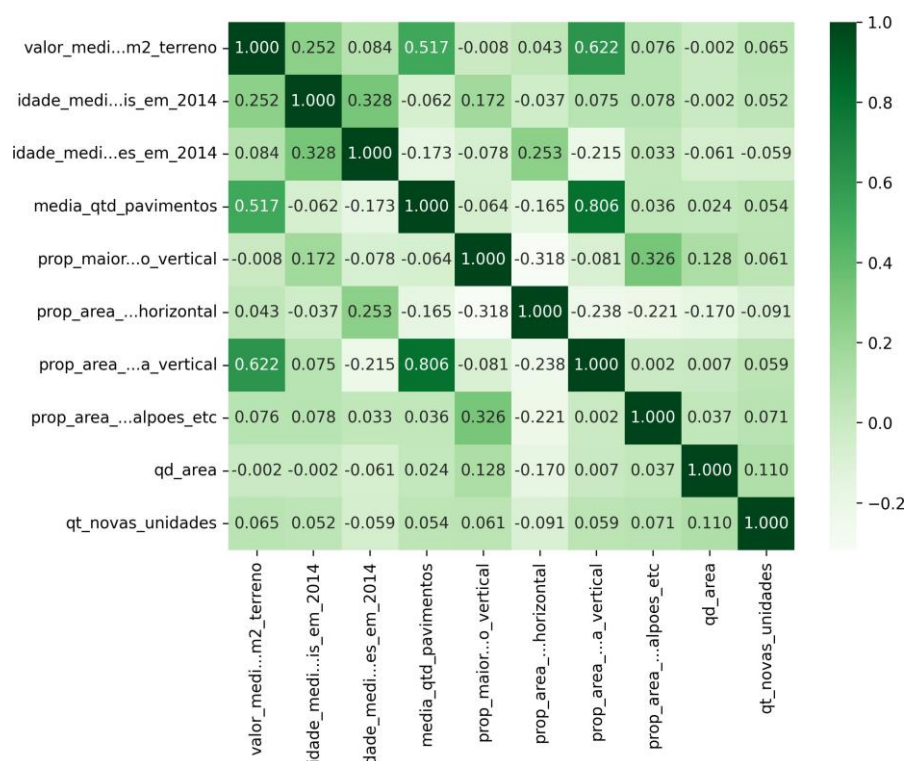
Figura 10 – *Matriz de Correlação de Pearson - Restrições Construtivas*



Fonte: Elaboração Própria

E.2 Base de dados - adensamento medido por metros quadrados de área construída

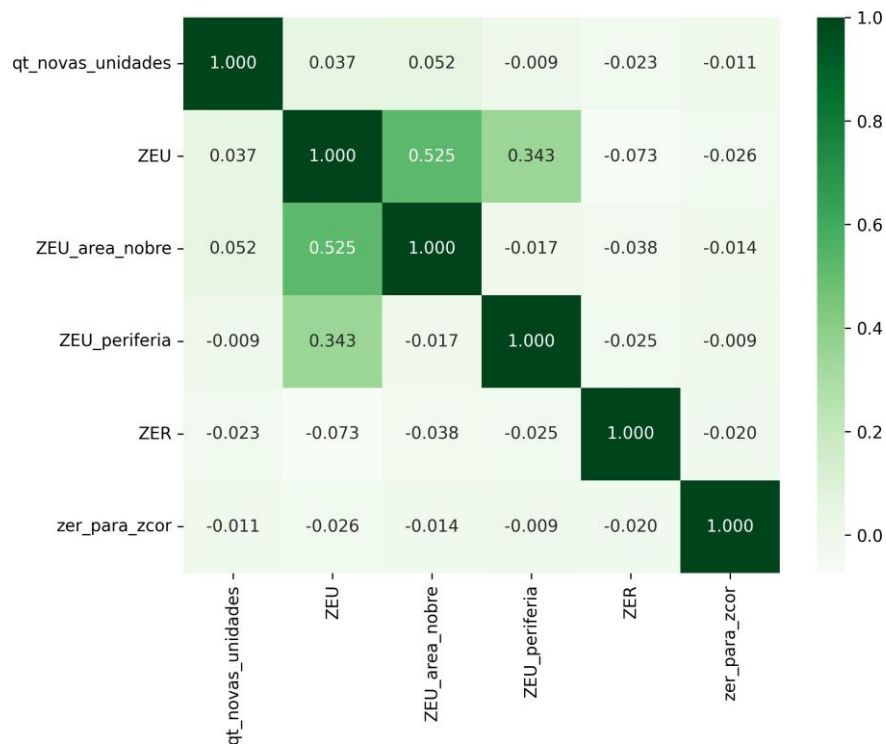
Esse banco de dados é analisado no Capítulo 4. Nele, cada registro representa um dos distritos da cidade de São Paulo. Ao todo, são 96 distritos e 13 variáveis. Os dados podem ser georreferenciados por meio da camada de distritos disponível no GeoSampa.

Figura 11 – Matriz de Correlação de Pearson - Potencial de Verticalização

Fonte: Elaboração Própria

A base de dados apresenta o total por distrito da área construída de todas as unidades de empreendimentos residenciais verticais produzidos na cidade no período 2007-2017, em metros quadrados, calculada conforme metodologia descrita na subseção D.1.3. Ela contém também os dados relativos à área edificável por distrito, obtida conforme apresentado na subseção D.5. Assim como o cálculo da acessibilidade ao mercado de trabalho por distrito, obtida por meio da metodologia apresentada na subseção D.7.

O Quadro 20 consiste no dicionário das variáveis construídas diretamente a partir das metodologias apresentadas nas subseções mencionadas acima. O Quadro 21 contém variáveis secundárias, que foram geradas por meio de operações realizadas entre as variáveis do 20. Por exemplo: a variável *coeficiente_aumentado* é calculada por meio da divisão do total da área construída no distrito (*area_constru*) pelo total da área do distrito que não possui impedimentos à verticalização (*area_verticalizavel*).

Figura 12 – *Matriz de Correlação de Pearson - Tratamentos e índices espaciais*

Fonte: Elaboração Própria

Quadro 20 – *BD Área Construída – Variáveis primárias*

Variável	Tipo	Descrição
distrito	string	Nome padronizado do distrito.
area	float	Área total do distrito, em metros quadrados.
area_terreno_verticalizado	float	Área total, em metros quadrados, dos terrenos que possuem edifícios com 15 ou mais pavimentos
area_verticalizavel	float	Área do distrito, em metros quadrados, que não tem impedimentos à verticalização.
area_verticalizavel_viavel	float	Área do distrito, em metros quadrados, que não tem impedimentos à verticalização e não está saturada de edifícios.
percentual_empregos_2007	float	Percentual do total de empregos da cidade acessível por viagens de até 60 minutos partindo do distrito em 2007.
percentual_empregos_2017	float	Percentual do total de empregos da cidade acessível por vi-

Quadro 21 – *BD Área Construída – Variáveis construídas*

Variável	Tipo	Descrição
dif_percentual	float	Variação no percentual de empregos acessível pelo distrito no período 2007-2017.
coef_area_verticalizavel	float	Proporção da área verticalizável em relação à área total do distrito.
coef_area_verticalizavel_viavel	float	Proporção da área verticalizável viável em relação à área total do distrito.
coeficiente_aumentado	float	Proporção da somatória da área construída de unidades residenciais verticais em relação à área verticalizável do distrito.
coeficiente_aumentado_viavel	float	Proporção da somatória da área construída de unidades residenciais verticais em relação à área verticalizável do distrito não saturada de edifícios.

Fonte: Elaboração Própria

A tabelas 4 apresenta a descrição das variáveis primárias dessa base de dados, assim como a variável dependente, *area_constru*, que representa a área construída vertical total por distrito. Já a tabela 5 apresenta a descrição das variáveis construídas listadas no Quadro 21.

Descrição da Base de Dados - Variáveis Primárias e Dependente								
index	count	mean	std	min	25%	50%	75%	max
area	96.000	15908594.959	26977435.268	2193567.836	7396560.507	9780573.408	13564560.712	20818714.280
area_terreno_verticalizado	96.000	189007.631	265142.791	0.000	5219.750	73829.933	255377.985	1150352.001
area_verticalizavel	96.000	5274593.057	2359897.339	682349.092	3932645.934	5114721.105	6410888.935	11648332.193
area_verticalizavel_viavel	96.000	5085585.426	2435405.370	410481.255	3796265.634	4873923.109	6341096.836	11615845.193
percentual_empregos_2007	96.000	0.411	0.230	0.047	0.205	0.393	0.639	0.807
percentual_empregos_2017	96.000	0.408	0.240	0.031	0.192	0.367	0.662	0.795
area_constru	96.000	507638.240	549891.112	0.000	118600.750	285192.500	749252.250	282690.000

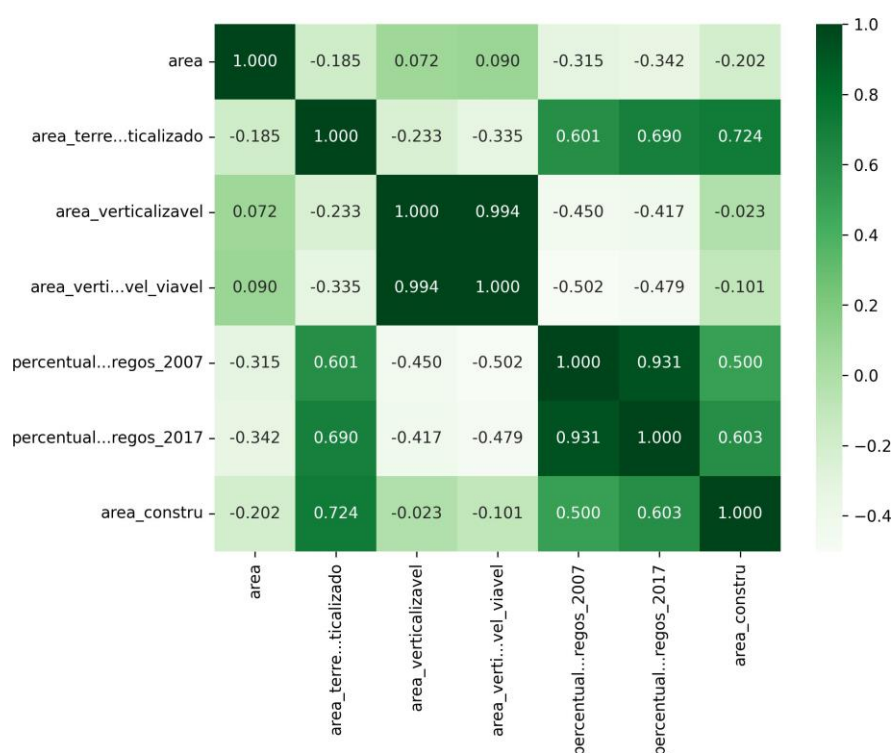
Tabela 4 – Descrição da Base de Dados - Variáveis Primárias e Dependente.

Descrição da Base de Dados - Variáveis Primárias e Dependente								
index	count	mean	std	min	25%	50%	75%	max
dif_percentual	96.000	-0.003	0.088	-0.255	-0.051	0.003	0.038	0.247
coeficiente_area_verticalizavel	96.000	0.499	0.169	0.003	0.393	0.541	0.632	0.707
coeficiente_area_verticalizavel_viavel	96.000	0.471	0.175	0.003	0.345	0.523	0.620	0.690
coeficiente_aumentado	96.000	0.118	0.131	0.000	0.022	0.065	0.182	0.601
coeficiente_aumentado_viavel	96.000	0.138	0.164	0.000	0.022	0.066	0.193	0.772

Tabela 5 – Descrição da Base de Dados - Variáveis Primárias e Dependente.

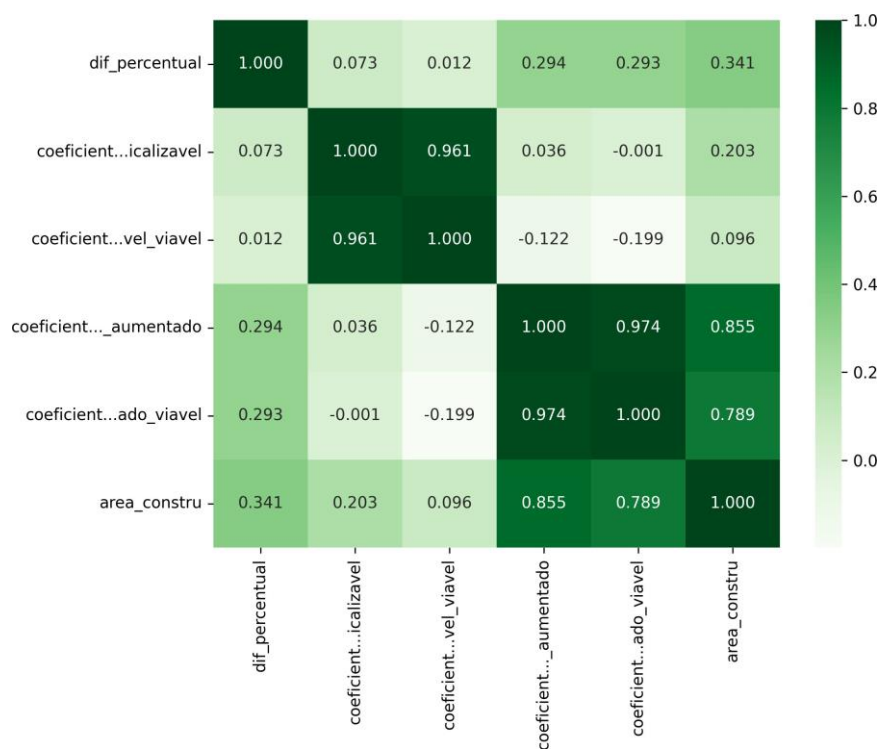
As Figuras 13 e 14 apresentam as matrizes de correlação de Pearson anotadas para cada um dos grupos de variáveis (primárias e construídas). Em ambas as matrizes foi incluída também a variável *y* do estudo, isto é, a área construída total de empreendimentos residenciais verticais por distrito.

Figura 13 – Matriz de Correlação de Pearson - Variáveis Primárias



Fonte: Elaboração Própria

Com isso, encerra-se a descrição das bases de dados que fundamentaram o presente estudo.

Figura 14 – *Matriz de Correlação de Pearson - Variáveis Construídas*

Fonte: Elaboração Própria