



O PAPEL DA MODELAGEM DA DIVISÃO MODAL NA ELABORAÇÃO DE PLANOS  
DE MOBILIDADE URBANA SUSTENTÁVEL

Marcello Victorino Junqueira de Souza

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Transportes, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Transportes.

Orientadores: Romulo Dante Orrico Filho

Marcelino Aurélio Vieira da Silva

Rio de Janeiro

Maio de 2016

O PAPEL DA MODELAGEM DA DIVISÃO MODAL NA ELABORAÇÃO DE PLANOS  
DE MOBILIDADE URBANA SUSTENTÁVEL

Marcello Victorino Junqueira de Souza

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO  
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)  
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS  
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM  
CIÊNCIAS EM ENGENHARIA DE TRANSPORTES.

Examinada por:

---

Prof. Romulo Dante Orrico Filho, Dr. Ing.

---

Prof. Marcelino Aurélio Vieira da Silva, D.Sc.

---

Prof. Angela Maria Gabriella Rossi, D.Sc.

---

Prof. Enílson Medeiros dos Santos, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

MAIO DE 2016

Souza, Marcello Victorino Junqueira de

O papel da modelagem da divisão modal na elaboração de planos de mobilidade urbana sustentável / Marcello Victorino Junqueira de Souza. – Rio de Janeiro: UFRJ/COPPE, 2016.

XII, 138 p.: il.; 29,7 cm.

Orientadores: Romulo Dante Orrico Filho

Marcelino Aurélio Vieira da Silva

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia de Transportes, 2016.

Referências Bibliográficas: p. 125-129.

1. Modelos de escolha discreta. 2. Plano de Mobilidade Urbana. 3. Priorização de investimentos. I. Orrico Filho, Romulo Dante *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Transportes. III. Título.

## DEDICATÓRIA

Aos meus pais, por se dedicarem à minha felicidade e existência.

À minha irmã, pela cumplicidade incondicional do amor fraterno.

Ao meu amor, por tornar cada alegria e conquista ainda mais saborosa.

Aos meus amigos, por me acompanharem em qualquer jornada, independente do destino.

Aos pessimistas, por me darem forças para me reerguer e continuar tentando.

E, principalmente, aos meus erros, por serem parte integrante do que vivi, aprendi e, de maneira contínua, contribuírem para construir a pessoa que me tornei.

*“Desconfiai do mais trivial, na aparência singelo*

*E examinai, sobretudo, o que parece habitual*

*Suplicamos expressamente: não aceiteis o que é de hábito como coisa natural, pois em tempo de desordem sangrenta, de confusão organizada, de arbitrariedade consciente, de humanidade desumanizada, nada deve parecer natural*

*Nada deve parecer impossível de mudar.”*

- Bertolt Brecht -

## AGRADECIMENTOS

Poucas pessoas têm a sorte de descobrirem, enquanto jovens, o que gostam de fazer. Mais raro ainda são aquelas que têm a oportunidade de atuarem, seja acadêmica ou profissionalmente, na sua área de interesse.

Assim, me considero extremamente afortunado pela sequência de experiências e acontecimentos na minha vida que me permitiram ser uma dessas poucas pessoas: feliz e realizado por estudar, aprender e colocar em prática aquilo que gosto (não necessariamente nessa ordem)!

Porém, me considero ainda mais afortunado por contar com diversas pessoas que contribuíram (e contribuem) diretamente com meu sucesso e felicidade:

Começo, assim, agradecendo aos meus pais, Seraphim Victorino e Terezinha Carvalho, que sempre acreditaram em mim - como quaisquer pais corujas - e sempre estiveram presentes nas minhas conquistas e vitórias.

À minha irmã, Alexandra Junqueira, por todas as "brigas" e alegrias que constituem a própria definição do Amor Fraternal!

À minha melhor amiga, namorada, noiva e esposa, Fernanda Matias, que teve papel fundamental nas etapas mais cruciais da vida acadêmica desse mestrando. A vontade de compartilhar alegrias com você me dá força e inspiração para puxar os limites cada vez mais longe e, obter assim, cada vez mais alegrias só para poder compartilhá-las! Sinceramente; muito obrigado por todas as dicas, paciência e apoio!

Aos professores do Programa de Engenharia de Transportes, que abriram as portas para um novo mundo de conhecimento que tem enorme importância e impacta diretamente a qualidade de vida da população. Quanto mais estudo, mais aprendo sobre novos campos de conhecimento a serem estudados!

Destaque especial aos professores orientadores que tive o prazer de conhecer e a oportunidade de adquirir conhecimento: Prof. Romulo Orrico Filho, por compartilhar tantos conselhos e dicas, além de esmiuçar os conceitos políticos e econômicos intrínsecos da área de transporte público; e Prof. Marcelino Aurélio, por me apresentar e conduzir o aprofundamento do conhecimento sobre as técnicas referentes ao planejamento de transportes, especificamente à etapa de modelagem.

Não poderia deixar de destacar a importância dos amigos de Mestrado, Gabriel Stumpf e Gabriel Tenenbaum, verdadeiros irmãos de consideração, por toda força,

parceria e ideias trocadas; seja durante o cafezinho pós-almoço ou alguma viagem para participar de congresso na área de Transportes. Além, é claro, dos encontros do “Delírio do PET”, os quais propiciou intensa troca de ideias e estreitou ainda mais esse vínculo de parceria e amizade, contando também com a ilustre participação dos Professores Carlos David Nassi e Glaydston Ribeiro.

Ao professor, amigo e ex-chefe, Eduardo Leal de Oliveira, que, atuando como supervisor/tutor durante minha experiência na área de consultoria, contribuiu significativamente para meu desenvolvimento profissional e pessoal.

Aos membros da Secretaria do Programa, Jane, Dona Helena e Natália, que, por inúmeras vezes, tornaram simples a resolução de situações que poderiam ser árduas e complexas.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão de bolsa de fomento à pesquisa ao longo do desenvolvimento desse trabalho.

Aproveito ainda para formalizar o agradecimento a pessoas que talvez nem saibam como foram úteis para o desenvolvimento dessa pesquisa: aos professores Orlando Strambi (USP) e Ana Larrañaga Uriarte (UFRGS), que participaram da sessão de acompanhamento de tese do congresso ANPET de 2015 e indicaram os “primeiros passos” a serem consultados para conhecer sobre a modelagem do Comportamento de Viagens; assim como à professora Lissy La Paix Puello (UTWENTE, Holanda), por permitir que sua aluna de doutorado, Flávia de Souza, compartilhasse o material desenvolvido sobre modelos de escolha discreta e workshops de utilização do BioGeme, além de responder atenciosamente às questões que surgiram pela utilização do programa.

Por último, mas não menos importante, agradeço encarecidamente aos membros da banca por dedicarem seu tempo e atenção à revisão dessa pesquisa de dissertação; sem dúvida alguma, contribuindo para seu aprimoramento.

A todos vocês, o meu muito e sincero obrigado.

Marcello Victorino Junqueira de Souza

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## O PAPEL DA MODELAGEM DA DIVISÃO MODAL NA ELABORAÇÃO DE PLANOS DE MOBILIDADE URBANA SUSTENTÁVEL

Marcello Victorino Junqueira de Souza

Maio/2016

Orientadores: Romulo Dante Orrico Filho

Marcelino Aurélio Vieira da Silva

Programa: Engenharia de Transportes

Essa pesquisa de dissertação tem o objetivo de comprovar o potencial de aplicação da modelagem da divisão modal para definir medidas que promovam a transferência modal desejada de forma mais eficaz e assim permitam apoiar a priorização de investimentos e determinação do plano de ações em sistemas de transportes, etapas obrigatórias em um Plano de Mobilidade Urbana Sustentável, melhor orientando a formulação de políticas públicas. Após introduzir o ferramental teórico mínimo necessário para a compreensão e replicação do método proposto, realiza-se uma revisão bibliográfica para identificar os fatores-chave utilizados com maior frequência na modelagem da divisão modal. Apresenta-se o procedimento proposto que permite ranquear os atributos que exercem maior influência sobre o processo de escolha modal, detalhando as etapas necessárias para o tratamento e consolidação dos dados. Por último, aplica-se a análise dos resultados obtidos em um ensaio ilustrativo com o intuito de exemplificar o potencial do procedimento na identificação de estratégias mais eficazes para promover a transferência modal desejada. Assim, a nova aplicação sugerida do procedimento proposto para apoiar a priorização de investimentos agrega transparência e respeita as características locais (*genius loci*), apoiando a elaboração de planos de mobilidade urbana sustentável e melhor orientando a formulação de políticas públicas, motivação desta pesquisa.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

THE ROLE OF MODAL SPLIT MODELING ON THE DEVELOPMENT OF  
SUSTAINABLE URBAN MOBILITY PLANS

Marcello Victorino Junqueira de Souza

May/2016

Advisors: Romulo Dante Orrico Filho

Marcelino Aurélio Vieira da Silva

Department: Transportation Engineering

The main objective of this research aims to prove the potential use of modal split modeling in order to identify strategies effectively promoting the desired modal shift, hence supporting the investment prioritization process and the action plan of transportation systems, which are obligated in the development of a Sustainable Urban Mobility Plan, therefore better guiding the formulation of public policies. After introducing the minimum theoretical concepts necessary for understanding and reproducing the proposed methodology, a bibliographic review is conducted in order to identify the key factors most often utilized in modal split modeling. The proposed procedure allowing the ranking of the most influent attributes over the mode choice process is presented. The necessary steps for data cleaning and its consolidation are described. Lastly, the analyses of the results are presented in order to exemplify the proposed procedure potential to identify the most effective strategies to promote the desired modal shift. Therefore, the new suggested application of the proposed procedure to aid the investment prioritization process aggregates transparency/defensibility while respecting the local characteristics (*genius loci*), aiding the development of sustainable urban mobility plans and better guiding the formulation of more effective public policies, which is the motivation of this research.

## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>1</b>
1.1 PROBLEMA E MOTIVAÇÃO.....	4
1.2 RELEVÂNCIA DO PROBLEMA .....	6
1.3 OBJETIVOS GERAL E ESPECÍFICOS .....	7
1.4 JUSTIFICATIVA .....	7
1.5 MÉTODO DE PESQUISA .....	9
1.6 PREMISSAS ADOTADAS .....	10
1.7 LIMITAÇÃO DA PESQUISA.....	10
1.8 ESTRUTURA DA PESQUISA.....	12
<b>2 MODELOS DE ESCOLHA DISCRETA .....</b>	<b>13</b>
2.1 TEORIA DA UTILIDADE ALEATÓRIA.....	14
2.1.1 Particularidades.....	16
2.1.2 Grupo Universal de Opções .....	17
2.1.3 Especificação da Componente Sistemática $V_{ij}$ .....	18
2.1.4 Derivada e Elasticidade: Direta e Cruzada .....	20
2.2 DETALHAMENTO DOS PRINCIPAIS MODELOS.....	23
2.2.1 Multinomial Logit (MNL).....	23
2.2.2 Nested Logit (NL) .....	25
2.2.3 Cross-Nested Logit (CNL) .....	28
2.2.4 Multinomial Probit (MNP).....	30
2.2.5 Mixed Logit (ML).....	31
2.2.6 Considerações finais sobre os Modelos.....	33
2.3 CALIBRAÇÃO .....	34
2.4 ETAPA DE TESTES .....	36
2.4.1 Teste Informal.....	36
2.4.2 Teste dos Coeficientes .....	37
2.4.3 Teste dos Modelos .....	38
2.4.4 Teste de Desempenho .....	39
2.5 EXEQUIBILIDADE .....	42
2.6 CONSIDERAÇÕES FINAIS .....	44
<b>3 FATORES-CHAVE DA MODELAGEM DA DIVISÃO MODAL .....</b>	<b>45</b>
3.1 ESTRATÉGIA DE BUSCA.....	46
3.2 AVALIAÇÃO GERAL DOS ARTIGOS .....	47
3.2.1 Categoria 1: Aplicações práticas .....	48
3.2.2 Categoria 2: Aprimoramento técnico.....	56
3.3 IDENTIFICAÇÃO DOS FATORES-CHAVE .....	62
3.3.1 Variável Dependente (Alternativas).....	62
3.3.2 Variáveis Independentes (Explicativas) .....	63

3.3.3	Motivo da Viagem.....	66
3.3.4	Modelos Utilizados e Indicadores de Desempenho ( <i>Goodness-of-Fit</i> ).....	66
3.3.5	Nível de Significância .....	68
3.3.6	Análises .....	69
3.4	CONSIDERAÇÕES FINAIS .....	70
<b>4</b>	<b>PROCEDIMENTO METODOLÓGICO .....</b>	<b>72</b>
4.1	LEVANTAMENTO DE DADOS - PESQUISA MULTIPROPÓSITO UFRJ.....	74
4.1.1	Elaboração do Questionário .....	74
4.1.2	Indicador de Acessibilidade – <i>PNT*</i> .....	75
4.2	TRATAMENTO DO BANCO DE DADOS .....	77
4.2.1	Remoção de Participações Incompletas .....	77
4.2.2	Remoção de Alternativas Escassas .....	77
4.2.3	Verificação de Consistência .....	78
4.2.4	Identificação e Remoção de <i>Outliers</i> .....	79
4.2.5	Estimação de Parâmetros .....	81
4.2.6	Consolidação do Banco de Dados .....	83
4.2.7	Definição do Tamanho Amostral .....	85
4.3	ESTIMAR O MODELO .....	87
4.3.1	Especificação/Construção .....	87
4.3.2	Calibração e Testes.....	88
4.3.3	Aperfeiçoamento Progressivo .....	95
4.4	MODELO ESTIMADO E VERIFICAÇÃO DE DESEMPENHO.....	101
4.5	CONSIDERAÇÕES FINAIS .....	104
<b>5</b>	<b>RESULTADOS E ANÁLISES .....</b>	<b>106</b>
5.1	ESTATÍSTICA DESCRITIVA.....	106
5.2	ANÁLISES .....	107
5.2.1	Valor do Tempo .....	108
5.2.2	Elasticidade Direta e Cruzada .....	109
5.2.3	Contribuição Metodológica Proposta: Ranqueamento de Atributos .....	113
5.2.4	Projeção do Efeito de Políticas Públicas .....	117
5.3	CONSIDERAÇÕES FINAIS .....	118
<b>6</b>	<b>CONCLUSÕES E RECOMENDAÇÕES FUTURAS .....</b>	<b>120</b>
<b>7</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>125</b>
<b>ANEXO A</b>	<b>.....</b>	<b>130</b>
<b>ANEXO B</b>	<b>.....</b>	<b>137</b>
<b>ANEXO C</b>	<b>.....</b>	<b>138</b>

## LISTA DE FIGURAS

Figura 1: Panorama da situação dos Planos de Mobilidade Urbana no Brasil (adaptado de IBGE, 2012) .....	3
Figura 2: Modelo Nested Logit: alternativas similares agrupadas em Ninhos .....	27
Figura 3: Cross-Nested Logit: alternativas simultaneamente em diferentes Ninhos .....	29
Figura 4: Relação empírica entre $\rho^2$ e $R^2$ (adaptado de Hensher, Rose e Greene (2015) .....	42
Figura 5: Exemplo de interface do programa NLOGIT 5 (adaptado de ECONOMETRIC SOFTWARE, 2015) .....	43
Figura 6: Exemplo dos resultados obtidos e ambiente de execução do BioGeme .....	44
Figura 7: Trecho do quadro sumário obtido pela revisão bibliográfica .....	45
Figura 8: Fluxograma da estratégia de busca da revisão sistemática .....	46
Figura 9: Principais variáveis explicativas agrupadas por características do Indivíduo, Modo, Ambiente e Outros, além de sua frequência .....	64
Figura 10: Fluxograma horizontal do procedimento metodológico proposto .....	73
Figura 11: Ilustração da análise espacial para cálculo do indicador <i>PNT</i> (CARVALHO, 2016). 76	
Figura 12: Box-Plot dos parâmetros Tempo e Custo de viagem, por alternativa .....	80
Figura 13: Fluxograma da verificação preliminar .....	89

## LISTA DE TABELAS

Tabela 1: Estatística t-Student em função do grau de liberdade e Nível de Confiança - adaptado de Fisher e Yates (1995) .....	37
Tabela 2: Estatística Chi-Quadrado em função do grau de liberdade e Nível de Confiança - adaptado de Fisher e Yates (1995) .....	39
Tabela 3: Frequência das variáveis independentes identificadas na revisão.....	62
Tabela 4 : Frequência das variáveis explicativas identificadas na revisão.....	65
Tabela 5: Frequência dos motivos de viagem identificados na revisão.....	66
Tabela 6: Frequência dos modelos utilizados e seus indicadores de desempenho.....	67
Tabela 7: Frequência dos níveis de significância identificados na revisão .....	69
Tabela 8: Frequência das análises identificadas na revisão .....	70
Tabela 9: Principais variáveis explicativas obtidas na pesquisa de Carvalho (2016).....	75
Tabela 10: Frequência de alternativas escassas.....	78
Tabela 11: Limites Superior e Inferior de <i>Outlier</i> do Banco#1 .....	80
Tabela 12: Quantidade de observações removidas pela análise de <i>Outlier</i> .....	80
Tabela 13: Desempenho médio de veículo automotor, por tipo de motor e ano - adaptado de MMA (2014) .....	82
Tabela 14: Total de observações e divisão modal, por Banco de Dados.....	85
Tabela 15: Tamanho amostral em função de diferentes parâmetros de aceitabilidade .....	87
Tabela 16: Variáveis explicativas, por grupo de características .....	88
Tabela 17: Calibração dos modelos de verificação preliminar, Banco#1 - Total.....	90
Tabela 18: Calibração do modelo de verificação preliminar, Banco#1 – Estratificado.....	91
Tabela 19: Calibração dos modelos de verificação preliminar, Banco#2 - Total.....	92
Tabela 20: Calibração do modelo de verificação preliminar, Banco#2 – Estratificado.....	94
Tabela 21: Coeficientes e verificações estatísticas, modelos 1, 2 e 3.....	96
Tabela 22: Coeficientes e verificações estatísticas, modelos 4, 5 e 6.....	98
Tabela 23: Coeficientes e verificações estatísticas, modelos 7, 8 e 9.....	99
Tabela 24: Significado dos atributos do modelo e seus respectivos coeficientes .....	102
Tabela 25: Percentual de acertos: escolhas observadas versus projeção do modelo.....	103
Tabela 26: Divisão modal real observada e estimada pelo modelo final, além de seu respectivo indicador de desempenho.....	104
Tabela 27: Sumário dos atributos incluídos no modelo, nos bancos Total e Alunos .....	106
Tabela 28: Elasticidade Direta e Cruzada, por atributo, por alternativa .....	110
Tabela 29: Ranqueamento dos atributos por ordem de influência sobre a probabilidade de utilização do Transporte Público.....	114
Tabela 30: Divisão modal original e simulada para cinco cenários .....	117

## 1. INTRODUÇÃO

De acordo com o relatório de perspectivas da urbanização global, em 2012, mais da metade da população mundial já vivia em cidades (ONU, 2012). A previsão é de que, até 2030, 60% da população mundial esteja concentrada em aglomerações urbanas (ONU, 2015a). Cada vez mais as pessoas buscam viver nos centros urbanos, se beneficiando das vantagens resultantes da concentração organizada de atividades, trabalho, comércio, saúde, educação, lazer, moradia, segurança, serviços etc.

No entanto, a forma como essas atividades se distribuem pela cidade geram, também, externalidades negativas que resultam em sérios problemas urbanos, como gentrificação, exclusão social, constantes e extensos congestionamentos, dentre outros. Não à toa, a Organização das Nações Unidas, em 2015, conseguiu aprovar uma nova agenda de ação até 2030: os objetivos para o desenvolvimento sustentável das cidades. Essa nova agenda define os 17 objetivos que devem ser buscados pelas cidades para que seja possível acabar com a pobreza, promover a prosperidade e o bem-estar de todos, assim como proteger o meio ambiente e outros (ONU, 2015b).

Dentre esses objetivos globais, destaca-se a relevância dos transportes urbanos, posto que afetam diretamente a equidade social, o acesso a oportunidades de trabalho, educação e lazer, a qualidade de vida da população, o desenvolvimento econômico e o meio ambiente. Estudos comprovam o impacto do setor de transportes sobre o consumo de energia e poluição atmosférica, além de se constituir em uma das principais fontes de emissão de CO<sub>2</sub> que agrava as mudanças climáticas (CARVALHO, 2011; COLVILE et al., 2001; FAIZ, 1993; KRZYŻANOWSKI; KUNA-DIBBERT; SCHNEIDER, 2005).

Em trecho da carta do evento Rio+20, realizado em 2012, ressalta-se a importância dos transportes urbanos: “Sistemas de transporte eficientes e efetivos são importantes para [...] reduzir a pobreza. Os atuais padrões de desenvolvimento de transporte não são sustentáveis e **podem agravar problemas** ambientais e de saúde.” (ONU, 2012, grifo do autor).

Reconhecendo a importância do setor do transporte sobre os problemas urbanos, o governo brasileiro aprovou, em 2012, a lei 12.587, que estabelece diretrizes com vistas a quebrar o paradigma do desenvolvimento urbano orientado ao uso do automóvel. A nova Política Nacional de Mobilidade Urbana (PNMU) orienta os municípios brasileiros a planejarem o sistema de transporte de forma integrada ao

planejamento do uso do solo, respeitando as diretrizes estabelecidas no Plano Diretor existente, mas influenciando suas atualizações futuras.

Os conceitos fundamentais para essa quebra de paradigma, conforme definidos pela PNMU, podem ser entendidos como a priorização do transporte coletivo público e o incentivo a utilização de modos não motorizados (deslocamentos a pé e por bicicleta) em detrimento ao uso do automóvel. Destacam-se ainda mecanismos que proporcionem equidade social ao acesso dos serviços da cidade, como por exemplo, garantia de gratuidades, busca pela modicidade tarifária e acessibilidade universal; além da importância da participação da população na elaboração dos Planos.

Anteriormente, exigia-se a elaboração de Plano de Mobilidade somente para os municípios com mais de 500 mil habitantes. No entanto, uma cidade deste porte tipicamente já apresentaria problemas urbanos e de mobilidade, com a presença constante de congestionamento e um processo de expansão do seu território já consolidado. Assim, esse tipo de cidade não poderia mais reorientar seu modelo de urbanização e de circulação de forma preventiva, demandando soluções complexas, caras e demoradas, que provavelmente prejudicariam a qualidade de vida da população.

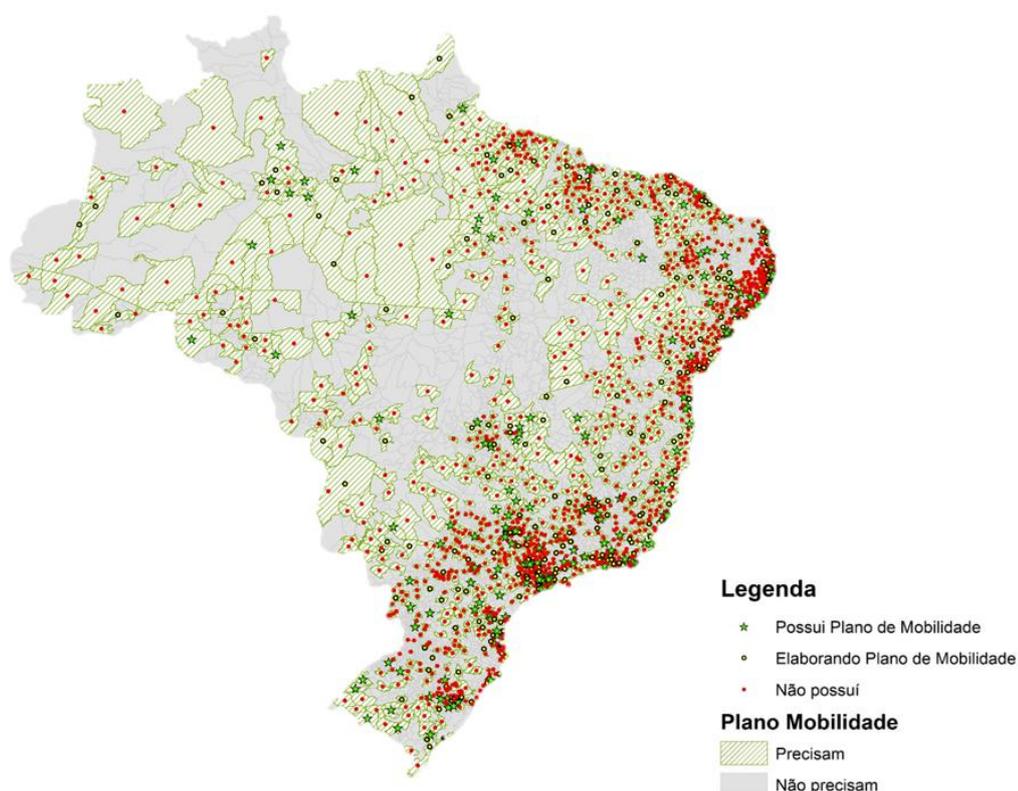
Reconhecendo esta falha, depois de onze anos sendo debatida, a nova política nacional foi aprovada com peso de lei, tornando obrigatória a elaboração de um Plano de Mobilidade Urbana Sustentável para todos os municípios com mais de 20 mil habitantes, como requisito para ter acesso a recursos federais destinados a investimentos em mobilidade urbana (BRASIL, 2012).

Embora a ideia de abordar o problema de forma preventiva se configure como uma estratégia interessante, o novo patamar estipulado pela lei representou um demasiado salto em relação ao padrão anterior, resultando no aumento brusco da quantidade de municípios que passaram a se enquadrar no perfil de exigência de cumprimento da Lei. Isso se configurou como um sério problema, tendo em vista a limitação da capacidade técnica disponível, tanto no setor público como no privado, para elaborar os planos de mobilidade respeitando o prazo de até três anos previsto em lei.

De fato, após o vencimento do prazo, limitado em 15 de abril de 2015, apenas cerca de 25% das cidades cumpriram sua obrigação administrativa. Ou seja, baseado nos dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2012 e com

informações atualizadas de 2015, dentre as 1.745 cidades enquadradas pelo novo patamar da lei, apenas 439 (25,1%) tinham ou estavam elaborando seus planos de mobilidade urbana. Ou seja, aproximadamente para cada quatro cidades brasileiras com mais de 20 mil habitantes, apenas uma cumpriu a obrigação da Lei 12.587/12.

Como pode ser visto na Figura 1, não é possível identificar um padrão espacial de concentração das cidades que atendem — ou não — à exigência da lei. Tampouco se pode identificar um padrão em função da quantidade de habitantes. Dentre os municípios que tem ou estão elaborando seu plano de mobilidade urbana, estão grandes cidades como São Paulo e Rio de Janeiro, mas também cidades pequenas, como, Armação dos Búzios, RJ e Campos do Jordão, SP, com pouco mais de 30 mil habitantes e 50 mil habitantes, respectivamente (IBGE, 2015).



**Figura 1: Panorama da situação dos Planos de Mobilidade Urbana no Brasil (adaptado de IBGE, 2012)**

Em função da importância do tema e das limitações que resultaram no reduzido número de cidades capazes de atender o prazo estipulado na lei, foi aprovada pela câmara dos deputados, no final de dezembro de 2015, uma proposta de revisão da Lei

de Mobilidade Urbana que, caso não seja vetada pelo Senado, prorrogaria em mais três anos o prazo de cumprimento: até abril de 2018.

De acordo com o relator da proposta de revisão, as principais dificuldades encontradas consistem na escassez de pessoal qualificado e na limitação de recursos financeiros (BRASIL, 2014).

Dessa forma, não só pela necessidade de mitigar as externalidades decorrentes dos problemas urbanos relacionados ao transporte, mas também pela obrigação legal de responsabilidade administrativa das prefeituras, acredita-se que muitos planos de mobilidade deverão ser elaborados nos próximos anos, principalmente a partir das eleições municipais de 2016. A elaboração desses planos é uma oportunidade para que seja possível mitigar os impactos decorrentes da rápida expansão urbana e dos problemas relacionados, direta e indiretamente, à mobilidade urbana.

## **1.1 Problema e Motivação**

A elaboração de um Plano de Mobilidade Urbana deve, ao seu fim, apresentar um Plano de Ações que identifique claramente os projetos e estratégias que devem ser executados para alcançar os objetivos estabelecidos para curto, médio e longo prazo (BRASIL, 2015).

Para investigar as medidas mais indicadas para atender determinados objetivos, diversos trabalhos na literatura acadêmica empregam métodos de modelagem da divisão modal, simulando o efeito de diferentes políticas públicas a partir da projeção agregada da demanda (AMOH-GYIMAH; AIDOO, 2013; BHAT; SARDESAI, 2006; DE JONG et al., 2003; EL-BANY et al., 2014; HESS et al., 2007).

No entanto, conforme apresentado como justificativa para a prorrogação do prazo de elaboração do Plano de Mobilidade Urbana, os recursos financeiros são escassos e limitados. Impossibilita-se, assim, a execução simultânea de todas as medidas identificadas como mais indicadas para atender aos objetivos estabelecidos no Plano de Mobilidade. Por isso, é fundamental que a etapa do Plano de Ações priorize as medidas que promovam resultados mais eficazes.

Porém, esses trabalhos se baseiam tipicamente na análise do sentido dos coeficientes estimados e na aplicação dos modelos para simular projeções agregadas. Dessa forma, avaliam apenas se as medidas promovem ou não resultado, sem levar

em consideração sua magnitude de contribuição. Consequentemente, negligenciam o nível de eficácia das medidas.

Alguns trabalhos chegam a captar a magnitude de resposta da propensão de utilização dos modos de transporte, realizando a análise da elasticidade direta, mas nenhum com o objetivo específico de listar os atributos que resultariam em medidas mais eficazes, apoiando o processo de tomada de decisão. Hammadou e Mahieux (2014) comparam a elasticidade obtida com a de outros trabalhos para destacar a singularidade da região de estudo, enquanto investigam como reduzir a utilização do automóvel em uma região degradada no Norte da França.

Román et al. (2014) estudam a escolha modal para obter maior conhecimento sobre a competitividade no mercado de transporte de passageiros entre Madri e Barcelona, as duas maiores cidades na Espanha. Com uma abordagem estratégica para ampliar a utilização do transporte aéreo, utilizam a análise de elasticidade somente para avaliar a variação da demanda frente à redução de tarifas. Identificam então esta demanda como inelástica e recomendam a adoção de outras medidas.

Já Sadri e Ukkusuri (2014) estudam o aprimoramento de planos de evacuação a partir da modelagem da escolha modal durante situação hipotética de evacuação de emergência em Miami Beach, nos EUA. Para isso, concentram-se na população mais vulnerável, que não possui alternativa própria de transporte, e projetam a parcela da população que dependeria de serviços públicos de socorro. No entanto, realizam análise da elasticidade direta apenas para inferir relação de causalidade entre atributos e a propensão de utilização de determinadas alternativas.

Assim, sem um procedimento metodológico quantitativo e transparente claramente definido, que seja capaz de apontar os atributos que exercem, direta e indiretamente, maior influência na escolha modal, coloca-se a seguinte questão: Como priorizar projetos e estratégias para melhorar a mobilidade urbana e torná-la mais sustentável?

Para responder essa pergunta, o problema principal que será estudado neste trabalho é justamente a lacuna metodológica para ranquear os atributos que exercem maior influência direta e cruzada sobre a escolha modal, o que permitiria identificar as medidas mais eficazes para promover a transferência modal desejada. Tal conhecimento apoiaria a etapa de priorização de investimento, permitindo melhor orientar a formulação de políticas públicas, que é a motivação dessa pesquisa.

## 1.2 Relevância do Problema

Se a formulação de políticas públicas não for orientada por um procedimento metodológico quantitativo e transparente capaz de priorizar os investimentos mais eficazes e eficientes, podem ocorrer duas situações: (1) os limitados recursos financeiros são aplicados em soluções que gerem baixo ou nenhum resultado, comprometendo a imagem do Plano de Mobilidade e da capacidade administrativa do gestor, enquanto os problemas existentes que não foram resolvidos tendem a se agravar; ou (2) os recursos são alocados corretamente em soluções que atendem às expectativas dos resultados esperados.

Na primeira situação, perde-se a oportunidade de aplicar corretamente os recursos financeiros, prejudicando a capacidade de mitigar as externalidades na medida em que os problemas urbanos se agravam. Além disso, compromete-se negativamente a imagem do Plano e da capacidade administrativa perante a sociedade, que é essencial para o sucesso de qualquer política pública que afete o cotidiano das pessoas.

Na segunda situação, os recursos são devidamente alocados, porém baseados em fatores como experiência, palpite ou mesmo, sorte. De qualquer forma, a ausência de um procedimento quantitativo e transparente-consultivo para embasar a priorização de investimentos reduz não só as chances de ocorrência de sucessos consecutivos, mas também o apoio de diferentes segmentos da população.

Se a oferta de transporte não consegue atender a demanda, ocorrem os congestionamentos. Com o agravamento dos problemas referentes à mobilidade urbana, observam-se impactos diretos nos aspectos econômicos, ambientais e sociais, como por exemplo, aumento do custo logístico, da poluição e dos tempos de deslocamentos, podendo resultar em piora na qualidade de vida da população (BANISTER; BERECHMAN, 2001; BANISTER, 2008; VAN WEE; ANNEMA; BANISTER, 2013).

Nesse sentido, o estudo realizado pela FIRJAN (2014) estima que os impactos dos congestionamentos nas regiões metropolitanas de São Paulo e Rio de Janeiro, durante o ano de 2013, provocaram custos na ordem de R\$ 98,4 bilhões – equivalente a 2% do PIB nacional. Como agravante, estudos na área de transportes e medicina correlacionam o aumento de doenças respiratórias com a maior exposição da poluição atmosférica oriunda dos congestionamentos (HEALTH EFFECTS INSTITUTE, 2010; MCADAM; STEER; PERROTTA, 2011).

Assim, destaca-se a relevância da elaboração de um Plano de Mobilidade Urbana de acordo com as diretrizes estabelecidas na PNMU para estimular o uso do transporte público e transporte não motorizado, em detrimento do uso do automóvel, e que seja feito de forma transparente e participativa para contar com o apoio da população, fundamental para o sucesso de qualquer plano urbano.

### **1.3 Objetivos Geral e Específicos**

O objetivo geral deste trabalho é comprovar o potencial de aplicação da modelagem da divisão modal para definir medidas que promovam a transferência modal desejada de forma mais eficaz e assim permitam apoiar a priorização de investimentos e determinação do plano de ações em sistemas de transportes, etapas obrigatórias em um Plano de Mobilidade Urbana Sustentável, melhor orientando a formulação de políticas públicas.

Para tal, alguns objetivos específicos se fazem necessários, sendo:

- ❖ Identificar as variáveis explicativas e os testes de desempenho mais utilizados para a modelagem da divisão modal;
- ❖ Elaborar um procedimento metodológico para modelar e posteriormente analisar, por meio de variáveis explicativas, o processo de escolha modal para viagens urbanas;
- ❖ Hierarquizar as variáveis explicativas por ordem de influência na escolha modal;
- ❖ Aplicar o procedimento metodológico proposto em um ensaio ilustrativo para exemplificar seu potencial na identificação das estratégias mais eficazes para promover a transferência modal desejada.

### **1.4 Justificativa**

Segundo Cascetta (2009), desde a década de 1980, os métodos de modelagem da divisão modal vêm sendo amplamente utilizados para apoiar e aprimorar o planejamento de transportes. De fato, inúmeros trabalhos utilizam modelos de escolha discreta para simular o impacto que a variação de atributos exerce sobre a divisão modal, avaliando diferentes medidas e políticas públicas para os mais diversos propósitos de viagem.

Palma e Rochat (2000) analisam os deslocamentos casa-trabalho na cidade de Genebra, concluindo que o tempo é o atributo mais importante para influenciar a escolha do modo de transporte. Assim, destacam que medidas voltadas a reduzir o tempo de viagem dos ônibus promoveriam maior atração de usuários para o sistema de transporte público do que políticas de isenção tarifária.

Cervero (2002) investiga a influência de características do ambiente construído sobre a escolha modal. Conclui empiricamente que a inclusão de atributos referentes à densidade, diversidade do uso do solo e infraestrutura urbana aprimoram o desempenho do modelo estimado, destacando a influência da qualidade das calçadas sobre a maior propensão de utilização do transporte público. Ao final, o autor destaca o potencial desse método para avaliar a importância relativa de cada atributo; promovendo maior conhecimento sobre o processo de escolha modal.

Su e Zhou (2012) avaliam o impacto de diferentes medidas de gerenciamento da demanda por viagens para desestimular o uso individual do automóvel, porém sob o ponto de vista da gestão corporativa. Concluem recomendando a redução da oferta de vagas no local de trabalho e o aumento da tarifa de estacionamento. Além disso, destacam que a bonificação financeira pode ser uma opção interessante para estimular que os funcionários passem a utilizar o transporte público em detrimento do carro. Ou seja, identificam alguns importantes atributos sobre o processo de escolha e recomendam medidas específicas para atender ao objetivo do estudo.

Já Ermagun e Samimi (2015) investigam como aumentar a utilização de transporte não motorizado por parte de jovens para deslocamentos até a escola, no Irã. Comparam o impacto da variação percentual na propensão de escolha de cada modo para identificar os atributos que exercem maior influência. Embora seja um dos trabalhos que mais avança na aplicação da análise de elasticidade para apoiar o planejamento de transportes, levando em consideração o nível de eficácia de diferentes atributos, ainda não promove sua classificação por ordem de influência, direta e/ou cruzada, no processo de escolha modal.

Identifica-se, então, uma interessante oportunidade de pesquisa para expandir o método de modelagem da divisão modal de forma a hierarquizar os atributos por ordem de influência no processo de escolha modal. Tal hierarquização possibilitaria identificar medidas mais eficazes para promover a transferência modal desejada (por exemplo, aumentando a participação de viagens de transporte público ou Transporte Não Motorizado).

Logo, essa expansão da modelagem da divisão modal poderia ser aplicada para apoiar a priorização de investimentos e a determinação de um plano de ações – etapas obrigatórias da elaboração de um Plano de Mobilidade Urbana Sustentável – contribuindo para melhor orientar a formulação de políticas públicas, que é a motivação da elaboração dessa pesquisa de dissertação.

Além disso, vale ressaltar que o caderno de referência para a elaboração de planos de mobilidade, desenvolvido em 2012 pelo Ministério das Cidades e atualizado em 2015, exige conteúdos mínimos em um Plano de Mobilidade (BRASIL, 2015). Dentre eles, destaca-se a realização de pesquisas e consultas para desenvolver o adequado diagnóstico da situação atual e prognóstico da situação futura. Esse processo engloba, dentre outros, a aplicação de questionários de caracterização das viagens realizadas por amostra representativa da população de estudo.

Ou seja, a própria elaboração de um Plano de Mobilidade promove a obtenção de um vasto banco de dados que contém informações sobre as características dos indivíduos e dos modos de transporte. Logo, acrescentando algumas perguntas ao questionário original, pode-se obter uma amostra representativa e rica em informações sobre atributos-chaves para a adequada modelagem da divisão modal. Assim, uma pequena complementação do questionário original viabilizaria a aplicação do procedimento metodológico proposto nessa dissertação.

Dessa forma, justifica-se também a necessidade de identificar os fatores-chave para a adequada modelagem da divisão modal. Principalmente as variáveis explicativas que devem ser acrescentadas ao questionário e os testes mais utilizados para avaliar o desempenho dos modelos.

## **1.5 Método de Pesquisa**

O desenvolvimento dessa pesquisa de dissertação adota a abordagem quantitativa que, segundo Fonseca (2002), consiste na utilização de linguagem matemática para reproduzir, de forma simplificada, um fenômeno de interesse e as relações entre suas variáveis. Essa abordagem permite realizar análises e inferir conclusões a partir da mensuração de dados; uma vez que a neutralidade no seu processo de obtenção seja garantida.

De acordo com os objetivos definidos, essa pesquisa pode ser considerada ainda como exploratória e de natureza aplicada:

- ❖ A pesquisa exploratória tem a finalidade de proporcionar maior familiaridade do pesquisador sobre o assunto, promovendo o aprimoramento de ideias e aplicações, além da identificação de lacunas e oportunidades de pesquisa (GIL, 2007). Principalmente quando não há conhecimento consolidado para a resolução de um problema específico;
- ❖ A pesquisa aplicada tem a finalidade de gerar conhecimentos que extrapolem o campo das ideias; podendo ser aplicados na prática para resolver problemas reais específicos (GERHARDT; SILVEIRA, 2009).

## 1.6 Premissas Adotadas

Existam diferentes métodos para modelar o comportamento individual de viagens (do inglês, *Travel behavior*), como por exemplo, os modelos de escolha discreta, hedônico, heurístico e derivados da teoria da inteligência artificial. No entanto, essa pesquisa se concentra no primeiro tipo: os modelos de escolha discreta.

Esses modelos se baseiam na teoria econométrica de maximização da função utilidade aleatória percebida. Assim, tanto em função do método utilizado como por motivos de simplificação, é importante definir claramente as premissas adotadas para a elaboração dessa pesquisa:

- ❖ Variáveis explicativas, como as características do indivíduo, dos modos de transporte e do ambiente podem ser empregadas em um procedimento metodológico para modelar o processo de escolha modal para viagens urbanas;
- ❖ Todo indivíduo é um ser racional e, portanto, toma decisões com o objetivo de maximizar a sua utilidade percebida;
- ❖ As decisões que implicam no maior consumo de tempo e dinheiro impactam negativamente na utilidade percebida pelas pessoas; e
- ❖ O modo de transporte selecionado por um indivíduo para o trajeto de ida é considerado o mesmo para o trajeto de volta.

## 1.7 Limitação da Pesquisa

Para atender um dos objetivos específicos, essa dissertação aplica, ao final, o procedimento metodológico proposto em um ensaio ilustrativo, com o intuito de

exemplificar seu potencial analítico. O banco de dados utilizado para esse propósito – já existente – foi obtido a partir da pesquisa de transporte multipropósito aplicada à Cidade Universitária, realizada em 2015 por Carvalho (2016) – e pode ser encontrado abertamente no repositório público apresentado no Anexo C.

A pesquisa *online* é do tipo Preferência Revelada (PR), logo incorpora as limitações intrínsecas desta técnica. Ou seja, não é possível avaliar a decisão dos entrevistados mediante situações hipotéticas, seja referente à elaboração de cenários com distintas combinações das principais variáveis de análise, ou mesmo frente a alternativas (modos de transporte) não existentes/disponíveis no local, naquele momento.

Outra limitação característica de uma PR é que o banco de dados consiste em decisões que foram tomadas sem que houvesse o pleno conhecimento sobre os atributos de todas as alternativas disponíveis. Tipicamente, o entrevistado conhece em mais detalhes o modo de transporte que utiliza com maior frequência, podendo até mesmo, em alguns casos, conhecer também atributos de outros modos, porém dificilmente conhece os de todas as alternativas disponíveis. Por isso, este tipo de pesquisa não possibilita mensurar, com significativo nível de sensibilidade, a influência que a variação marginal de determinado atributo exerceria sobre a probabilidade de escolha de uma alternativa, para cada indivíduo.

Para suprir ambas as limitações, seria necessário empregar a técnica de Pesquisa Declarada (PD) para substituir ou, preferencialmente, complementar a PR. Neste último caso, caberia adotar cuidados especiais para tratar as diferentes escalas dos dados coletados e permitir sua consolidação em um único banco de dados, do tipo PR-PD.

De qualquer forma, os métodos de escolha discreta podem ser aplicados sobre banco de dados obtidos a partir de PR, PD ou da combinação de ambas (PR-PD). As limitações aqui apresentadas, portanto, não se configuram como uma barreira impeditiva para o adequado desenvolvimento dessa pesquisa, nem para a aplicação dos métodos selecionados, assim como também não impedem a verificação dos resultados e não interferem nas conclusões extraídas.

## **1.8 Estrutura da Pesquisa**

Este capítulo 1 apresenta a contextualização do problema que motivou a pesquisa, sua relevância e os objetivos propostos para a dissertação.

O capítulo 2 apresenta o mínimo ferramental teórico que entendemos possa garantir a compreensão e replicação da pesquisa elaborada. Introduce e comenta os principais modelos utilizados para a modelagem da escolha discreta. Apresenta também o que se entende por boas práticas com o intuito de desmistificar o tema, facilitando e ampliando sua utilização em casos reais e em desenvolvimento de pesquisas.

O capítulo 3 apresenta a revisão sistemática realizada, detalhando a estratégia de busca adotada; analisando trinta e dois artigos científicos selecionados e comentando os fatores-chave utilizados com maior frequência para o adequado desenvolvimento da modelagem da divisão modal. O quadro que discrimina os fatores-chave por artigo revisado é apresentado no Anexo A.

Já o capítulo 4 apresenta a sequência de atividades realizada para a elaboração deste trabalho, comentando as etapas necessárias para o tratamento dos dados e apresentando uma alternativa para a estimação de atributos referente às características da viagem (independente do valor informado pelos entrevistados). Além disso, apresenta também a representação esquemática do procedimento proposto, elaborada com vistas a facilitar seu entendimento e daí sua replicação e aperfeiçoamento futuros.

O Capítulo 5 apresenta a descrição estatística dos dados utilizados na modelagem da divisão modal e exemplifica o potencial do procedimento metodológico proposto. Interpretam-se os resultados obtidos pelo cálculo do Valor do Tempo e da elasticidade. Classifica os atributos por ordem de influência no processo de escolha modal e apresenta como este procedimento de hierarquização permite apoiar o processo de priorização de investimentos em um Plano de Mobilidade, melhor orientando a formulação de políticas públicas. Ao final, simula diferentes cenários e avalia o impacto da variação de determinados atributos sobre a projeção da divisão modal.

As principais conclusões e recomendações futuras são destacadas no Capítulo 6. Indica como os objetivos da pesquisa foram atendidos e apresenta observações especiais resultantes do desenvolvimento da pesquisa. Por último, sintetiza as recomendações para trabalhos futuros.

## 2 MODELOS DE ESCOLHA DISCRETA

Este capítulo consiste na contextualização teórica do método utilizado no desenvolvimento dessa pesquisa. Apresenta-se o ferramental teórico mínimo necessário para sua compreensão, além de comentários explicativos e boas práticas recomendadas para a modelagem da divisão modal. Tem o intuito de desmistificar o tema e facilitar/ampliar a utilização dessa técnica que, embora consolidada academicamente, ainda é pouco utilizada em aplicações práticas.

De fato, os modelos de escolha discreta têm sido amplamente utilizados em estudos de economia, transportes e marketing. Eles possibilitam descrever, explicar e até mesmo prever o comportamento de indivíduos mediante processos de escolhas entre alternativas específicas, como por exemplo, a marca de refrigerante a ser comprado ou o modo de transporte utilizado para chegar ao trabalho.

Muitos estudos foram realizados nos últimos 40 anos, utilizando modelos de escolha discreta para determinar a probabilidade de escolha de determinado modo de transporte em função de variáveis explicativas (BEN-AKIVA; LERMAN, 1985; BEN-AKIVA et al., 1997; MCFADDEN; TYE; TRAIN, 1977; MCFADDEN, 1974, 1987; PAULSSEN et al., 2014; REVELT; TRAIN, 1998; VAN CAN, 2013).

De acordo com Cascetta (2009), os modelos de escolha discreta são importantes para o planejamento de transporte, pois permitem obter representações detalhadas dos diferentes e complexos aspectos da demanda de transporte, respaldados por fortes justificativas teóricas.

Esses modelos são considerados desagregados, pois se baseiam no nível individual de escolha. Isto é, a escolha de cada participante (e suas respectivas características) representa um ponto de dados a ser analisado. Destacam-se algumas peculiaridades que os tornam superiores aos modelos de nível agregado:

- ❖ Por serem estáveis, podem ser aplicados/expandidos no tempo e espaço (para a previsão de cenários futuros e tendências ou podendo ser replicados em regiões similares);
- ❖ São mais eficientes, tendo em vista que requerem quantidade significativamente menor de dados a serem coletados (conforme comentado anteriormente, cada participante da pesquisa é analisado como um ponto de observação, ao invés de se agrupar vários para obter apenas um valor representativo de determinado grupo);

- ❖ As variáveis explicativas utilizadas no modelo têm coeficientes estimados explicitamente, possibilitando avaliar o sentido e a magnitude de influência de cada parâmetro adotado no modelo.

Segundo Órtúzar e Willumsen (2011), os modelos de escolha discreta apresentam o seguinte postulado: “A probabilidade de um indivíduo escolher uma determinada alternativa é uma função de suas características socioeconômicas e da atratividade relativa da alternativa”.

Essa “atratividade relativa” é representada pelo conceito microeconômico da Utilidade Percebida, enquanto o processo de escolha de cada indivíduo é regido pela maximização da utilidade percebida referente a cada alternativa.

No entanto, antes de especificar a estruturação e características de aplicação dos principais modelos de escolha discreta, é importante aprofundar o conhecimento sobre a teoria de qual derivam: a Utilidade Aleatória.

## 2.1 Teoria da Utilidade Aleatória

A partir de Koppelman e Bhat (2006), Cascetta (2009), Train (2009) e Órtúzar e Willumsen (2011), apresentam-se detalhes comentados sobre a teoria da utilidade aleatória com o intuito de desmistificar o tema, que fundamenta os modelos de escolha discreta.

Esta teoria apresenta o seguinte postulado:

- ❖ Os indivíduos são seres racionais e, portanto, tomam decisões selecionando alternativas que maximizem sua utilidade líquida pessoal;
- ❖ Existe um grupo de alternativas disponíveis  $A = \{A_1, \dots, A_j, \dots, A_N\}$  e um grupo vetorial de atributos observados  $X$ . Cada indivíduo  $i$  possui um grupo particular de atributos  $x \in X$  e, de forma geral, um grupo específico de alternativas disponíveis  $A(i) \in A$ . Assim, o grupo de alternativas disponíveis pode variar para cada indivíduo, por exemplo: se uma pessoa não possui habilitação para dirigir, a alternativa “conduzir automóvel” não deve estar disponível no seu grupo de escolha;
- ❖ Cada alternativa  $A_j \in A$  tem associada uma utilidade líquida  $U_{ij}$  para cada tomador de decisão  $i$ . Como o modelador não controla o

experimento, mas apenas observa o processo de escolha de cada indivíduo, a função utilidade  $U_{ij}$  pode ser expressa por duas componentes, sendo:

- a. Uma componente sistemática  $V_{ij}$  que é composta pelos atributos observados  $x$ ; e
- b. Uma componente aleatória  $\varepsilon_{ij}$  que representa as particularidades de cada indivíduo (variações de gosto), os erros relacionados à mensuração dos atributos e das observações feitas pelo modelador.

Destaca-se a importância da inclusão da componente aleatória  $\varepsilon_{ij}$ , descrita no terceiro item do postulado, pois permite explicar dois fenômenos que aparentemente seriam “irracionais” no processo de escolha: 1) duas pessoas, com características socioeconômicas idênticas, expostas às mesmas restrições e com acesso ao mesmo grupo de opções podem escolher alternativas distintas; e ainda, 2) uma pessoa pode não escolher aquela que seria, a partir da influência dos atributos, esperada como a melhor alternativa.

Assim, a função utilidade de uma alternativa  $A_j$  percebida por um indivíduo  $i$  é:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad (1)$$

A componente sistemática  $V_{ij}$  pode assumir diversas formas diferentes. No entanto, por questões de conveniência analítica e estatística, adota-se tipicamente a forma linear aditiva dos atributos  $x \in X$ , dada pela Equação 2. Os parâmetros  $\beta$  são constantes para todos os indivíduos (por isso o índice  $i$  não aparece no primeiro termo), mas podem variar para as diferentes alternativas  $A_j \in A$ .

$$V_{ij} = \sum_k \beta_{jk} \cdot x_{ijk} \quad (2)$$

Conforme visto anteriormente, o indivíduo  $i$  escolhe a alternativa  $A_j$  com maior utilidade associada, como representado na Equação 3:

$$U_{ij} \geq U_{im}, \forall A_m \in A(i) \quad (3)$$

Aplicando a Equação 1 na Equação 3, temos que:

$$(V_{ij} + \varepsilon_{ij}) \geq (V_{im} + \varepsilon_{im}) \forall A_m \in A(i) \quad (4)$$

Como as componentes de erro  $\varepsilon$  não são conhecidas, não é possível afirmar com certeza qual a alternativa  $A_j$  que resulta na maior utilidade percebida  $U_{ij}$ . Por isso utiliza-se o conceito de probabilidade, como apresentado na Equação 5:

$$P_{ij} = Prob[(V_{ij} + \varepsilon_{ij}) \geq (V_{im} + \varepsilon_{im}), \forall A_m \in A(i)] \quad (5)$$

Essa equação pode ser reescrita, isolando  $\varepsilon_{im}$ , ajustando-a para a estrutura formal de probabilidades:

$$P_{ij} = Prob[\varepsilon_{im} \leq \varepsilon_{ij} + (V_{ij} - V_{im}), \forall A_m \in A(i)] \quad (6)$$

Logo, se a distribuição dos erros for conhecida, torna-se possível calcular a probabilidade de escolha da alternativa  $A_j$ , que resultaria na maior utilidade percebida.

Assim, os modelos de Escolha Discreta são construídos a partir da Equação 6, arbitrando uma determinada distribuição para os erros  $\varepsilon$  para que seja possível, então, obter uma função de probabilidade calculável. Os principais modelos de Escolha Discreta são apresentados na seção 2.2. Porém, destacam-se primeiramente algumas particularidades da Teoria da Utilidade Aleatória.

### 2.1.1 Particularidades

Para que a Equação 1 seja válida, é necessário que haja certa homogeneidade na amostra de estudo, determinando que todos os indivíduos estejam sujeitos às mesmas restrições e tenham acesso ao mesmo grupo de alternativas (*choice set*). No entanto, nem sempre essa condição é aplicável. Para resolver essa limitação de forma simples, adota-se o conceito do Grupo Universal de Opções para todos os indivíduos, descrito no item 2.1.2.

A construção da componente sistemática  $V_{ij}$  requer que o modelador assuma comportamentos específicos para os diferentes atributos a serem estimados, que podem ainda variar para cada alternativa  $j$ . Esse assunto é abordado no item 2.1.3. Como a complexidade das técnicas de busca para a determinação da forma e tipo de atributos das diferentes alternativas a serem estimados foge do objetivo desta pesquisa, recomenda-se que mais informações sejam obtidas em Órtúzar e Willumsen (2011), especificamente no capítulo 8.

Por último, analisando a Equação 6, chama-se atenção para seu o último termo, dado por  $(V_{ij} - V_{im})$ . Pode-se notar então que, para o cálculo da probabilidade de escolha de uma alternativa, a parcela importante é a diferença das utilidades e não o valor absoluto de cada uma.

Destaca-se então que a Utilidade é adimensional e, para que os atributos de diferentes unidades (como tempo e distância) possam ser somados, seus coeficientes devem assumir automaticamente unidade inversa, como  $tempo^{-1}$  e  $distância^{-1}$ . Por esse motivo, a magnitude dos coeficientes não apresenta qualquer significado, devendo apenas ser avaliado seu sentido de contribuição (se positivo ou negativo).

Assim, para avaliar a magnitude de contribuição, deve-se obter a elasticidade direta (ou cruzada) dos atributos, ou seja, a variação marginal da probabilidade de escolha de uma alternativa em decorrência da variação marginal do valor de um atributo. As elasticidades direta e cruzada são apresentadas no item 2.1.4.

### 2.1.2 Grupo Universal de Opções

O Grupo Universal de Opções é uma solução simples para atender a exigência de homogeneidade dentre os indivíduos do estudo. Conhecida também como “abordagem implícita”, consiste em adicionar ao modelo uma variável *dummy* de disponibilidade, que assume valor zero ou um, referente a cada uma das alternativas (nesse caso, modo de transporte).

Por exemplo, se uma pessoa não possui carteira de habilitação para conduzir veículo automotor, verificado pelo atributo *dummy*  $x_{habilitação}$ , a alternativa “conduzir automóvel” não deve fazer parte de seu Grupo de Opções disponíveis. Nesse caso, define-se o critério de disponibilidade  $A_{i,auto}$  de acordo com a função:

$$A_{i,auto} = \begin{cases} 1 & \text{se } x_{i,habilitação} = \text{"Sim"} \\ 0 & \text{se } x_{i,habilitação} = \text{"Não"} \end{cases} \quad (7)$$

A Equação 8 pode ser generalizada, determinando critérios específicos de disponibilidade para cada uma das alternativas. Ou seja, caso a alternativa  $j$  esteja disponível para o indivíduo  $i$ , a variável *dummy* de disponibilidade  $A_{ij}$  assume valor um; caso contrário, assume o valor zero.

Assim, adotando a solução do Grupo Universal de Opções, pode-se reescrever a Equação 3 somando-se a função  $\log$  das variáveis de disponibilidade especificadas:

$$P_{ij} = Prob[(U_{ij} + \log A_{ij}) \geq (U_{im} + \log A_{im}), \forall A_m \in A(i)] \quad (8)$$

Se a alternativa não for disponível,  $A_{ij}$  assume valor zero. Como  $\log(0)$  é uma operação não definida, o primeiro termo da Equação 8 não pode ser calculado, resultando em probabilidade zero; conforme esperado. No entanto, caso a alternativa seja disponível,  $A_{ij}$  assume valor um. Como o  $\log(1)$  é igual a zero, a variável de controle  $A_j$  não exerce nenhuma influência sobre a utilidade da alternativa  $U_{ij}$ , permitindo calcular normalmente a probabilidade de escolha dessa alternativa.

Logo, pode-se especificar qualquer modelo de escolha discreta adotando o conceito do Grupo Universal de Opções. Para isso, basta que o modelador determine o critério de disponibilidade e especifique a adição ao modelo do termo  $\log(A_{ij})$  para as alternativas que se deseje controlar.

### 2.1.3 Especificação da Componente Sistemática $V_{ij}$

Os atributos podem ser de três tipos diferentes: 1) relacionados às características socioeconômicas (ou do indivíduo), como por exemplo, renda, idade, gênero, quantidade de carros no domicílio e se possui ou não carteira de habilitação; 2) relacionados às características do modo (ou nível de serviço), como por exemplo, frequência, tempo de viagem total, parcela do tempo de viagem em espera, parcela do tempo de viagem dentro do veículo ou custos; e 3) relacionados às características do ambiente (ou uso do solo), como por exemplo, local de origem, presença de estações de transporte público até 800 metros de distância, presença de ciclovia, extensão da malha ciclável ou a quantidade de lojas no local de destino.

Os atributos  $x$  podem ser, basicamente, de três tipos:

- ❖ Genérico: quando o atributo aparece na função utilidade de duas ou mais alternativas  $j$  e seus coeficientes  $\beta$  apresentam a mesma forma e valor; sendo:  $\beta_{1k} = \beta_{2k} = \beta_k$ . Por exemplo, admitindo-se que os atributos Tempo e Custo de Viagem exercem o mesmo impacto sobre as utilidades percebidas, independente da alternativa;
- ❖ Específico: quando o atributo aparece somente na função utilidade de uma alternativa específica  $j$ . Por exemplo, a Tarifa de Estacionamento

só se aplica à alternativa “Carro”. Ou ainda, quando um atributo se aplica a mais de uma alternativa, porém seu coeficiente  $\beta_j$  apresenta forma e/ou valor específico para cada uma. Por exemplo, o tempo de espera pelo transporte público se aplica às alternativas “ônibus” e “metrô”, mas se o nível de serviço de um for superior ao outro, o atributo deve ser tratado com coeficientes específicos para cada alternativa, ao invés de forma genérica; ou ainda

- ❖ Constante: caso especial do tipo “Específico”, conhecido como *Alternative Specific Attribute* (ASA), que pode ser entendido como um termo constante que representa a “preferência modal”. O atributo artificial *dummy*  $x_j$  assume valor 1 para a alternativa  $j$  e 0 para as demais. Assim, seu coeficiente  $\beta_j$ , denominado como *Alternative Specific Constant* (ASC), atua como um termo constante na componente sistemática  $V_j$ . Esse atributo pode ser interpretado como a diferença entre a utilidade média e a porção que é explicada pela contribuição dos demais atributos. Logo, quanto maior for o valor deste termo Constante, menor é o poder de explicação decorrente da contribuição das variáveis explicativas adotadas no modelo, e vice-versa.

Conforme descrito anteriormente, a partir da Equação 6, pode-se observar que a parcela importante para a probabilidade de escolha de uma alternativa é a diferença entre as utilidades ( $V_{ij} - V_{im}$ ) e não o valor absoluto de cada uma delas. A partir dessa propriedade da Teoria da Utilidade Aleatória e das características dos atributos, destacam-se algumas considerações importantes para a especificação da Componente Sistemática:

- ❖ Se todas alternativas apresentarem um atributo Específico, não é possível determinar todos os coeficientes, sendo necessário que uma das alternativas assuma valor nulo. Essa alternativa passa a ser denominada “base” ou “referência”, enquanto os demais atributos passam a ser interpretados como a diferença de utilidade relativa à alternativa base;
- ❖ O mesmo vale para o caso especial de atributos Constantes. Se todas as  $J$  alternativas possuírem um atributo Constante, é necessário adotar a estratégia de uma “Alternativa Base” para que seja possível estimar os  $(J - 1)$  coeficientes;

- ❖ Somente os atributos Genéricos, com um coeficiente  $\beta_k$ , podem estar presentes em todas as alternativas, contanto que o valor dos atributos não seja o mesmo dentre as alternativas, isto é,  $x_{1k} \neq x_{2k} \neq x_{3k}$ . Caso contrário, a diferença entre as utilidades resultaria em um coeficiente  $\beta_k$  multiplicando um fator nulo, o que impossibilitaria estimar o coeficiente  $\beta_k$ . Por exemplo, admitindo que o Tempo de Viagem seja um atributo que exerce igual influência sobre todas as alternativas (Carro, Bicicleta e Ônibus), é importante verificar que os tempos de viagem referentes a cada alternativa não sejam idênticos. Se forem ( $T_c = T_b = T_o = 15 \text{ min}$ ), passa a ser necessário atribuir o Tempo de Viagem como atributo Específico para cada uma das alternativas, ou adotar a estratégia da “Alternativa Base”.

#### 2.1.4 Derivada e Elasticidade: Direta e Cruzada

Os modelos de escolha discreta são representações da probabilidade de escolha de uma alternativa em função dos valores dos atributos que compõem sua utilidade percebida.

Assim, uma significativa contribuição deste método resulta da análise da variação da probabilidade de escolha de determinada alternativa,  $P_j$ , em decorrência da variação do valor de determinado atributo  $x_{jk}$ , para todos os indivíduos. Destaca-se que essa variação é considerada de forma agregada, calculada pela média das variações individuais ou pela variação aplicada nos valores médios; por isso o índice  $i$  não aparece em suas equações.

Para exemplificar, tal análise permite determinar qual seria a redução média esperada na probabilidade de uso do automóvel em decorrência do aumento do custo de estacionamento; ou o aumento na probabilidade de utilização do transporte público em função da melhoria do serviço, promovendo redução de 10 minutos do tempo de viagem.

Para obter a sensibilidade desta “taxa de resposta” da probabilidade em função da variação unitária de um atributo, calcula-se a derivada da função probabilidade em relação ao atributo em questão. Aplicando a regra da cadeia, obtém-se a Equação 9:

$$\frac{\partial P_j}{\partial x_{jk}} = \frac{\partial V_j}{\partial x_{jk}} \cdot P_j \cdot (1 - P_j) \quad (9)$$

No caso em que a componente sistemática da utilidade tem forma linear aditiva, como por exemplo,  $V_j = \beta_0 + \beta_1 \cdot x_{j1} + \beta_2 \cdot x_{j2} + \dots + \beta_k \cdot x_{jk}$ , pode-se ver claramente que  $\frac{\partial V_j}{\partial x_{jk}} = \beta_k$ , ou seja, equivalente ao valor do coeficiente do atributo  $x_{jk}$ . Assim, a Derivada Direta é dada pela Equação 10:

$$\frac{\partial P_j}{\partial x_{jk}} = \beta_k \cdot P_j \cdot (1 - P_j) \quad (10)$$

Observa-se que a Derivada Direta é proporcional ao sentido do coeficiente  $\beta_k$ . Logo, uma variação positiva no atributo  $x_{jk}$  resulta no aumento da probabilidade de escolha da alternativa  $j$ , caso seu coeficiente  $\beta_k$  seja positivo; e vice-versa.

Para verificar a influência da variação de um atributo  $x_{jk}$  relativo à alternativa  $j$  sobre a probabilidade de escolha das demais alternativas  $P_m$  (excluindo a alternativa  $j$ ), tem-se que:

$$\frac{\partial P_m}{\partial x_{jk}} = -\beta_{jk} \cdot P_m \cdot P_j, \quad \forall m \neq j \quad (11)$$

Ou seja, a Derivada Cruzada é dada pela Equação 11.

Nesse caso, observa-se que a Derivada Cruzada é inversamente proporcional ao sentido do coeficiente  $\beta_{jk}$ . Logo, uma variação positiva no atributo  $x_{jk}$  resulta na redução da probabilidade de escolha das demais alternativas (exceto da alternativa  $j$ ), caso seu coeficiente  $\beta_{jk}$  seja positivo; e vice-versa.

No entanto, é importante destacar que a Derivada se refere apenas à sensibilidade de variação da função probabilidade em relação à variação unitária de um determinado atributo. Para que seja possível quantificar essa “taxa de resposta”, é preciso normalizar a Derivada Direta (ou Cruzada) em relação à unidade do atributo em si, obtendo então a Elasticidade Direta (ou Cruzada).

Dessa forma, a Elasticidade é definida como a variação percentual da probabilidade de escolha de uma alternativa em decorrência da variação percentual de um atributo. De acordo com Koppelman e Bhat (2006), para eliminar os problemas decorrentes da ambiguidade de determinação do ponto de referência para a normalização (se no valor inicial, final ou a média da diferença), recomenda-se adotar variações muito pequenas, analisando de forma pontual.

Assim, considerando um intervalo de variação infinitesimal, obtém-se a Elasticidade Pontual a partir da Derivada obtida anteriormente:

$$E_{d_{x_{jk}}}^{P_j} = \frac{\partial P_j}{\partial x_{jk}} \cdot \left( \frac{x_{jk}}{P_j} \right) \quad (12)$$

Aplicando então a Equação 10, elimina-se o termo  $P_j$ , obtendo a Elasticidade Direta, dada pela Equação 13:

$$E_{d_{x_{jk}}}^{P_j} = \frac{\partial P_j}{\partial x_{jk}} \cdot \left( \frac{x_{jk}}{P_j} \right) = [\beta_k \cdot (1 - P_j)] \cdot x_{jk} \quad (13)$$

Pode-se reparar que, diferentemente da Derivada Direta, a Elasticidade é função de um termo adicional, relacionado ao próprio valor do atributo analisado,  $x_{jk}$ .

De forma similar, a Elasticidade Pontual Cruzada é obtida a partir da Derivada Cruzada, multiplicando a Equação 11 pelo termo  $(x_{jk}/P_m)$ .

$$E_{c_{x_{jk}}}^{P_m} = \frac{\partial P_m}{\partial x_{jk}} \cdot \left( \frac{x_{jk}}{P_m} \right) = -\beta_{jk} \cdot P_m \cdot x_{jk} , \forall m \neq j \quad (14)$$

A Elasticidade Cruzada é dada pela Equação 14.

Um importante fato a se observar é que a diferença total entre as Derivadas ou Elasticidades deve sempre ser nula, tendo em vista que a soma das probabilidades de escolha das alternativas não pode ser superior a 100%. Isto é, para que alguma alternativa apresente aumento na sua probabilidade de escolha é imprescindível que haja redução de outras. De acordo com Train (2009), embora essa propriedade seja simples e intuitiva, muitos modeladores cometem o erro de construir um modelo admitindo que, por exemplo, a utilização do transporte coletivo por ônibus aumente, porém sem afetar a utilização do automóvel ou de transportes não motorizados.

Por motivo de simplicidade analítica, a formulação da Elasticidade apresentada se baseia no caso mais simples (Multinomial Logit, utilizado nessa pesquisa), no qual se assume a independência entre as alternativas. Assim, a fórmula final da Elasticidade Direta (ou Cruzada) varia para os demais modelos (que se baseiam em premissas mais complexas), embora o seu cálculo se baseie no mesmo raciocínio detalhado no item 2.1.4.

É importante ressaltar que o procedimento comentado anteriormente, referente à Elasticidade Pontual, se aplica apenas para atributos contínuos, como renda, custo e tempo de viagem; uma vez que, por definição, variáveis de natureza

discreta não podem ser derivadas. Para esses casos, adota-se o conceito Elemental da elasticidade, computando a proporção da variação incremental da probabilidade de uma alternativa sobre a proporção da variação unitária do atributo de natureza discreta, ambas referentes ao valor intermediário das respectivas diferenças (KOPPELMAN; BHAT, 2006). Tal procedimento é denominado de Elasticidade no Arco (ou no Ponto Médio) e definido matematicamente na Equação 15.

$$Elasticidade - Arco = \frac{\frac{(P_2 - P_1)}{(P_1 + P_2)/2}}{\frac{(X_2 - X_1)}{(X_1 + X_2)/2}} \quad (15)$$

Assim, os conceitos da Derivada e Elasticidade (direta ou cruzada) se aplicam a todos os modelos de escolha discreta, configurando-se como importantes ferramentas para a análise do impacto na divisão modal decorrente de mudanças em determinadas variáveis explicativas (resultantes, por exemplo, da implantação de políticas públicas específicas), assim como sua magnitude de influência sobre o processo de escolha.

Uma vez que a teoria da utilidade tenha sido fundamentada e algumas de suas peculiaridades devidamente comentadas, apresenta-se a seguir, em maiores detalhes, os procedimentos necessários para a construção dos principais modelos de escolha discreta.

## 2.2 Detalhamento dos Principais Modelos

Apresentam-se, a seguir, as suposições adotadas para a especificação dos principais modelos de escolha discreta, assim como suas características, vantagens e limitações, além das situações recomendadas de aplicação.

### 2.2.1 Multinomial Logit (MNL)

Conforme visto na Equação 6, a construção de um modelo de escolha discreta depende de suposições arbitradas em relação às componentes de erro,  $\varepsilon$ . O modelo mais simples, MNL, é construído a partir das seguintes suposições sobre a distribuição das componentes de erro:

- ❖ Apresentam distribuição Gumbel;

- ❖ Idênticas e independentemente distribuídas entre as alternativas;
- ❖ Idênticas e independentemente distribuídas entre os indivíduos.

Embora o procedimento comum para tratar dados desconhecidos seja admitir uma distribuição do tipo Normal, para o caso específico do cálculo de probabilidades essa suposição não resulta em uma função de forma fechada, o que demandaria significativa capacidade de processamento computacional para sua solução. De fato, os modelos Probit são construídos a partir dessa suposição de distribuição e são apresentados em maiores detalhes no subitem 2.2.4.

A distribuição Gumbel admitida é similar à Normal, porém apresenta funções densidade e de distribuição acumulada de probabilidades mais simples, que resultam em uma forma fechada para o cálculo da probabilidade. Isso significa dizer que a probabilidade pode ser obtida a partir de uma equação simples e de interpretação intuitiva, ao invés de necessitar de integração numérica multi dimensional (Probit) ou de processos de simulação aleatória (Mixed Logit).

Assim, a partir da suposição feita, a função da probabilidade de escolha de uma alternativa  $j$  ( $j = 1, 2, \dots, m$ ) é dada pela Equação 16.

$$P_j = \frac{\exp(V_j)}{\sum \exp(V_m)}, j = (1, 2, \dots, m) \quad (16)$$

#### 2.2.1.1 Binomial Logit

No caso especial em que o Grupo de Opções seja composto por apenas duas alternativas, o modelo MNL colapsa em uma nova estrutura, conhecida como Binomial Logit. Com as mesmas propriedades e restrições do modelo mais amplo, a sua função probabilidade é ainda mais simples, dada pela Equação 17:

$$P_a = \frac{1}{1 + \exp(V_b - V_a)} \quad (17)$$

#### 2.2.1.2 Independência entre Alternativas Irrelevantes

Em decorrência das suposições arbitradas para sua construção, principalmente o segundo e terceiro itens, o modelo MNL apresenta uma propriedade restritiva conhecida como “Independência entre Alternativas Irrelevantes” (do acrônimo em

inglês, *IIA*). Essa propriedade define que, para qualquer indivíduo, a proporção entre a probabilidade de escolha de duas alternativas específicas é independente da existência de uma terceira alternativa (e seus atributos).

Essa propriedade pode ser vista como uma vantagem, pois possibilita a inclusão ou remoção de alternativas sem afetar o modelo que já tenha sido estimado. Assim, o MNL costuma ser utilizado para projetar a divisão modal decorrente da inclusão de uma nova alternativa, como por exemplo, o impacto na divisão modal em função da construção de um novo modo de transporte.

Por outro lado, essa propriedade é restritiva, pois o modelo falha caso haja correlação entre as alternativas disponíveis. No clássico exemplo do “paradoxo do ônibus azul/vermelho”, a probabilidade de escolha do automóvel é reduzida em decorrência da inclusão de uma terceira alternativa (o ônibus de cor azul), embora ela apresente as mesmas características da alternativa já existente (ônibus vermelho).

Nesse caso, a simples alteração da cor do veículo não afeta a percepção dos indivíduos perante o nível de serviço das alternativas iniciais, “carro” e “ônibus”, provando que se as alternativas forem correlacionadas, o MNL não deve ser aplicado (MCFADDEN, 1974).

É importante observar que, em decorrência dessa independência entre alternativas, no caso específico do MNL a Elasticidade Cruzada apresenta valores idênticos para a variação da probabilidade de escolha das demais alternativas em função da variação de um atributo.

Conclui-se então que o modelo Multinomial Logit deve ser aplicado somente para os casos em que as alternativas disponíveis sejam significativamente distintas entre si, podendo ser ponderadas de forma clara e independente pelos indivíduos.

### 2.2.2 Nested Logit (NL)

O modelo Nested Logit foi desenvolvido justamente para superar a restrição imposta pela propriedade *IIA*, relaxando-a parcialmente para permitir tratar a correlação entre alternativas, porém ainda apresentando uma forma fechada da função probabilidade, que simplifica seu cálculo.

Esse modelo parte da suposição que algumas alternativas são mais similares do que outras, podendo ser agrupadas, apresentando assim termos em comum nas

componentes de erros. Essa “similaridade” que permite agrupar determinadas alternativas é capturada pela covariância entre os resíduos aleatórios das alternativas.

Ou seja, a componente de erros  $\varepsilon$  da função utilidade  $U_j$  pode ser decomposta em duas partes: uma referente ao grupo/ninho (do inglês, *nest*) de alternativas similares, denominada  $\varepsilon_N$  (que pode assumir valores diferentes para cada ninho); e outra referente à alternativa em si, denominada  $\varepsilon_{j|N}$ , conforme a Equação 18:

$$U_j = V_j + \varepsilon_N + \varepsilon_{j|N} \quad (18)$$

A probabilidade de escolha de uma alternativa  $j$  pode ser representada, então, pela probabilidade marginal de escolha do ninho  $n$ , multiplicada pela probabilidade condicional de escolher  $j$  dado que o ninho  $n$  fora escolhido.

$$P_j = P_n \cdot P_{j|n} \quad (19)$$

De forma similar ao modelo MNL, admite-se que as componentes de erros sejam idênticas e independentes, porém somente entre as alternativas pertencentes ao mesmo ninho, e que apresentam distribuição Gumbel (que resulta na forma fechada) com fator de escala  $\lambda_n$ . Considerando:

$$\theta_n = 1/\lambda_n \quad (20)$$

$$\Gamma_n = \log \sum_{j' \in j(n)} \exp \left( \frac{V_{j',n}}{\theta_n} \right) \quad (21)$$

$$P_j = \frac{\exp(\theta_n \cdot \Gamma_n)}{\sum_{n' \in N} \exp(\theta_{n'} \cdot \Gamma_{n'})} \cdot \frac{\exp(V_{j,n}/\theta_n)}{\sum_{j' \in J} \exp(V_{j',n}/\theta_n)} \quad (22)$$

Sendo:

$n$ : ninho ( $n = 1, 2, \dots, N$ );

$\Gamma_n$ : função log-soma;

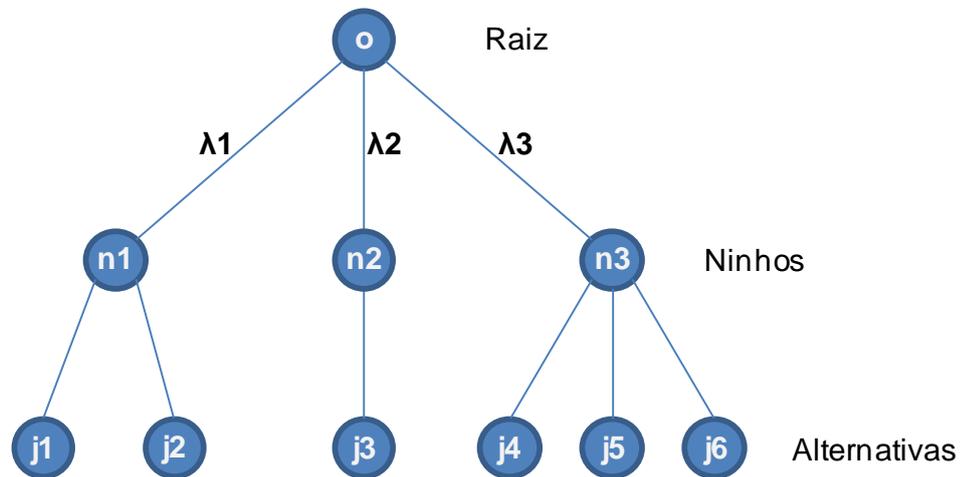
$j$ : alternativa ( $j = 1, 2, \dots, J$ );

$\theta_n$ : parâmetro de “log-soma”.

Assim, a função probabilidade do modelo Nested Logit é dada pela Equação 22. Essa função pode ser interpretada como um processo de escolha sequencial, conhecido como árvore de decisão. Como ilustrado na Figura 2, a decisão parte do nível mais amplo (raiz), selecionando determinados Ninhos em sequência, até chegar na alternativa escolhida, no nível mais baixo. É importante ressaltar que essa é

apenas uma representação didática do modelo. O processo de escolha do indivíduo não necessariamente ocorre em uma sequência de etapas.

Destaca-se ainda que, diferentemente do MNL, a elasticidade direta (ou cruzada) desse modelo pode assumir valores distintos entre as alternativas (KOPPELMAN; BHAT, 2006).



**Figura 2: Modelo Nested Logit: alternativas similares agrupadas em Ninhos**

A função log-soma, dada pela Equação 21, representa o Máximo Esperado das Utilidades das alternativas  $j'$  contidas no Ninho  $n$ . Ou seja, é a função *log* da soma da exponencial da utilidade referente às alternativas agrupadas em um determinado Ninho, divididas pelo parâmetro de “log-soma” do Ninho.

Conforme apresentado por Koppelman e Bhat (2006), o parâmetro da função log-soma,  $\theta_n$ , também conhecido como “grau de dissimilaridade”, representa o grau de correlação das componentes de erro entre as alternativas agrupadas em um Ninho. Para que haja consistência com os princípios de maximização da Teoria da Utilidade Aleatória, esse valor é restrito entre zero e um,  $0 \leq \theta_n \leq 1$ . O Grau de Dissimilaridade pode ser interpretado da seguinte forma:

- ❖  $\theta_n > 1$  : Não respeita a restrição imposta. Descarta-se o modelo;
- ❖  $\theta_n = 1$  : Correlação nula. Não existe similaridade entre as alternativas, logo não há formação de Ninhos. Nesse caso especial, o modelo Nested colapsa no MNL;

- ❖  $0 \leq \theta_n \leq 1$  : Existe correlação entre as alternativas. Quanto mais próximo de zero, maior a similaridade entre as alternativas. Quanto mais próximo de um, menor a similaridade;
- ❖  $\theta_n = 0$  : Correlação perfeita. Significa que a escolha entre as alternativas contidas em um Ninho é determinística;
- ❖  $\theta_n < 0$  : Não respeita a restrição imposta. Descarta-se o modelo.

De forma geral, o modelo Nested Logit lida parcialmente com a restrição imposta pela propriedade *I/A* ao permitir que haja correlação entre alternativas, agrupando-as em Ninhos. Esses Ninhos agrupam as características não observadas e demais erros decorrentes da modelagem de escolha. Por isso, sua aplicação é recomendada em situações de escolha bidimensional, como por exemplo, escolhas “modo-destino”, em que tipicamente existe influência das características do Destino sobre o modo de transporte a ser escolhido. Ou ainda, em casos que algumas alternativas apresentem claramente alto grau de similaridade entre si e, portanto, podem ser agrupadas.

No entanto, a construção da “árvore de decisão” pode assumir diversas estruturas e níveis diferentes. Embora o conhecimento do modelador permita excluir alguns agrupamentos indevidos, ainda assim é necessário um extensivo processo de busca pela melhor estrutura do modelo, isto é, o arranjo de ninhos e alternativas.

Assim, esse modelo apresenta a desvantagem de demandar significativa quantidade de tempo, além de não ser capaz de captar a variação de gosto entre os indivíduos, nem tratar alternativas que apresentem heteroscedasticidade. Mas sua principal limitação é a restrição de que uma alternativa pertença a apenas um único Ninho, inviabilizando a modelagem de alternativas resultantes da combinação de modos de transporte (integração modal).

### 2.2.3 Cross-Nested Logit (CNL)

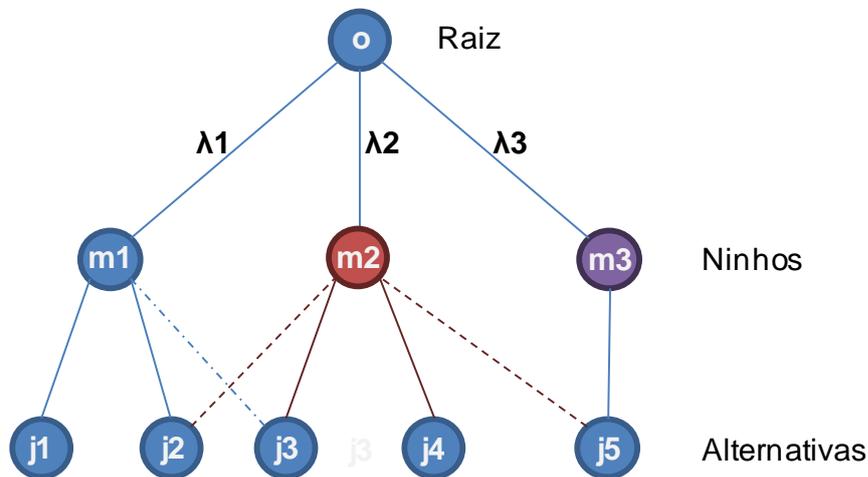
Esse modelo foi desenvolvido para superar a principal limitação do Nested Logit, permitindo que as alternativas possam pertencer, ao mesmo tempo, a mais de um Ninho. Ou seja, permite que exista correlação de alternativas entre Ninhos diferentes. Embora sua construção seja similar ao do modelo anterior, a inclusão de um termo adicional de somatório garante que cada alternativa possa pertencer a mais de um Ninho (HESS et al., 2010).

De acordo com os mesmos autores, respeitando as restrições  $0 \leq \alpha_{jm} \leq 1$  e  $\sum \alpha_{jm} = 1, \forall j$ ; a função de probabilidade de escolha de uma alternativa  $j$ , que pode pertencer a mais de um Ninho  $m$ , é dada pela Equação 23:

$$P_j = \sum_{m=1}^M \left( \frac{\left( \sum_{j \in S_m} (\alpha_{jm} \cdot \exp(V_j))^{1/\lambda_m} \right)^{\lambda_m}}{\sum_{m=1}^M \left( \sum_{j \in S_l} (\alpha_{jl} \cdot \exp(V_j))^{1/\lambda_l} \right)^{\lambda_l}} \cdot \frac{(\alpha_{jm} \cdot \exp(V_j))^{1/\lambda_m}}{\sum_{j=1}^J (\alpha_{jm} \cdot \exp(V_j))^{1/\lambda_m}} \right) \quad (23)$$

Esse modelo é significativamente flexível, sendo aplicado em situações de escolha multidimensional, principalmente quando não é possível identificar com clareza as similaridades entre as alternativas, ou quando o fenômeno estudado consiste justamente na interação entre alternativas de características distintas. Como por exemplo, casos de integração intermodal ou *park-and-ride* (quando o indivíduo conduz seu automóvel até um estacionamento e completa a viagem embarcando em um modo de transporte coletivo).

No entanto, da mesma forma que o modelo anterior, sua construção pode assumir diversas estruturas diferentes. Considerando que uma alternativa pode pertencer simultaneamente a mais de um ninho, a busca pela estrutura ideal pode resultar em um processo ainda mais demorado. A Figura 3 apresenta um exemplo da complexidade que a estrutura de ninhos e alternativas pode assumir.



**Figura 3: Cross-Nested Logit: alternativas simultaneamente em diferentes Ninhos**

Além desse ponto negativo, assim como os anteriores, esse modelo não é capaz de reproduzir a heterogeneidade de gosto entre indivíduos, nem trata

alternativas que apresentem heteroscedasticidade. Para isso, são necessário os modelos mais complexos, apresentados a seguir.

#### 2.2.4 Multinomial Probit (MNP)

Ao admitir que a componente de erro,  $\varepsilon$ , apresente distribuição Normal multivariada, a construção de sua matriz de covariância não impõe qualquer restrição, possibilitando que as variâncias sejam diferentes e que as alternativas sejam correlacionadas de qualquer forma. Isto é, não requer que a distribuição seja Idêntica e Independentemente Distribuída, nem entre as alternativas, nem entre os indivíduos. Por isso, as restrições da propriedade //A não se aplicam.

Se por um lado essa distribuição Normal confere maior flexibilidade para o modelo, por outro, dificulta sua utilização, tendo em vista que não resulta em uma função de probabilidade de forma fechada. Logo, a aplicação desse modelo é mais complexa, pois requer o uso de simulação (ou aproximações) para que seja possível obter uma solução numérica.

$$P_i = \int_{u_1=-\infty}^{u_i} \int_{u_2=-\infty}^{u_i} \dots \int_{u_j=-\infty}^{u_i} f(\mathbf{u}) du_j \dots du_1 \quad (24)$$

Como pode ser visto na Equação 24, sua função de probabilidade é composta de integrais múltiplas da função de densidade  $f(\mathbf{u})$ , que deve ser calculada para a utilidade de todas as alternativas. Os programas computacionais tipicamente adotam a abordagem de simulação para suprir essa dificuldade numérica. Uma das soluções mais utilizadas é o simulador GHK, que emprega avançadas técnicas de integração a partir de simulações do tipo Monte Carlo.

Como a complexidade de seu desenvolvimento foge do escopo deste trabalho, para informações mais detalhadas sobre o processo de utilização do Probit, recomenda-se consultar Ortúzar e Willumsen (2011) e Train (2009), seções 8.5 e 5.6, respectivamente.

De forma geral, esse modelo é flexível, permitindo capturar a heterogeneidade de gosto entre indivíduos e tratar alternativas que apresentem heteroscedasticidade ou que sejam correlacionadas. Além disso, pode ser aplicado para pesquisas longitudinais, do tipo *Panel Data* (banco de dados formado por observações feitas pelos mesmos indivíduos, porém ao longo do tempo).

No entanto, sua aplicação prática é limitada pela capacidade de processamento computacional, demandando cuidado para a interpretação dos resultados, tendo em vista que a normalização da escala dos parâmetros não é obtida de forma automática, como nos modelos da família Logit.

#### 2.2.5 Mixed Logit (ML)

O modelo Mixed Logit pode ser considerado como o mais avançado e que possui maior flexibilidade, tido atualmente como o estado da arte na modelagem de escolhas. Diferentemente do Probit, que assume distribuição Normal, esse modelo não restringe a distribuição da componente de erros.

Assim, além de capturar a heterogeneidade de gosto entre indivíduos, tratar variáveis correlacionadas e poder ser aplicado em pesquisas longitudinais (*Panel Data*), sua construção pode assumir distribuições que resultem em forma fechada da função de probabilidade. Embora essa característica simplifique significativamente seu processo de cálculo, ainda se faz necessário empregar técnicas de simulação para obter soluções numéricas.

Sua formulação pode ser obtida de duas maneiras distintas. No entanto, por ser mais didático, adota-se a abordagem da Componente de Erros. Essa versão se baseia na inclusão de uma componente de erros na função Utilidade tradicional e é descrita a seguir.

Como visto na Equação 1, a Utilidade é composta por duas componentes: 1) a sistemática, na qual os atributos observáveis  $X_{ij}$  são multiplicados por parâmetros fixos  $\beta_j$  que não variam entre os indivíduos; e 2) a de erros aleatórios,  $\varepsilon_{ij}$ , que apresenta distribuição independente e idêntica.

A Equação 25 reescreve a função Utilidade introduzindo uma componente adicional de erro ( $\varphi_{ij} \cdot Y_{ij}$ ), onde  $\varphi_{ij}$  são parâmetros desconhecidos e não observados para cada indivíduo  $i$  e alternativa  $j$ , enquanto  $Y_{ij}$  é uma distribuição arbitrada pelo modelador.

$$U_{ij} = \beta_j \cdot X_{ij} + \varepsilon_{ij} + (\varphi_{ij} \cdot Y_{ij}) \quad (25)$$

A liberdade de poder assumir qualquer distribuição para essa componente adicional de erro é que proporciona maior flexibilidade ao modelo Mixed Logit,

podendo assumir qualquer forma. De fato, assumindo definições específicas, esse modelo é capaz de replicar qualquer um dos modelos da família Logit.

Pode-se, até mesmo, reescrever modelos Logit tradicionais e incorporar diferentes estruturas de erro, conferindo características especiais, como por exemplo, um Multinomial Logit que seja capaz de tratar alternativas correlacionadas e que apresentem heteroscedasticidade.

Assim, de acordo com Ortúzar e Willumsen (2011), a obtenção da função de probabilidade do modelo Mixed Logit parte de um modelo Logit conhecido, incorporando diferentes parâmetros que variam aleatoriamente entre os indivíduos, além da influência de atributos desconhecidos que são distribuídos de forma arbitrária. A função de probabilidade é dada pela Equação 26:

$$P_{ij} = \int L_{ij}(\beta).f(\beta)d\beta \quad (26)$$

Ou seja, o modelo Mixed Logit pode ser interpretado como a média ponderada de um modelo Logit específico,  $L_{ij}(\beta)$ , obtido a partir da influência dos parâmetros  $\beta$ , que assumem aleatoriamente diferentes valores entre os indivíduos  $i$ . Sendo que o peso de cada parâmetro,  $\beta_i$ , é dado por uma função densidade arbitrada,  $f(\beta)$ .

Essa função densidade é denominada “distribuição mista” e pode assumir valores contínuos e/ou discretos, sendo este último responsável pela construção de modelos de classe Latente (quando as variáveis explicativas são intangíveis e, portanto, não podem ser observadas), muito utilizados em psicologia e marketing (TRAIN, 2009).

De forma geral, o modelo ML é o mais flexível e apresenta as mesmas vantagens do Probit, com o benefício de requisitar menor capacidade de processamento. No entanto, conforme comentado anteriormente, quanto mais complexo o modelo, maior sua dificuldade de utilização e interpretação.

Assim, conforme destacado por Ortúzar e Willumsem (2011), dependendo do objetivo do estudo e do conhecimento existente a priori para sua modelagem, o nível de flexibilidade proporcionado por esse modelo pode ser desnecessário.

## 2.2.6 Considerações finais sobre os Modelos

Conforme apresentado, existem diversos modelos derivados da Teoria da Maximização da Utilidade Aleatória, cada um apresentando vantagens e situações recomendadas de aplicação, assim como desvantagens e cuidados necessários para sua utilização. Essa contextualização teórica abrangeu os principais modelos de escolha discreta, partindo do mais simples (MNL) para o mais complexo (Mixed Logit).

É importante ressaltar o dilema da flexibilidade versus praticidade de um modelo: quanto mais flexível, maior sua complexidade de utilização, demandando significativo processamento computacional para a obtenção de soluções numéricas.

De fato, Ortúzar e Willumsem (2011) destacam que os modelos Probit e o Mixed Logit apresentam sérias restrições para aplicações em bancos de dados extensos ou em situações com mais de 4 alternativas, limitando sua aplicação prática em casos reais. Enquanto Train (2009) comenta que o potencial máximo de utilização do Mixed Logit é limitado pela capacidade de processamento computacional disponível.

Destaca-se, no entanto, que se um modelo for construído a partir de uma sólida teoria econométrica, confere-se elevado poder de explicação do fenômeno estudado aos atributos observados, fazendo com que as componentes aleatórias dos erros tenham pouca relevância. Essa redução na contribuição das componentes de erro praticamente inutiliza a maior flexibilidade proporcionada pelos modelos avançados.

Outro ponto importante é o conhecido problema estatístico em que os modelos mais complexos tipicamente apresentam melhor ajuste para determinada amostra, embora esse padrão de desempenho nem sempre se mantenha para sua aplicação em projeções futuras, diferentemente dos modelos mais simples, como o MNL e Nested Logit (CARSON et al., 2015).

Assim, cabe ao modelador a liberdade de escolher o modelo mais adequado para atender ao objetivo do estudo, sendo tipicamente recomendada a abordagem de aperfeiçoamento progressivo. Isto é, parte-se do modelo mais simples, o MNL, e verifica-se gradualmente se a utilização de modelos mais complexos resulta em um melhor poder explicativo, ou se o modelo mais simples já é suficientemente adequado.

No entanto, a especificação da estrutura e forma funcional dos modelos é apenas parte da modelagem da divisão modal. Para que o modelo possa ser aplicado

para realizar análises e estimar projeções futuras, deve-se primeiro executar as importantes etapas de calibração e testes de verificação do desempenho.

### 2.3 Calibração

A componente sistemática da utilidade é composta por coeficientes  $\beta$ , multiplicados pelos atributos explicativos  $x$ . A componente dos erros incorpora os coeficientes estruturais  $\lambda$ , que dependem da construção do modelo.

O processo de calibração consiste justamente na obtenção dos valores para esses coeficientes,  $\beta$  e  $\lambda$ , que respeitando as restrições impostas e a estrutura definida, o modelo resulte no melhor desempenho possível. Isto é, de acordo com os atributos observados em uma amostra, o modelo seja capaz de prever com elevado percentual de acertos as alternativas escolhidas pelos diversos indivíduos.

No entanto, os modelos de escolha discreta não podem ser calibrados por métodos convencionais de ajuste de regressão, como o método dos Mínimos Quadrados. Diferentemente de casos convencionais, na modelagem da divisão modal a variável dependente é uma probabilidade de escolha não observável, que deve ser confrontada com as escolhas feitas de fato pelos indivíduos que compõem a amostra de estudo.

Por isso, para os modelos que apresentam uma forma fechada de função de probabilidade, adota-se o método do estimador de máxima verossimilhança. Para sua utilização, é preciso executar dois passos: 1) definir uma função de probabilidade conjunta baseada nas escolhas previstas pelo modelo e nas escolhas realmente feitas na amostra observada, isto é, uma função de verossimilhança; somente então 2) estimar os parâmetros que maximizem essa função.

A função de verossimilhança da probabilidade de escolha de uma alternativa para uma amostra com  $I$  indivíduos, cada um com  $J$  alternativas disponíveis, pode ser escrita como o produto das probabilidades:

$$L(\beta) = \prod_{i \in I} \prod_{j \in J} P_{ij}(\beta)^{z_{ij}}$$

Sendo:

$P_{ij}$  a probabilidade de um indivíduo  $i$  selecionar a alternativa  $j$ ;

$\zeta_{ij}$  um indicador dummy de escolha que assume valor um se  $j$  for de fato a alternativa selecionada pelo indivíduo  $i$  na amostra, e zero caso contrário.

Para executar o segundo passo e estimar os parâmetros que maximizam essa função, aplica-se o conceito do Ponto Crítico. Isto é, deriva-se a função de verossimilhança em relação ao parâmetro desejado ( $\beta$  ou  $\lambda$ ) e resolve-se a equação igualando-a a zero. Para certificar que trata-se de um ponto de valor máximo (e não um valor mínimo), deve-se obter a segunda derivada com valor negativo.

No entanto, por questões de simplicidade e para evitar trabalhar com produtos de funções de probabilidade (que tipicamente apresentam exponenciais), é comum aplicar a função *log* para linearizar a função de verossimilhança. Obtém-se então o somatório de uma função significativamente mais simples, apresentada a seguir, denominada Log-Verossimilhança:

$$LL(\beta) = \log(L(\beta)) = \sum_{i \in I} \sum_{j \in J} \zeta_{ij} \cdot \ln(P_{ij}(\beta))$$

Vale ressaltar que, além do desenvolvimento matemático para obter o estimador que maximiza a função, pode-se utilizar também uma heurística de busca que identifica o valor máximo a partir de uma sequência de iterações e comparações com os valores obtidos. A maioria dos *softwares* adota essa segunda abordagem.

No caso de modelos que não possuem função de probabilidade de forma fechada, como o Probit e Mixed Logit, adota-se a verossimilhança simulada, na qual a função  $P_{ij}(\beta)$  é obtida a partir das técnicas avançadas de simulação, comentadas anteriormente.

Porém, para não prejudicar a etapa de calibração e, conseqüentemente, afetar o desempenho do modelo, recomenda-se tratar preliminarmente a amostra observada. Esse processo é denominado de “limpeza de dados” e consiste em analisar o banco de dados para identificar e remover valores indevidos. Podem ser adotadas estratégias de verificação da consistência das respostas, valores significativamente diferentes da distribuição esperada (*outliers*) e/ou a existência de colinearidade entre as variáveis explicativas.

Uma vez que o banco de dados tenha sido tratado e o modelo devidamente calibrado, procede-se então à última (e não menos importante) etapa da modelagem da divisão modal, a etapa de testes para verificação de seu desempenho e comportamento.

## 2.4 Etapa de Testes

Após especificar e calibrar o modelo, é importante testar seu desempenho para garantir que sua especificação esteja adequada e, somente então, seja possível aplicá-lo para realizar análises sobre o processo de escolha e fazer projeções futuras.

Essa etapa de verificação consiste na aplicação de uma sequência de testes que avaliam a coerência dos resultados obtidos, a significância dos coeficientes e o desempenho do modelo, isto é, sua capacidade de prever corretamente as alternativas selecionadas referentes a uma amostra observada.

### 2.4.1 Teste Informal

Esse é o primeiro teste a ser aplicado sobre o resultado do modelo calibrado. Verifica-se a coerência do sentido de contribuição de cada atributo, confrontando-os com o sentido esperado a priori. Por exemplo, considerando que todas as demais características sejam idênticas, espera-se que a alternativa que demandar maior custo e maior gasto de tempo para realizar o deslocamento seja considerada menos “atrativa”. Ou seja, espera-se, a priori, que o sinal de contribuição dos atributos relacionados a custo e tempo de viagem seja negativo.

É importante ressaltar que este teste é informal, logo é possível que os coeficientes apresentem sentido contrário ao esperado. No entanto, esses casos são raros e devem ser verificados com atenção pelo modelador, podendo indicar a existência de algum erro na especificação do modelo ou revelar um comportamento específico/diferenciado, que deve ser devidamente justificado.

Outro teste informal é a verificação de coerência da relação entre coeficientes de diferentes atributos. Um exemplo muito utilizado é a relação entre o tempo e o custo de viagem. Esta relação é interpretada como o “Valor do Tempo” (do acrônimo em inglês, *VOT*) e deve ser comparada com os valores tipicamente obtidos em trabalhos similares na literatura acadêmica. Outras relações com interpretação similar podem ser encontradas, como por exemplo, o “Valor Subjetivo do Tempo” (*SVT*) ou o “Valor do Tempo Economizado” (*VTS*).

Essa relação pode ainda ser generalizada, confrontando o coeficiente de custo com qualquer atributo, inclusive o tempo, obtendo-se assim a “Disposição em Pagar” (do inglês, *Willingness-To-Pay*). Essa relação avalia quanto um indivíduo estaria disposto a pagar para ganhar uma unidade de determinado atributo (podendo ser

tempo ou, por exemplo, conforto). No entanto, a análise sobre essas relações entre coeficientes são mais recomendadas para amostras obtidas a partir da técnica de Pesquisa Declarada, pois permitem captar com maior sensibilidade a taxa marginal de substituição (*trade-off*) de cada indivíduo (ORTÚZAR; WILLUMSEN, 2011).

#### 2.4.2 Teste dos Coeficientes

Para amostras que sejam consideradas significativamente grandes, podem-se aplicar técnicas de teste de hipótese para avaliar a significância de contribuição de determinado atributo, como por exemplo, o teste “t de Student”.

Neste teste, admite-se que a contribuição de determinado atributo equivale à zero, configurando-se como a hipótese nula a ser testada. Como pode ser visto na Equação 27, divide-se o coeficiente estimado,  $\beta_k$ , pelo seu desvio padrão,  $DP$ , confrontando-o com o valor obtido na tabela t-Student em função dos parâmetros grau de liberdade e nível de confiança.

$$t_k = \frac{\beta_k}{DP(\beta_k)} \quad (27)$$

O grau de liberdade de um banco de dados é determinado pela diferença entre a quantidade de amostras obtidas e os coeficientes a serem estimados. Quando o grau de liberdade for superior a 120, considera-se a amostra significativamente grande e adota-se o valor limite da distribuição t de Student, equivalente ao valor infinito.

Para esse caso específico e arbitrando nível de confiança em 95%, a hipótese nula é rejeitada com 95% de chance de não cometer o erro do tipo I (rejeitar uma suposição verdadeira) para valores em que a estatística t de Student esteja fora do intervalo  $[-1,96; +1,96]$ , conforme apresentado na Tabela 1.

**Tabela 1: Estatística t-Student em função do grau de liberdade e Nível de Confiança - adaptado de Fisher e Yates (1995)**

Grau de Liberdade	Nível de Confiança			
	99%	98%	95%	90%
1	63,66	31,82	12,71	6,31
2	9,99	6,96	4,30	2,92
3	5,84	4,54	3,18	2,32
120	2,62	2,36	1,98	1,66
Infinito	2,58	2,33	1,96	1,65

Assim, quando o módulo do valor da estatística t de Student de um coeficiente (obtido pela Equação 27) for maior do que 1,96, pode-se afirmar com nível de confiança de 95% que o coeficiente estimado é diferente de zero e, portanto, apresenta contribuição significativa. Caso contrário, prevalece a hipótese nula de que o coeficiente é igual à zero, logo, o respectivo atributo deve ser descartado do modelo.

#### 2.4.3 Teste dos Modelos

Similar ao teste anterior, esta verificação avalia a significância da melhoria do modelo. No entanto, ao invés de avaliar a contribuição de cada atributo, esse teste avalia o modelo como um todo.

Neste teste, admite-se que a melhoria no ajuste do novo modelo proposto equivale à zero, configurando-se como a hipótese nula a ser testada. Como pode ser visto na Equação 28, multiplica-se por dois a diferença entre os valores finais da Log-verossimilhança (*log-likelihood*) obtidos em ambos os modelos, sendo  $Y'$  o modelo a ser testado. Como a Log-verossimilhança apresenta distribuição Chi-quadrado, confronta-se o valor obtido na Equação 28 com a tabela Chi-quadrado, em função dos parâmetros grau de liberdade e nível de confiança.

$$\chi^2 = -2 \times [LL_{final}(Y) - LL_{final}(Y')] \quad (28)$$

O grau de liberdade dos modelos é determinado pela diferença entre a quantidade de parâmetros adotados no modelo a ser testado,  $Y'$ , e o modelo base,  $Y$ . Adotando a abordagem do aprimoramento progressivo, os modelos devem ser estimados incluindo um atributo de cada vez. Dessa forma, o grau de liberdade seria equivalente a um.

Para esse caso específico e arbitrando nível de confiança em 95%, a hipótese nula é rejeitada com 95% de chance de não cometer o erro do tipo I (rejeitar uma suposição verdadeira) para valores em que a estatística de verossimilhança esteja fora do intervalo  $[-3,84; +3,84]$ , como pode ser visto na Tabela 2.

**Tabela 2: Estatística Chi-Quadrado em função do grau de liberdade e Nível de Confiança - adaptado de Fisher e Yates (1995)**

Grau de Liberdade	Nível de Confiança			
	99%	98%	95%	90%
1	6,63	5,41	3,84	2,71
2	9,21	7,82	5,99	4,61
3	11,34	9,84	7,81	6,25
4	13,28	11,67	9,49	7,78
5	15,09	13,39	11,07	9,24

Assim, quando o módulo do valor da estatística Chi-quadrado da diferença entre dois modelos (obtido pela Equação 28) for maior do que 3,84, pode-se rejeitar a hipótese nula, afirmando com nível de confiança de 95% que o novo modelo estimado apresenta melhoria no desempenho significativamente diferente de zero. Caso contrário, a hipótese nula não pode ser rejeitada, logo, o respectivo modelo não apresenta contribuição significativa e deve ser descartado.

#### 2.4.4 Teste de Desempenho

Para que um modelo possa ser considerado adequado, é preciso atender a três critérios importantes: 1) produzir resultados melhores do que um “palpite”; 2) seja capaz de prever com elevado percentual de acerto as alternativas escolhidas; e 3) seja conciso, contendo somente as variáveis explicativas que exercem significativa influência no processo de escolha.

Os dois primeiros são óbvios e se referem diretamente à capacidade de predição do modelo, enquanto o terceiro critério é mais sutil e se refere à aplicação prática do modelo, buscando minimizar a quantidade de dados a serem coletados.

Esse último critério influencia não só nos gastos e, conseqüentemente, na viabilidade de realização de uma pesquisa, mas também na qualidade das respostas obtidas. Tipicamente na realização de pesquisas, quanto mais longo for o questionário, maior é o risco de “fadigar” o participante, o que pode resultar em elevada quantidade de participações incompletas ou, no pior dos casos, respostas incorretas/incoerentes.

Assim, apresentam-se três testes usualmente adotados para verificar o desempenho de um modelo.

#### 2.4.4.1 Percentual de Acertos (Hit Ratio)

Esse teste verifica o atendimento aos dois primeiros critérios, avaliando a capacidade do modelo em prever corretamente as alternativas observadas na amostra pesquisada.

$$\% \text{ acertos} = \frac{N_{\text{acertos}}}{N_{\text{Total}}} \quad (29)$$

Sendo  $N_{\text{acertos}}$  a quantidade de predições corretas obtidas pelo modelo e  $N_{\text{Total}}$  a quantidade total de observações da amostra analisada. Sua aplicação é simples e muito utilizada na literatura acadêmica, funcionando como um adequado padrão de comparação (*benchmarking*).

#### 2.4.4.2 Probabilidade média de predições corretas

Assim como o anterior, esse teste também verifica o atendimento aos dois primeiros critérios. No entanto, ele é mais completo, pois é capaz de capturar e avaliar a variação de probabilidades de escolha entre diferentes modelos.

Por exemplo, admitindo um caso hipotético em que dois modelos apresentem o mesmo Percentual de Acertos para uma determinada amostra  $N$ . Esse teste é capaz de verificar que o modelo que selecionou a alternativa errada, embora a chance de escolher a certa fosse maior (digamos, 52%), é superior ao modelo que selecionou a alternativa errada, que apresenta elevada probabilidade de ser selecionada (digamos, 90%). Ou seja, quanto maior for a amostra, maior será a quantidade de erros do segundo modelo.

$$PMPC = \frac{1}{N} \cdot \sum (P_{ij} \cdot \zeta_{ij}) \quad (30)$$

Sendo  $N$  a quantidade de observações da amostra;  $P_{ij}$  a probabilidade do indivíduo  $i$  escolher a alternativa  $j$ ; e  $\zeta_{ij}$  a variável *dummy* que assume valor um caso a alternativa  $j$  seja de fato observada na amostra, e zero caso contrário.

Embora esse teste seja o mais completo, é pouco utilizado na literatura acadêmica, provavelmente por apresentar processo mais trabalhoso de obtenção.

### 2.4.4.3 Proporção de Log-Verossimilhança Ajustada ( $\bar{\rho}^2$ )

Os testes anteriores avaliam o desempenho do modelo, porém não abrangem o terceiro critério apresentado para um bom modelo: concisão. Esse teste é amplamente utilizado na literatura acadêmica, embora seja importante ressaltar que seu resultado não possui interpretação simples.

A função de verossimilhança, utilizada para calibrar o modelo, possui ainda uma característica peculiar, que é utilizada para avaliar o desempenho do modelo.

- ❖ Seus valores são sempre negativos;
- ❖  $LL(x) = 0$  : modelo perfeito. Acerta 100% dos casos;
- ❖  $LL(x) \rightarrow -\infty$  : pior modelo possível. Erra 100% dos casos.

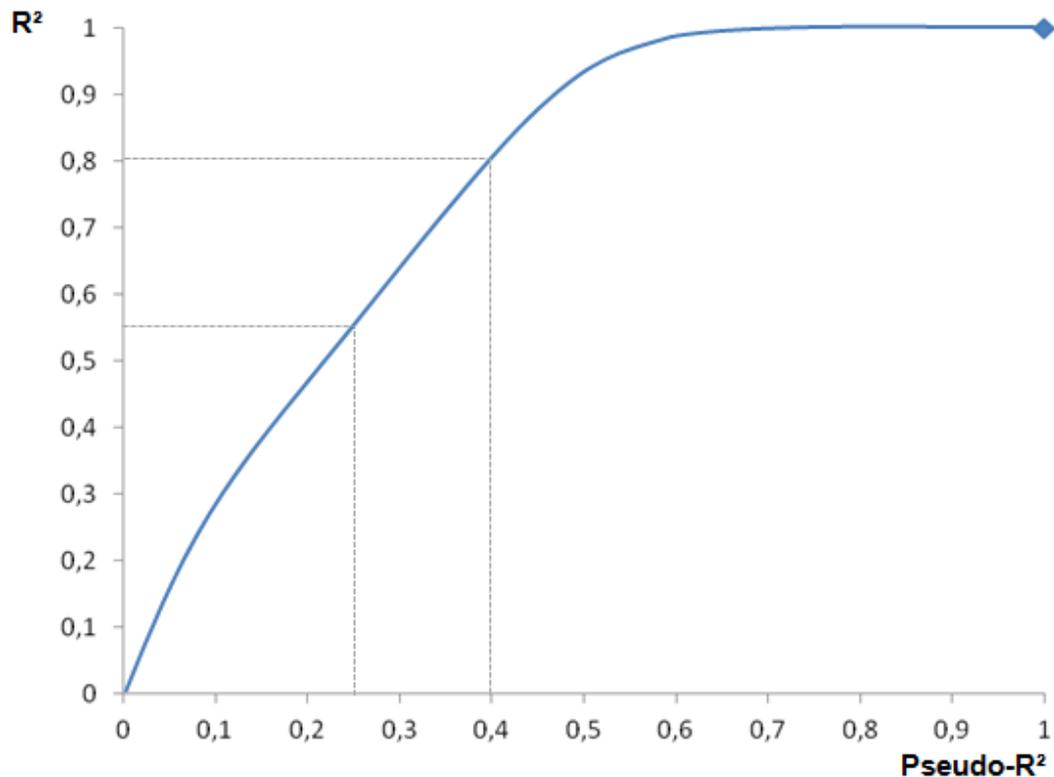
Existem três casos especiais para a função de Log-verossimilhança: 1) o modelo sem parâmetros,  $LL(0)$ , que representa o “palpite”; 2) o modelo composto somente por constantes,  $LL(C)$ , sempre superior ao primeiro; e 3) o modelo estimado,  $LL(\beta)$ , composto por constantes e variáveis explicativas.

Logo, para que o modelo estimado tenha bom desempenho, espera-se que o valor de sua função de Log-verossimilhança seja significativamente superior ao do modelo composto apenas por constantes. Para avaliar ainda o quesito concisão, subtrai-se do modelo estimado  $LL(\beta)$  a quantidade de parâmetros explicativos  $K$  (exceto as constantes), de acordo a fórmula:

$$\bar{\rho}^2 = 1 - \left( \frac{LL(\beta) - K}{LL(C)} \right) \quad (31)$$

Como pode ser visto na Equação 31, se o valor da função de Log-verossimilhança do modelo estimado for próximo ao do modelo constante, então  $\bar{\rho}^2$  tendo a zero, representando um péssimo desempenho. Se o modelo estimado for muito bom, seu valor tende a zero, fazendo com que  $\bar{\rho}^2$  tenda a um, representando desempenho perfeito.

Embora seus valores limites sejam claros, sua interpretação para valores intermediários não é trivial. Diferentemente do fator de determinação  $R^2$  de uma regressão linear generalizada, este indicador não é proporcional. De fato, Ben-Akiva e Lerman (1985) consideram como bons os modelos que resultem em valores superiores a 0,25; enquanto Ortúzar e Willumsen (2011) consideram como excelentes os valores próximos de 0,4.



**Figura 4: Relação empírica entre  $\rho^2$  e  $R^2$  (adaptado de Hensher, Rose e Greene (2015))**

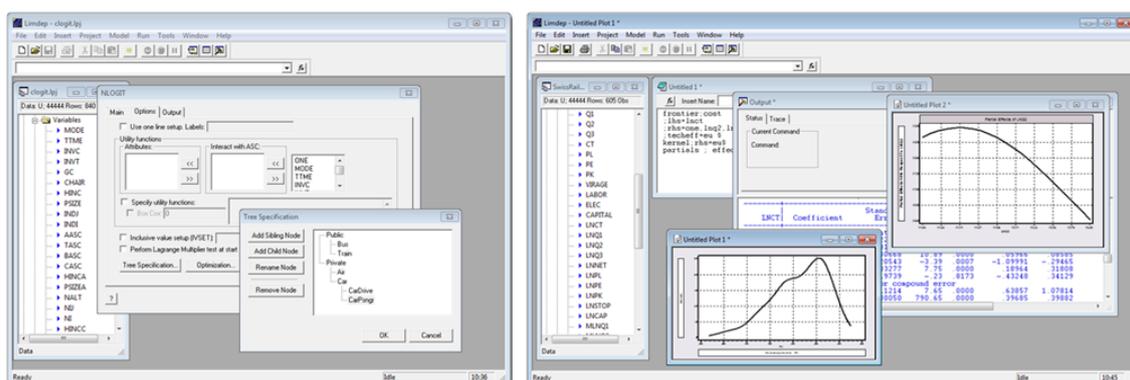
De fato, a Figura 4 apresenta a relação empírica entre o fator de ajuste dos modelos de escolha discreta,  $\rho^2$  (ou Pseudo- $R^2$ ), e o coeficiente de determinação dos modelos lineares,  $R^2$ , desenvolvida por Domencich e McFadden (1975). Verifica-se que os valores recomendados anteriormente pelos outros autores, com  $\rho^2$  entre 0,25 e 0,4, corresponderiam a  $R^2$  entre 55% e aproximadamente 80%, confirmando este intervalo como ajuste adequado a excelente de um modelo de escolha discreta.

## 2.5 Exequibilidade

Como visto nessa contextualização teórica, a estimação de um modelo (especificação, calibração e etapa de testes) é um processo longo que requer a realização de diversos cálculos, sempre trabalhando com um extensivo banco de dados. Dessa forma, é essencial contar com o apoio de programas computacionais específicos para essa finalidade, viabilizando a utilização dos modelos de escolha discreta para aplicações práticas em casos reais.

Existem alguns programas que podem ser utilizados, como por exemplo, o SPSS, SAS, STATA, NLOGIT ou até mesmo o R e o BioGeme, sendo esses dois últimos de licença aberta. No entanto, vale ressaltar que, até a conclusão dessa pesquisa de dissertação, identificaram-se poucos programas capazes de estimar os modelos Probit e Mixed Logit: apenas o NLOGIT e o BioGeme (em sua versão especial em Python).

Apesar de o programa NLOGIT possuir avançada interface de interação, cálculo e visualização de resultados (como pode ser observado na Figura 5), sua versão de licença acadêmica é demasiada restrita, limitando a avaliação de 1.500 entradas de dados. Além disso, o programa trabalha com os dados no formato *long*, no qual diversas linhas de dados, contendo informação sobre o mesmo indivíduo, representam apenas uma observação; ao invés do tradicional formato *wide*, onde cada linha corresponde a uma observação. Dessa forma, considerando a modelagem com duas alternativas de transporte, a quantidade de observações permitida é dividida por dois, limitada a apenas 750 observações.



**Figura 5: Exemplo de interface do programa NLOGIT 5 (adaptado de ECONOMETRIC SOFTWARE, 2015)**

Assim, em função da significativa limitação imposta pela versão acadêmica do software NLOGIT, essa pesquisa de dissertação utiliza o programa de licença aberta, BioGeme (BIERLAIRE, 2003). A Figura 6 ilustra o ambiente de execução do programa (à direita) e alguns dos resultados obtidos (à esquerda).

```

Model: Logit
Number of estimated parameters: 10
Number of observations: 1048
Number of individuals: 1048
Null log likelihood: -726.418
Cte log likelihood: -709.464
Init log likelihood: -726.418
Final log likelihood: -583.419
Likelihood ratio test: 285.998
Rho-square: 0.197
Adjusted rho-square: 0.183
Final gradient norm: +3.778e-004
Diagnostic: Convergence reached...
Iterations: 8
Run time: 00:01
Variance-covariance: from analytical hessian
Sample file: banco2_a.dat

```

```

-> time interval [10:01:53.10:01:53]
Run time for var/covar computation: 00:00
BIOGEME Input files
=====
Parameters: default.par
Model specification: mn1_fundao_AperfeicoamentoProgressivo.mod
Sample 1: banco2_a.dat
BIOGEME Output files
=====
Estimation results: mn1_fundao_AperfeicoamentoProgressivo.rep
Estimation results (HTML): mn1_fundao_AperfeicoamentoProgressivo.html
Estimation results (Latex): mn1_fundao_AperfeicoamentoProgressivo.tex
Estimation results (Blogit): mn1_fundao_AperfeicoamentoProgressivo.FI2
Result model spec. file: mn1_fundao_AperfeicoamentoProgressivo.res
Sample statistics: mn1_fundao_AperfeicoamentoProgressivo.sta
BIOGEME Debug files
=====
Log file: mn1_fundao_AperfeicoamentoProgressivo.log
Parameters debug: parameters.out
Model debug: model.debug
Model spec. file debug: _specFile.debug
Model informations: Logit Model
=====
The minimum argument of exp was -4.54454
Run time for estimation: 00:01
Total run time: 00:01
C:\hiogeme\MU>hiogeme mn1_fundao_AperfeicoamentoProgressivo banco2_a.dat

```

**Figura 6: Exemplo dos resultados obtidos e ambiente de execução do BioGeme**

Apesar de requerer cuidados especiais quanto à entrada de linhas de código do modelo e formatação do banco de dados, os resultados são obtidos de forma simples, em um arquivo no formato *HTML*. Tais resultados podem ser copiados para um processador de planilhas com o intuito de realizar cálculos mais avançados, como a elasticidade direta e cruzada, ou aprimorar a visualização. Além disso, o programa conta ainda com um simulador, BioSim, que permite obter o impacto na divisão modal decorrente da variação de determinados parâmetros, uma vez que o modelo tenha sido devidamente calibrado.

## 2.6 Considerações Finais

Um modelo é uma representação matemática de um fenômeno real que se deseja estudar. Para que um modelo de escolha discreta possa ser utilizado para realizar análises e projeções futuras de demanda ou divisão modal, é essencial que passe por um extensivo processo de tentativa e erro, no qual o modelador repete ciclicamente as etapas de estimação, calibração e testes de desempenho até obter um resultado considerado como satisfatório.

Essas etapas são detalhadas ao longo da contextualização teórica apresentadas no Capítulo 2, porém a boa prática da modelagem recomenda que o pesquisador desenvolva seu trabalho a partir de modelos econométricos consolidados.

Assim, faz-se necessário buscar na literatura acadêmica maiores informações sobre trabalhos similares, principalmente para identificar as principais variáveis explicativas para a adequada modelagem da divisão modal. A etapa de revisão bibliográfica é apresentada no próximo capítulo.

### 3 FATORES-CHAVE DA MODELAGEM DA DIVISÃO MODAL

A boa prática recomenda que a modelagem da divisão modal seja construída a partir de modelos econométricos devidamente consolidados, identificando na literatura os fatores-chave que são frequentemente utilizados em trabalhos e pesquisas similares.

Com esse intuito, realizou-se uma revisão sistemática para identificar os modelos e variáveis explicativas utilizados com maior frequência na modelagem da divisão modal, obtendo ainda um *benchmarking* dos indicadores de desempenho que permite comparar o poder explicativo do modelo desenvolvido nesse trabalho.

A Revisão Sistemática é um procedimento estruturado para consultar a literatura acadêmica, composta pelas etapas de planejamento de busca, sua execução e posterior análise dos materiais selecionados. A adoção de uma estratégia de busca explícita proporciona uma revisão não tendenciosa e que pode ser replicada por outros pesquisadores interessados, assegurando assim sua qualidade (PAI et al., 2003).

Este capítulo apresenta a estratégia de busca adotada na revisão sistemática realizada. Contém ainda a avaliação geral dos trinta e dois artigos selecionados, além do detalhamento dos principais fatores chaves identificados.

Resultado do atendimento de um dos objetivos específicos dessa pesquisa de dissertação, a Figura 7 apresenta um trecho do quadro que discrimina as principais informações identificadas na revisão.

#	Autor / Ano	Variáveis		Motivo Viagem	Modelo	Goodness-of-Fit		Nível Sig.	Análise
		Dependente	Independente			Pseudo-R <sup>2</sup>	% Correto		
15	Bekhore Shiftan (2009)	Carro	Tempo de viagem (total)	Casa-Trabalho	MNL	0,261	95%	Distribuição P(modal) x Variável	
		Carona	Preço do estacionamento		Nested Logit	0,268			
		Trem ( Kiss and Ride, Park and	Tempo de viagem (caminhada)		Cross-Nested Logit	0,269			
		Ônibus	# carros no domicílio Custos ou Tarifa Tempo de viagem (veículo) Tempo de viagem (espera)		Mixed MNL	0,269			

**Figura 7: Trecho do quadro sumário obtido pela revisão bibliográfica**

O quadro completo que discrimina os fatores-chave de cada artigo é apresentado no Anexo A, organizando-os em ordem cronológica e alfabética.

### 3.1 Estratégia de Busca

A estratégia de busca consiste na definição explícita de uma série de fatores que são essenciais para garantir a replicabilidade da revisão, como as bases de dados consultadas, a lógica de pesquisa e os critérios de seleção; que são apresentados a seguir.

Consultaram-se as bases de dados da *Scopus* e *Science Direct*, buscando por artigos científicos em inglês que tenham sido publicados nos últimos vinte anos. Essa definição do idioma e tipo do documento tem o objetivo de assegurar a qualidade dos trabalhos pesquisados, uma vez que precisaram passar por rigorosa etapa de revisão de pares (*peer-review*) para terem publicação aceita em periódicos internacionais.

A Figura 8 apresenta o fluxograma da estratégia de busca, detalhada a seguir.

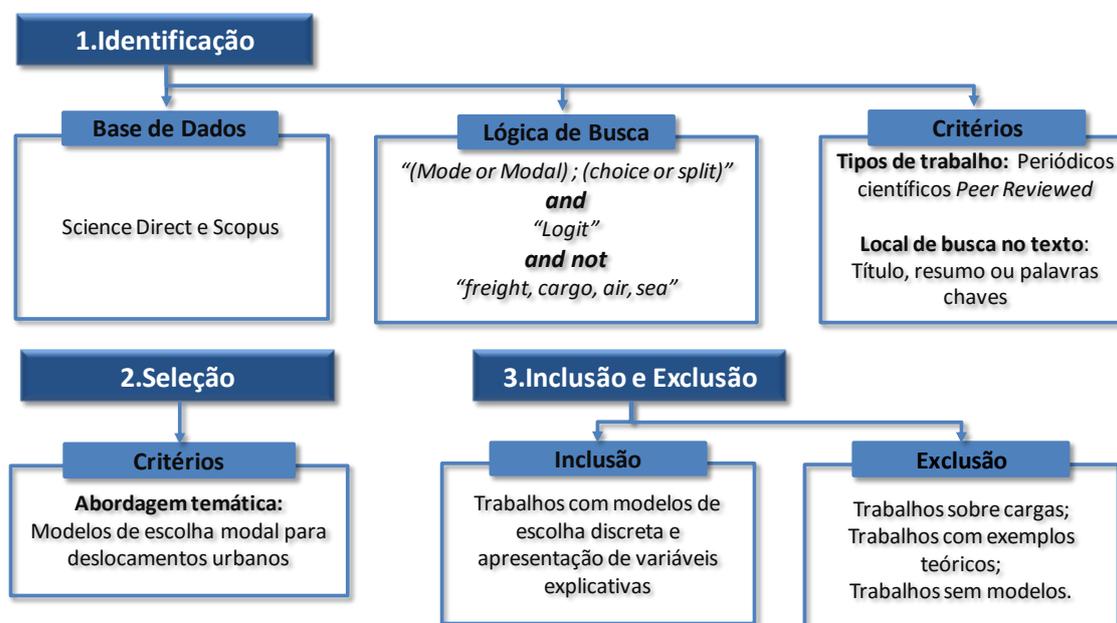


Figura 8: Fluxograma da estratégia de busca da revisão sistemática

A lógica de busca foi aplicada ao título, resumo e palavras-chave dos artigos, procurando pela combinação das exatas palavras em inglês "Mode Choice" ou "Modal Split" mais a presença da palavra "Logit", e que não apresentasse as palavras "Freight" ou "Air" ou "Cargo" ou "Sea". Isto é, buscou-se por artigos que tratassem especificamente de dois assuntos: divisão modal e modelos de escolha discreta da família Logit. Além disso, como o foco da dissertação é sobre transporte urbano de

passageiros, refinou-se a busca descartando os artigos que se referiam à aplicação sobre transporte de cargas.

O critério definido para a seleção dos artigos foi a abordagem temática de modelos de escolha modal aplicados a deslocamentos urbanos. Incluíram-se todos os trabalhos que apresentassem a especificação de um modelo de escolha discreta e que identificassem as variáveis explicativas utilizadas para sua construção. Excluíram-se os trabalhos que não especificavam claramente o modelo utilizado, que eram aplicados somente sobre exemplos teóricos ou que, mesmo depois do refinamento de busca, ainda tratavam sobre o transporte de cargas.

Dessa forma, executando o processo de revisão entre junho e dezembro de 2015, foi possível identificar 40 (quarenta) artigos que atendiam à lógica de busca. Excluíram-se 8 (oito) trabalhos que não obedeciam aos critérios de seleção. Então, a revisão sistemática permitiu selecionar 32 (trinta e dois) artigos, que são avaliados a seguir, comentando os fatores-chaves identificados, além das lacunas e oportunidades de pesquisa observadas.

### **3.2 Avaliação Geral dos Artigos**

Os trinta e dois artigos incluídos na revisão sistemática podem ser agrupados basicamente em duas categorias de objetivos distintos. Embora ambas utilizem modelos de escolha discreta e posteriormente realizem análises a partir do modelo estimado, é possível identificar claramente as distintas abordagens: 1) avaliação de efeitos de políticas públicas; e 2) aprimoramento metodológico.

A primeira é mais prática, focada no planejamento de transporte. Concentra-se na avaliação de políticas públicas, aplicando um modelo de escolha discreta para que seja possível estimar, e conseqüentemente avaliar, o impacto na divisão modal em decorrência de situações específicas. Ou seja, simulam efeitos de políticas públicas ao considerar determinada variação de atributos específicos, como por exemplo, o investimento na melhoria do transporte coletivo, resultando na maior frequência do serviço ofertado. Representa 56% da revisão (18 artigos).

A segunda categoria é mais técnica, com foco no aprimoramento da modelagem em si. Concentra-se basicamente em contribuições metodológicas que aperfeiçoam modelos existentes, ou até mesmo propõem modelos novos, aplicando-os em um caso prático com o objetivo de comparar seus indicadores de desempenho.

Ou seja, para um mesmo banco de dados, aplicam diferentes modelos e avaliam qual deles apresenta maior poder explicativo do fenômeno estudado. Tipicamente empregam o Multinomial Logit como o modelo mais simples para servir de base de comparação para avaliar o desempenho de suas contribuições técnicas. Representa 44% da revisão (14 artigos).

### 3.2.1 Categoria 1: Aplicações práticas

Em seu artigo, de Palma e Rochat (2000) aplicam um modelo Nested Logit para verificar que a inclusão de atributos relacionados ao indivíduo e ao ambiente, além de características do modo de transporte, é importante para a modelagem da divisão modal. O banco de dados é composto por 726 observações de viagens Casa-Trabalho, em pesquisa realizada na cidade de Genebra, em 1994. Utilizando a análise da elasticidade direta, identificaram que, neste caso específico, o tempo de viagem exerce maior influência sobre o processo de escolha modal do que o custo. Comentam então que políticas públicas voltadas para reduzir o tempo de viagem do transporte público seriam mais eficientes do que a redução de sua tarifa ou a cobrança de pedágio urbano para os condutores de automóvel. Embora não seja feita a comparação entre diferentes tipos de modelo considerando diferentes atributos, os autores afirmam que a inclusão dos atributos mencionados aprimora o modelo, destacando a acessibilidade e a quantidade de veículos no domicílio. Comentam ao final que as elasticidades obtidas nesse trabalho foram ainda utilizadas para calibrar o METROPOLIS, modelo dinâmico de tráfego utilizado para o planejamento de transporte na cidade de Genebra durante a década de 1990. Esse modelo dinâmico era empregado para avaliar o impacto de medidas de gerenciamento de transporte e tinha o objetivo de identificar a combinação de medidas que maximizaria os resultados desejados. Embora não seja o enfoque principal do artigo, a utilização desse modelo dinâmico se assemelha ao objetivo dessa dissertação, configurando-se como contribuição interessante para seu desenvolvimento.

Cervero (2002) investiga a influência que características do ambiente construído exercem sobre a escolha modal. Propõem ainda uma padronização para avaliar as características do ambiente, concentrando nas suas três dimensões principais: densidade, diversidade e desenho. Aplica então modelos binomiais e MNL para comparar o desempenho de modelos básicos frente aos incrementados com características do ambiente. O banco de dados utilizado é composto por 1.960 observações obtidas a partir de pesquisa domiciliar da região metropolitana de

Washington, realizada em 1994, complementada por informações detalhadas do uso do solo do condado de Montgomery. Comprova empiricamente então que as características do ambiente aprimoram o desempenho do modelo, destacando a avaliação da intensidade do uso do solo, se apresenta diversidade de usos e a qualidade das calçadas. Ao final, destaca o potencial da análise da elasticidade como uma interessante ferramenta que permite obter maior conhecimento sobre a importância relativa de cada atributo do modelo no processo de escolha modal, corroborando o potencial de utilização da expansão metodológica desenvolvida nessa dissertação.

Com o objetivo de atualizar o modelo de planejamento de transporte da Holanda, de Jong et al. (2003) estimam um modelo Mixed Logit de escolha conjunta entre o modo de transporte e o horário em que a viagem é realizada. O banco de dados utilizado é obtido a partir de pesquisa declarada que entrevistou mais de mil viajantes e foi aplicado para diferentes motivos de viagens. O artigo constata que o horário da viagem realmente é um fator de influência sobre a escolha do modo de transporte. Por meio de simulações de cenários distintos para projeções da divisão modal, os autores verificam que a escolha do horário da viagem é sensível a mudanças na duração e no custo do deslocamento dentro do horário de pico. Concluem então que esses atributos seriam os mais eficientes para influenciar no processo de decisão de viagens, mitigando a concentração de veículos na hora de pico. Embora não seja o objetivo principal do artigo, essa análise sobre os atributos que exercem maior influência no processo de escolha se assemelha ao objetivo dessa pesquisa de dissertação.

De forma similar, Hess et al. (2007) propõem um modelo de escolha conjunta entre modo de transporte e horário de viagem, com o intuito de avaliar a eficácia de medidas de gerenciamento da demanda que contribuam para a mitigação da concentração de viagens durante o horário de pico. No entanto, o destaque principal desse artigo é sobre a viabilidade prática de utilização do modelo. Por isso, os autores abrem mão da maior flexibilidade do Mixed Logit e utilizam o Nested Logit. Diferentes estruturas são testadas, confrontando os resultados com um modelo MNL (base de comparação). Assim, concluem que o Nested Logit é recomendável para aplicações práticas em casos reais que envolvem muitas variáveis e atributos. Além disso, verificaram que os indivíduos são mais sujeitos a alterarem o horário de realização da viagem do que o modo de transporte escolhido. Mais uma vez, similar ao objetivo dessa dissertação, os autores avaliam o efeito de políticas públicas por meio da

projeção agregada da divisão modal, simulando diferentes cenários com a variação de diferentes atributos.

Schwanen e Mokhtarian (2005) aprofundam a investigação sobre a influência das características da forma urbana na escolha modal, criticando conclusões de trabalhos anteriores por não levarem em consideração o efeito da “Auto-Seleção” (*Self-Selection*). Isto é, quando o indivíduo escolhe o local de morar em função da preferência por determinado modo de transporte, e não o inverso. O banco de dados é composto de 1.358 observações de viagens casa-trabalho, obtidas a partir de pesquisa domiciliar na região de São Francisco, Califórnia, realizada em 1998. Aplicam então os modelos MNL e Nested Logit, corroborando a expectativa inicial de que a consonância entre o local de moradia e a preferência modal do indivíduo exerce influência na escolha modal. Embora o artigo conclua que a relação de causalidade entre forma urbana e escolha modal seja complexa, constata que a inclusão de características sobre a forma urbana e da estrutura física do ambiente aprimora o desempenho dos modelos de escolha discreta.

Bhat e Sardesai (2006) utilizam o modelo Mixed Logit para avaliar a influência de dois aspectos sobre a escolha modal: a necessidade de realizar paradas durante o deslocamento e a confiabilidade sobre a duração das viagens. O banco de dados é composto pela combinação de pesquisa revelada e declarada realizada em Austin, Texas, que contou com 679 participantes. Os autores adotam a abordagem do aprimoramento progressivo para definir os atributos que devem ser incluídos na modelagem. Refinaram a disponibilidade das alternativas por meio de uma variável de controle para bicicletas e viagens a pé, sendo disponíveis para viagens com distância total inferior a 13 km e 4 km, respectivamente. Ao final, analisam o efeito de políticas públicas com o objetivo de identificar os atributos que exercem maior influência na escolha modal, assemelhando-se assim ao objetivo dessa dissertação. No entanto, os autores realizam essa análise por meio da projeção agregada da divisão modal, simulando diferentes cenários com a variação de diferentes atributos.

Müller, Tscharaktschiew e Haase (2008) investigam se o fechamento de locais de ensino na Alemanha, com o intuito de redução de custos, poderia causar impactos negativos em função da transferência modal resultante do remanejamento de alunos. Para tal, aplicam o MNL para viagens do tipo casa-educação e verificam que a distância total, condições climáticas e disponibilidade de automóvel são os fatores que mais influenciam no processo de escolha. O banco de dados é composto por 9.300 observações obtidas a partir de vasta pesquisa realizada em 2004, na cidade de

Dresden, Alemanha, que contou com a participação de 4.650 indivíduos reportando seus hábitos de deslocamento durante o verão e inverno. De fato, os autores simulam o impacto resultante do remanejamento de alunos e comprovam que haveria distorção na divisão modal, com significativa transferência de viagens realizadas por transporte não motorizado para o transporte público e, principalmente, para o automóvel. Vale ressaltar que esse trabalho obteve um modelo com excelente desempenho ( $\bar{\rho}^2 > 0,4$  e média de acertos superior a 80%), mesmo utilizando o mais simples MNL. Ainda, para obter mais informação sobre a influência dos atributos no processo de escolha modal, realizaram análise gráfica da distribuição da probabilidade de escolha de cada modo de transporte em função da variação da distância e de dois parâmetros binários: a disponibilidade de automóvel e condição climática.

Su e Zhou (2012) especificam um modelo Nested Logit com o objetivo de avaliar o impacto na transferência modal decorrente de diferentes estratégias de gerenciamento da demanda por viagens, porém sob uma abordagem da gestão corporativa da mobilidade. A modelagem adota atributos sobre as características dos indivíduos, dos modos de transporte e do ambiente. O vasto banco de dado é composto por 62.346 observações, obtidas a partir da pesquisa de redução de viagens para o trabalho no estado de Washington, realizada em 2005 nos Estados Unidos. Embora sejam consideradas na modelagem muitas características referentes ao ambiente, os autores concluem que as medidas que promovem maior redução do uso individual do automóvel são o aumento da tarifa de estacionamento, aumento do desconto para estacionamento de veículos com mais de um ocupante e a redução da oferta/disponibilidade de vagas. Além disso, identificaram que os indivíduos responderiam positivamente à premiação direta com subsídio financeiro, configurando-se como uma interessante medida para reduzir a utilização individual do automóvel. Similar ao objetivo e procedimento metodológico dessa dissertação, o artigo utilizou a análise da elasticidade direta para avaliar o impacto dos atributos sobre o processo de escolha.

Amoh-Gyimah e Aidoo (2013) especificam um modelo Conditional Logit para verificar a influência dos atributos sobre o processo de escolha modal para viagens do tipo Casa-Trabalho, porém sob o ponto de vista de funcionários do governo de Gana. O banco de dados é composto por apenas 372 observações, obtidas a partir da combinação de pesquisa revelada e declarada realizada em 2012. Destacam que as características individuais exercem significativa influência na escolha modal, principalmente o tamanho da família, renda, nível de escolaridade e estado civil. Embora o trabalho seja interessante, não utilizou nenhum dos indicadores

convencionais sobre desempenho do modelo, dificultando sua avaliação. Destaca-se ainda que o reduzido banco de dados pode afetar negativamente o modelo, enviesando-o e comprometendo sua decorrente aplicação para análises e projeções futuras. As análises sobre a influência dos atributos foi feita apenas pelos coeficientes e pela razão das probabilidades (*odds ratio*) de escolha entre alternativas.

Dave, Raykundaliya e Shah (2013) modelam a escolha modal de crianças pequenas na Índia com o intuito de obter maiores informações sobre os atributos que influenciam na decisão, contribuindo para melhor orientar a formulação de políticas públicas. O banco de dados é composto de 821 observações a partir de pesquisa em seis escolas na cidade de Vadodara, Índia. Aplicam então o modelo Binomial Logit, um para cada uma das quatro políticas públicas avaliadas, considerando diferentes combinações dos atributos. Concluem que a renda familiar e o custo das alternativas de transporte escolar exercem maior influência na decisão. Indicam ainda que quanto maior o tempo de caminhada, a idade da criança e a quantidade de pessoas empregadas na família, menor e a probabilidade de utilização de um serviço proposto de ônibus escolar coordenado.

Um dos trabalhos que mais se assemelham ao objetivo dessa dissertação, Santos et al. (2013) investigam quais atributos exercem influência sobre a escolha modal de viagens para o trabalho, porém como foco em cidades médias na Europa. Assim, os autores estimam três modelos de escolha discreta: MNL, NL e Mixed Logit. Embora o último (mais flexível) apresente melhor desempenho ( $\bar{\rho}^2 = 0,333$ ), não é significativamente diferente do obtido pelos demais ( $\bar{\rho}^2 = 0,331$ ). O banco de dados utilizado é composto por diversas fontes oficiais sobre 112 cidades europeias, com valores agregados médios dentre os anos de 2001 a 2004. Ao final, propõem a formulação de políticas públicas a partir da influência de determinados atributos capazes de alterar a parcela de divisão modal do automóvel, transporte público e transporte não motorizado. No entanto, é importante destacar que essa análise se refere apenas à obtenção de resultados de cada medida, não levando em consideração seu nível de eficácia, uma vez que avalia apenas o sentido de contribuição dos coeficientes e não sua magnitude.

El-Bany et al. (2014) investigam o impacto que a implantação hipotética de um sistema de BRT causaria na divisão modal da cidade de Porto Said, Egito. Para isso, desenvolvem três modelos de escolha discreta: MNL, NL e Mixed Logit. O banco de dados é composto de 2.400 observações, obtidas a partir de pesquisa declarada realizada em 2014, que contou com a participação de 100 pessoas do local de estudo.

Os autores identificam que a maior renda está diretamente relacionada com a maior probabilidade de utilização do automóvel em detrimento do transporte público ou táxi. Destacam ainda que diferentemente do observado em países desenvolvidos, o tempo fora do veículo (em espera) é mais importante do que o tempo dentro do veículo. Inferem, então, que a baixa qualidade das instalações do sistema de transporte público e a reduzida acessibilidade, características dos países em desenvolvimento, influenciam na escolha modal. Principalmente sobre o processo de escolha de pessoas que não se configurem como usuários cativos, reduzindo a probabilidade de utilização de alternativas além do automóvel. Embora o trabalho destaque o potencial dos modelos parametrizados de escolha discreta para apoiar o processo de tomada de decisão, a eficácia das medidas não foi levada em consideração.

Enaux e Gerber (2014) investigam se a consciência dos indivíduos em relação à energia e sustentabilidade exerce influência sobre a escolha modal. Para isso, especificam um modelo Binomial Logit que incorpora características socioeconômicas, do modo de transporte e do ambiente, incluindo ainda uma avaliação psicológica sobre crenças relacionadas a energia e meio ambiente. O banco de dados é composto por 3.313 observações, obtidas a partir de pesquisas de transporte que incluiu técnica de livre associação em relação à palavra “energia”, realizada em 2010 e 2011 sobre cidadãos de Luxemburgo que realizavam viagens a trabalho que atravessavam a fronteira. Concluem quem o nível de consciência sobre energia exerce influência sobre o processo de decisão, embora sua contribuição não seja significativa. Identificam que a garantia de disponibilidade de vagas de estacionamento no local de emprego está diretamente relacionada com a maior utilização do automóvel, assim como horário flexível de trabalho aumenta a propensão de utilização do transporte público. Na mesma linha dessa dissertação, os autores destacam o potencial dos modelos parametrizados de escolha discreta para melhor orientar a formulação de políticas públicas mais eficientes, comentando, dentre outros, que no caso específico estudado, traria mais resultado promover uma política de redução da oferta de estacionamento do que investir recursos financeiros na melhoria do sistema de transporte público.

Ermagun e Samimi (2015) investigam os atributos que podem influenciar na decisão modal para aumentar a parcela de utilização de transporte não motorizado para jovens em viagens do tipo casa-educação. Para isso, especificam um MNL e um Nested Logit de três níveis, incorporando características socioeconômicas, dos modos de transporte e do ambiente. O banco de dados é composto por 2.653 observações obtidas a partir de pesquisa revelada realizada na cidade de Teerã, Irã, em 2011. Similar à expansão metodológica proposta nessa dissertação, os autores realizam

análise da elasticidade direta e cruzada para avaliar o impacto na propensão de utilização modal em decorrência da variação percentual de determinados atributos. Destacam o potencial da análise da elasticidade a partir de modelos parametrizados para melhor orientar a formulação de políticas públicas. No entanto, de forma similar aos demais trabalhos, esse artigo não leva em consideração o ranqueamento dos atributos que resultariam em medidas mais eficientes.

Hammadou e Mahieux (2014) investigam quais estratégias de gerenciamento da demanda por viagem (GDV) seriam mais recomendadas para influenciar a transição para uma mobilidade mais sustentável, especificamente em uma região degradada no Norte da França. Especificam um modelo MNL incorporando características dos indivíduos, dos modos de transporte e do ambiente. O vasto banco de dados é composto por 15.628 observações obtidas por meio de duas pesquisas domiciliares de transporte realizadas em 2005 e 2006 na região do estudo. Os autores analisam a elasticidade direta para obter a sensibilidade da escolha modal em função da variação dos atributos custo, tempo de viagem e frequência do transporte público. Comparam esses resultados frente a outros estudos, destacando a singularidade da região estudada. Ao final, por meio da simulação de diferentes cenários, avaliam o efeito de algumas estratégias de GDV (e suas combinações) sobre a divisão modal, identificando que a redução na tarifa do ônibus promoveria maior transferência modal do que o vasto investimento de recursos para melhorar o serviço de transporte público. Embora esse trabalho se assemelhe ao objetivo e procedimento metodológico da dissertação, os autores utilizam a análise da elasticidade direta com o intuito de comprovar a singularidade da região de interesse (zona degradada no Norte da França), comparando os resultados frente a outros trabalhos e valores médios referentes à França.

Neste interessante trabalho, Román et al. (2014) modelam a divisão modal com o objetivo de obter melhores informações sobre a competitividade no mercado de passageiros entre as duas maiores cidades da Espanha, Madri e Barcelona. Especificam um modelo MNL (base de comparação) e um Mixed Logit, incorporando características dos indivíduos e dos modos de transporte, com especial destaque para o detalhamento sobre os tempos nas diferentes etapas das viagens: em espera, embarque e dentro do veículo. O banco de dados é composto por 1.043 observações válidas, obtidas por meio de pesquisa revelada realizada em 2009, aplicada de forma presencial por equipe treinada. Os autores destacam a utilização da *Expenditure Rate*, uma relação entre a renda domiciliar e o tempo livre disponível para gastar, em detrimento de somente considerar a renda domiciliar. Realizam análise de elasticidade

direta e cruzada, obtendo sempre valores inferiores a 1, revelando a inelasticidade da demanda neste mercado. Isto é, a redução percentual do custo ou tempo de viagem não promove aumento percentual equivalente (ou superior) no uso da alternativa. Os autores realizam ainda diversas análises sobre o valor do tempo (VOT) nas diferentes etapas da viagem, identificando que a parcela em espera é mais impactante do que o tempo de embarque, que por sua vez supera o tempo dentro do veículo. Indicam então a importância de adotar estratégias para reduzir o tempo nos pontos de verificação de segurança e de *check-in* para aumentar a competitividade do transporte aéreo. Vale ressaltar que, embora este trabalho utilize análise da elasticidade, seu intuito é investigar o comportamento da demanda e não a influência dos atributos.

Sadri et al. (2014) modelam o comportamento de escolha modal para uma situação de evacuação de emergência em função de um furacão hipotético, com foco no processo de decisão de indivíduos que não possuem alternativas próprias de modo de transporte. Para isso, especificam um modelo Nested Logit que incorpora características dos indivíduos e sobre os modos de transporte, levando em consideração ainda o destino de evacuação e a prévia experiência em situações de emergência. O banco de dados é composto por apenas 132 observações válidas, obtidas por meio de entrevista presencial com moradores da região de Miami Beach, EUA. Concluem que o serviço gratuito de ônibus especiais se configura como a alternativa com maior probabilidade de utilização, principalmente entre indivíduos que se destinam a hotéis ou abrigos. Além disso, a presença de pessoas idosas no domicílio está diretamente relacionada ao aumento da propensão de buscar carona com pessoas conhecidas, enquanto famílias com maior renda estariam mais propensas a utilizar táxi. Dessa forma, destacam o potencial do modelo parametrizado de escolha discreta para apoiar a elaboração de planos de evacuação que atendam à necessidade da população, disponibilizando oferta compatível com a divisão modal estimada. Embora esse artigo se assemelhe ao objetivo dessa dissertação, a orientação para melhor formulação de políticas públicas não leva em consideração o nível de eficácia de diferentes medidas, utilizando a análise de elasticidade somente para inferir relação de causalidade entre os atributos pesquisados e a propensão de utilização de cada alternativa de evacuação.

Shahin, Hüseyin e Kemal (2014) investigam a preferência modal para viagens realizadas em eventos especiais que movimentam elevada quantidade de pessoas, como competições esportivas. Especificam um modelo Binomial Logit, incorporando características socioeconômicas e sobre a percepção dos torcedores, atribuindo-as como variáveis *dummy* na construção do modelo. O banco de dados é composto por

apenas 199 observações obtidas a partir de pesquisa de preferência revelada, por meio de entrevistas pessoais em três ocasiões diferentes: dias de jogo das três principais equipes de futebol de Istambul. Para suprir a falta de informação sobre as alternativas não selecionadas pelos entrevistados, os autores propõem a utilização da geração de dados, adotando valores médios para o custo e tempo de viagens das demais alternativas. Concluem que a maioria dos torcedores está inclinada a utilizar o sistema de transporte público em dias de eventos especiais, independente do gênero. No entanto, indivíduos com elevada renda e faixa etária superior a 40 anos são mais propensos a utilizar o automóvel. Assim, destacam o potencial do método para apoiar o planejamento de transporte com o objetivo de mitigar o congestionamento decorrente de grandes eventos, recomendando que as medidas de gerenciamento da demanda leve em consideração os diferentes grupos de idade, concentrando principalmente nas pessoas com mais de quarenta anos.

### 3.2.2 Categoria 2: Aprimoramento técnico

Caldas e Black (1997) desenvolvem um método para aprimorar a modelagem da divisão modal, propondo um processo de tratamento do banco de dados que, embora reduza a quantidade de observações, tende a melhorar o desempenho do modelo ao remover respostas identificadas como incoerentes. O banco de dados utilizado é composto por 907 respostas obtidas a partir de Pesquisa Revelada sobre utilização da linha de trem de subúrbio, realizada no Rio de Janeiro em 1993. O método de análise geométrica identifica limites e regiões de maior consistência, removendo as observações fora dessa região. Assim, o trabalho promove redução de 30% no banco de dados (240 observações eliminadas) e melhora o desempenho em 13%. No entanto, é importante destacar que, no caso de aplicação utilizado, o tratamento proposto promoveu demasiada redução do banco de dados, o que pode comprometer a qualidade do banco de dados ao introduzir viés nas observações.

De forma similar, o trabalho de Hensher (2012) tem o objetivo de aprimorar a utilização dos dados coletados para a estimação de modelos de escolha discreta. O autor apresenta um método para a utilização do modelo Mixed Logit que seja capaz de permitir a heterogeneidade da escala de diferentes fontes de dados, importante para a combinação de pesquisas reveladas e declaradas. O banco de dados é composto de 2.668 observações obtidas a partir de diferentes pesquisas declaradas realizadas durante 1995, em seis cidades australianas. Assim, o artigo comprova que o melhor

desempenho é obtido ao permitir a heterogeneidade de escala não só entre as observações de uma pesquisa, mas também entre diferentes fontes.

Bhat (1998) apresenta a limitação do modelo Nested Logit que, embora seja muito aplicado para modelagem de situações que envolvam escolhas multidimensionais, não permite agrupar simultaneamente os atributos não observados de ambas as dimensões, apenas uma de cada vez. Propõem então a utilização do Mixed Logit para tratar concomitantemente os fatores não observados no caso bidimensional, modo de transporte e horário de realização do deslocamento. O banco de dados é composto por 3.000 amostras de viagens com origem no domicílio e de motivo social/recreacional, obtidas pela pesquisa domiciliar da Bacia de São Francisco, feita em 1990. Compara o resultado de quatro modelos, sendo um MNL (base de comparação) e três ML, um para a dimensão Modo, outro para Tempo e o último para a combinação de ambos. Comprova então, com resultados empíricos, que capturar os atributos não observados de ambas as dimensões simultaneamente resulta em melhor desempenho do modelo, além de reproduzir com maior fidelidade o processo de escolha.

De forma similar ao trabalho anterior, Bhat (2000) propõem a utilização de um modelo Mixed Logit para permitir capturar atributos não observados em mais de uma dimensão simultaneamente; neste caso, sobre a Zona de Origem e a de Destino. Trabalhando sobre a mesma fonte de dados, a amostra utilizada neste artigo é composta por 1.617 observações de viagens Casa-Trabalho. O autor apresenta em detalhe o processo metodológico de simulação, necessário para a utilização do modelo ML. Posteriormente, compara o resultado de três modelos, sendo um MNL (base de comparação), um Nested Logit e o ML proposto. Comprova mais uma vez com resultados empíricos que capturar os atributos não observados de ambas as dimensões simultaneamente aprimora o resultado do modelo e ainda é mais indicado para a avaliação de políticas públicas. De fato, esse trabalho vai além do aprimoramento metodológico e utiliza análise de projeção agregada da divisão modal para avaliar o efeito de três estratégias distintas de Política Pública por: 1) aumento do custo médio por meio de pedágio urbano, 2) redução do tempo médio de viagem do transporte público em 10 minutos, e 3) redução do tempo médio de espera pelo transporte público em 10 minutos. Pela análise realizada, conclui que a medida 2 seria capaz de promover maior redução no uso do carro e aumento do transporte público.

Koppelman e Wen (2000) desenvolve um modelo capaz de superar a restrição imposta pelo Nested Logit, com flexibilidade suficiente para permitir que haja

correlação entre as alternativas, independente do agrupamento em ninhos de similaridade. Além disso, o modelo proposto possui função de probabilidade de forma fechada, promovendo maior facilidade para sua utilização. Os autores detalham a formulação do Paired Combinatorial Logit (PCL), suas propriedades, método de estimação dos coeficientes e similaridades com outros modelos da família Logit. O banco de dados utilizado é composto por 2.769 observações obtidas a partir da pesquisa VIA Rail, realizada em 1989 para estudo de implantação de um trem de alta velocidade entre as cidades de Toronto e Montreal, no Canadá. Compara então o resultado de três modelos, um MNL (base de comparação), um NL e o PCL proposto, comprovando empiricamente que o aprimoramento metodológico apresentado é superior aos demais. No entanto, destaca-se que as vantagens obtidas pelo modelo proposto são similares às do já consolidado modelo Mixed Logit.

Hess, Bierlaire e Polak (2005) comentam a importância do indicador Valor do Tempo (VOT) para o planejamento de transportes. Destacam, porém, que sua obtenção a partir de modelos Mixed Logit não é trivial, pois a definição da distribuição da componente de erros pode ser incompatível com a teoria microeconômica, o que culminaria na obtenção de resultados inadequados. Utilizando o mesmo banco de dados do trabalho anterior, aplicam quatro modelos, sendo um MNL (base de referência) e três Mixed Logit, variando sua distribuição das componentes de erro. Identificam então o modelo que resulta no melhor desempenho, propondo a padronização da utilização da distribuição triangular para a componente de erros de modelos ML, que restringe o Valor do Tempo a valores sempre negativos.

De forma similar, Cirillo e Axhausen (2006) experimentam diferentes distribuições para a componente de erros do modelo Mixed Logit, com o intuito de verificar qual resulta no melhor desempenho, comparando-os com um MNL (base de referência). O banco de dados é composto por 5.795 respostas, obtidas por meio de uma pesquisa longitudinal realizada em 1999 na Alemanha, na qual 360 participantes preencheram um diário de viagens realizadas ao longo de uma semana. No entanto, os autores verificaram que alguns indivíduos apresentam sinal invertido ao esperado em relação ao Valor do Tempo (VOT). Isto é, estão dispostos a pagar mais caro para que a viagem seja mais longa, como por exemplo, pessoas que apreciam o trajeto percorrido. Concluem então que, diferentemente do que estava sendo adotado como prática comum, não se deve adotar distribuições que restrinjam o valor do tempo percebido pelos indivíduos, refutando a recomendação do artigo anterior.

Krygsman, Arentze e Timmermans (2007) investigam a relação entre a escolha modal e a realização de atividades intermediárias em deslocamentos do tipo casa-trabalho. Para tal, propõem uma abordagem metodológica, denominada *Co-Evolutionary*, que complementa o modelo MNL permitindo obter indicação sobre a ordem de tomada de decisão. O banco de dados utilizado é composto por 2.757 observações obtidas a partir da extensa pesquisa longitudinal realizada na Holanda, em 2001, parte do projeto AMADEUS. Concluem então que, no caso estudado, primeiro o indivíduo decide sobre a realização de viagens intermediárias e somente então escolhe o modo de transporte. Vale destacar que, com a devida especificação do modelo, obtiveram excelente desempenho ( $\bar{\rho}^2$  superior a 0,4), mesmo com o simples MNL.

Bekhor e Shiftan (2009) comentam a dificuldade na modelagem da escolha modal quando as alternativas são compostas pela integração de modos de transporte, como por exemplo, os casos de *Park-and-Ride* e *Kiss-and-Ride*. Para avaliar o modelo mais indicado para esse tipo de situação, os autores estimam seis modelos diferentes: MNL, Nested Logit e Cross-Nested Logit, além de suas versões do Mixed Logit. O banco de dados é composto por 3.588 observações de viagens com propósito Casa-Trabalho, obtidas a partir de pesquisa declarada realizada em 2000, na cidade de Tel Aviv. O artigo comprova que o modelo mais flexível (CNML – Cross Nested Mixed Logit) apresenta melhor desempenho ( $\bar{\rho}^2 = 0,269$ ), embora o resultado não seja significativamente diferente do NL ( $\bar{\rho}^2 = 0,268$ ) ou mesmo do MNL ( $\bar{\rho}^2 = 0,261$ ). Ainda, os autores verificaram que as alternativas compostas devem ser compreendidas como majoritariamente de transporte público, uma vez que essa estrutura de agrupamento obteve melhores resultados.

Pulugurta, Arun e Errampalli (2013) propõem a utilização de inteligência artificial para a modelagem da divisão modal, justificando que a escolha individual é um processo complexo que seria melhor representado por lógica Fuzzy. O banco de dados é composto por 5.822 observações obtidas a partir de pesquisa domiciliar realizada em 2012, na cidade de Port Blair, Índia. Contendo características do indivíduo e do modo de transporte, os autores compararam o desempenho do modelo Fuzzy frente a um MNL e comprovam que a técnica de inteligência artificial obtém melhor desempenho no percentual de predições corretas. Ao final, avaliam o impacto de diferentes políticas públicas por meio da simulação da divisão modal em função da variação de diferentes atributos. No entanto, embora esse artigo se assemelhe ao objetivo dessa dissertação, o método utilizado apresenta duas significativas limitações: por não ser parametrizado, não é possível obter informações mais detalhadas sobre a

influência de cada atributo; e, além disso, por não ser embasado em uma teoria microeconômica, o modelo não tem respaldo teórico que valide sua expansão para análises de projeções futuras.

Sun, Zhu e Chen (2013) adotam uma abordagem diferenciada ao propor um modelo dinâmico, no qual a componente sistemática da função utilidade incorpora não só atributos referentes à demanda, mas também sobre a oferta dos modos de transporte. O caso de aplicação utiliza dados referentes ao corredor de passageiros entre as cidades de Zhengzhou e Xian, na China, levando em consideração a distância, tarifa, tempo de viagem e custo de construção de cada uma das alternativas. A divisão da demanda é estimada por meio da simulação de diversos parâmetros, incorporando de forma dinâmica que a maior quantidade de usuários de determinado modo resultaria na redução relativa do custo de construção, afetando a escolha modal. O ponto de equilíbrio é identificado por meio de um algoritmo de busca. Embora o modelo proposto permita estimar a divisão modal por meio da simulação da variação de determinados atributos, não é possível obter informações mais precisas sobre a influência de cada um relativamente ao processo de decisão. Além disso, não é possível avaliar a consistência dos coeficientes estimados. Ressalta-se, então, a importância da utilização de modelos parametrizados.

Habib e Weiss (2014) propõem um método para avaliar a evolução do comportamento da escolha modal ao longo do tempo. Aplicado a viagens destinadas ao trabalho, fazem uso de pesquisa longitudinal e incorporam variáveis latentes ao modelo para representar o “potencial de cativar dos modos de transporte”; isto é, a atratividade não observável de cada alternativa. O modelo proposto estende a formulação do Paired Combinatorial Logit para acomodar a função de escala parametrizada da componente Logit adotada no avançado modelo *Captivity Logit*. O banco de dados é composto por 199.092 observações, obtidas a partir de três edições da pesquisa domiciliar de transporte realizada nas cidades de Toronto e Hamilton, Canadá, em 1996, 2001 e 2006. Ao final, comparam o desempenho obtido frente a um MNL, comprovando empiricamente a superioridade do modelo proposto. Destacam, então, a importância de variáveis latentes relacionadas à atratividade dos modos de transporte para a adequada modelagem e análise da divisão modal.

Van, Choocharukul e Fujii (2014) investigam a contribuição de fatores psicológicos sobre o processo de escolha modal, focando em decisões sobre viagens para o trabalho feitas por jovens universitários de países desenvolvidos e em desenvolvimento. O banco de dados é composto por 1.096 observações obtidas por

meio de pesquisas voluntárias realizadas em universidades de engenharia de seis países asiáticos, em 2005 e 2006. Especificam então sete modelos MNL, um para cada país e um considerando a junção de todos os dados. Incorporam os fatores psicológicos em escala qualitativa com cinco níveis entre pares de valores antagônicos, como por exemplo, “ruim-bom” ou “barato-caro”. Esses fatores são agrupados em três aspectos distintos: simbólicos/afetivos, instrumentais e ordenamento social. Concluem afirmando que todos os aspectos psicológicos considerados se provaram significativos na modelagem da escolha modal. Identificam a importância de aspectos simbólicos/afetivos sobre a baixa utilização do transporte público, recomendando que sejam implantados programas específicos para melhorar a imagem desse modo, principalmente no sistema educacional de crianças em países em desenvolvimento. No entanto, embora a utilização de fatores psicológicos possa contribuir de forma interessante para ampliar o conhecimento sobre os atributos que influenciam na escolha modal, destaca-se o baixo desempenho obtido pelo modelo estimado ( $\bar{\rho}^2 < 0,25$ ), o que compromete a avaliação de seus parâmetros. Destaca-se então a necessidade de cuidado especial sobre a forma de levantamento e tratamento desses dados, ou mesmo uma maior quantidade de observações para cada uma das pesquisas aplicadas.

Flügel et al. (2015) propõem um método para a modelagem da divisão modal considerando a implantação de uma nova alternativa de transporte, tratando as diferenças de escala para a combinação de banco de dados de pesquisa revelada e declarada. O banco de dados do exemplo de aplicação é composto por 8.402 observações obtidas pela participação de 607 pessoas sobre viagens com motivo recreacional. A pesquisa foi aplicada de forma online e as escolhas eram apresentadas de forma binária, confrontando as características do modo de transporte tipicamente utilizado e do sistema de Trem de Alta Velocidade a ser implantado, na Noruega. Para que fosse possível tratar esse banco de dados, os autores utilizaram o modelo Cross-Nested Logit. Apresentam então a construção matemática do método proposto. Ao final, comentam a importância de combinar PR e PD para agregar maior realidade aos estudos de projeção de demanda por um novo modo de transporte. No entanto, ressaltam a complexidade desse tipo de estudo, uma vez que a disponibilidade de uma nova alternativa pode alterar completamente a estrutura de competição no mercado, o que desqualificaria os dados obtidos pela PR. Concluem então que o método proposto é mais recomendado para casos em que a pesquisa de preferência declarada contemple, pelo menos, três alternativas, mesmo que isso implique em maior dificuldade para o adequado preenchimento do questionário.

### 3.3 Identificação dos Fatores-Chave

Com base na revisão bibliográfica, foram identificados os fatores-chave utilizados com maior frequência no processo de modelagem da escolha modal, dentre as variáveis dependentes, variáveis independentes, análises realizadas, modelos empregados e seus respectivos indicadores de desempenho.

Tais fatores são apresentados em detalhes a seguir.

#### 3.3.1 Variável Dependente (Alternativas)

Os principais modos de transporte adotados como alternativas para a modelagem da divisão modal são: carro, carona, transporte coletivo, ônibus, trem e transporte não motorizado (englobando bicicleta e deslocamentos a pé).

**Tabela 3: Frequência das variáveis independentes identificadas na revisão**

Variável Dependente					
Carro	27	<b>84%</b>	Ônibus Fretado	2	<b>6%</b>
Transporte Coletivo	15	<b>47%</b>	Rickshaw	2	<b>6%</b>
Carona	13	<b>41%</b>	TAV	2	<b>6%</b>
Ônibus	13	<b>41%</b>	Táxi	2	<b>6%</b>
Trem	12	<b>38%</b>	VLT	2	<b>6%</b>
TNM	9	<b>28%</b>	BRT	1	<b>3%</b>
Avião	5	<b>16%</b>	Transporte Individual	1	<b>3%</b>
A pé	3	<b>9%</b>	Van	1	<b>3%</b>
Bicicleta	3	<b>9%</b>	Outros	1	<b>3%</b>
Motocicleta	3	<b>9%</b>			

De acordo com a Tabela 3, poucos trabalhos consideram as alternativas "a pé" e "bicicleta", optando pelo seu agrupamento em "transporte não motorizado" (TNM).

O inverso ocorre com o Transporte Individual, para o qual, na maior parte dos casos, opta-se por discriminar as alternativas em "carro", "carona", "motocicleta" e "táxi", indicando que esses modos de transporte apresentam características específicas e significativamente distintas, por isso recomenda-se analisá-los separadamente.

Quanto ao Transporte Coletivo, não é possível identificar um padrão claro. Muitos estudos optam pela alternativa em sua forma agrupada, mais genérica (47%), enquanto outros discriminam as alternativas em "ônibus" e "trem" (cerca de 40%). Entende-se, portanto, que a decisão sobre a melhor forma de apresentar a alternativa

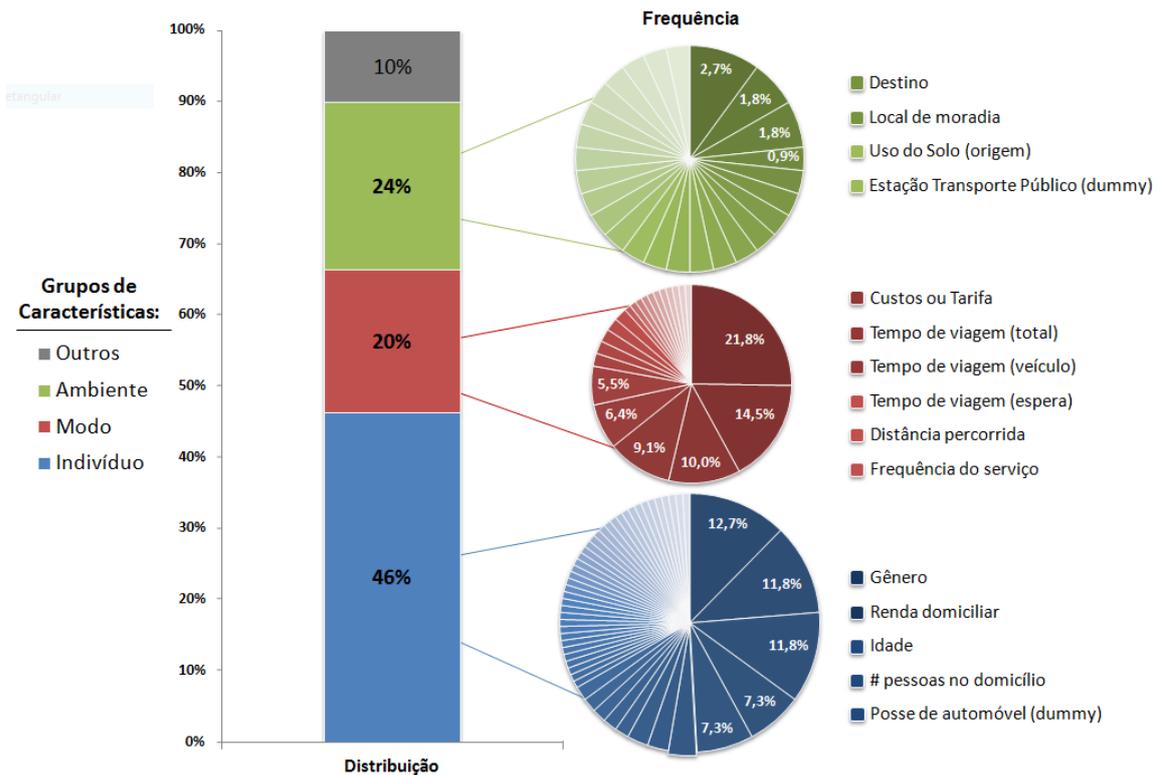
referente aos transportes coletivos depende do objetivo do estudo e da similaridade entre os atributos observáveis (cobertura/atendimento, preço da tarifa etc.) e os não observáveis (nível de serviço, conforto etc.) de cada modo.

Por último, observa-se a baixa frequência da utilização de alternativas como “van”, “rickshaw”, “VLT”, “BRT” ou “ônibus fretado”. Pode-se inferir que a baixa frequência de utilização dessas alternativas decorre, de forma geral, da reduzida oferta desses modos de transporte nos estudos realizados. Assim, sua utilização depende tipicamente das características locais da região que o estudo se aplica (por exemplo, países desenvolvidos ou em desenvolvimento).

### 3.3.2 Variáveis Independentes (Explicativas)

Foram identificadas 110 diferentes variáveis explicativas utilizadas nos trabalhos de modelagem da divisão modal. Estas variáveis podem ser agrupadas, basicamente, em quatro grandes grupos, sendo três deles referentes às características dos indivíduos (46%), do modo de transporte (20%) e do ambiente (24%); e o último se refere a um grupo genérico, aqui denominado “Outros” (10%).

A Figura 9 apresenta a distribuição das variáveis explicativas por grupo de característica, assim como a frequência de participação das principais variáveis que os compõem. Destaca-se que a soma da frequência excede o valor de cem por cento, tendo em vista que a modelagem incorpora a combinação de diferentes atributos.



**Figura 9: Principais variáveis explicativas agrupadas por características do Indivíduo, Modo, Ambiente e Outros, além de sua frequência**

As variáveis utilizadas com maior frequência nos artigos examinados estão listadas na Tabela 4. Dentre elas, destacam-se alguns dados socioeconômicos (gênero, idade e renda domiciliar), o custo total relacionado ao deslocamento e a duração da viagem, considerando o tempo total ou sua discriminação em duas parcelas específicas: dentro do veículo e em espera.

Essa separação em duas parcelas parte de um conceito mais sutil, referente à percepção individual sobre a relatividade do tempo. Isto é, uma unidade de tempo pode assumir valores distintos dependendo da situação que o indivíduo se encontra. De fato, Ortúzar e Willumsen (2011) afirmam que o valor do tempo em espera é significativamente superior ao valor do tempo dentro do veículo.

Tal segregação se mostra essencial caso o objetivo do estudo requeira maior sensibilidade do modelo, como por exemplo, para avaliar o impacto do aumento da frequência do serviço de ônibus sobre a divisão modal, tendo em vista que a medida afetaria apenas a parcela “tempo em espera”.

**Tabela 4 : Frequência das variáveis explicativas identificadas na revisão**

<b>Grupo</b>	<b>Variável Explicativa</b>		
Modo	Custos ou Tarifa	24	22%
Modo	Tempo de viagem (total)	16	15%
Indivíduo	Gênero	14	13%
Indivíduo	Renda domiciliar	13	12%
Indivíduo	Idade	13	12%
Modo	Tempo de viagem (veículo)	11	10%
Modo	Tempo de viagem (espera)	10	9%
Indivíduo	# pessoas no domicílio	8	7%
Indivíduo	Possui automóvel (dummy)	8	7%
Modo	Distância percorrida	7	6%
Modo	Frequência	6	5%
Indivíduo	# carros no domicílio	4	4%
Indivíduo	Estado Civil (dummy)	3	3%
Indivíduo	Educação (categorias)	3	3%
Ambiente	Destino	3	3%

Embora de forma agrupada as características relacionadas ao ambiente represente significativa participação na modelagem da divisão modal (24%), não é possível identificar, de forma isolada, as variáveis explicativas utilizadas com maior frequência. De fato, apenas os atributos “destino” e “origem” dos deslocamentos aparecem em mais de um trabalho.

Pode-se deduzir que isso ocorre por não haver a padronização de indicadores específicos para representar características sobre o ambiente construído, identificando assim uma lacuna na modelagem da divisão modal. De fato, (CERVERO, 2002) propõem a padronização para modelar as características do ambiente, denominando de “3 D’s”, sendo: Densidade, Diversidade e Desenho.

Dentre os trabalhos revisados que consideram a influência do ambiente na modelagem, podem-se citar alguns exemplos de indicadores *proxies* utilizados para avaliar a acessibilidade do transporte público ou a facilidade de utilização de modos não motorizados: a existência de estações de transporte de média-alta capacidade; e a extensão da malha ciclável de determinada região. Justamente essa flexibilidade para definir indicadores *proxies* é que dificulta a padronização de variáveis para avaliar as características do ambiente.

### 3.3.3 Motivo da Viagem

Como esperado, a maioria dos estudos concentra sua aplicação para examinar os deslocamentos de motivo Casa-Trabalho (59%). Isso pode ser explicado, basicamente, pela elevada relevância dos impactos decorrentes deste tipo de viagem, que tipicamente resulta na concentração de elevada quantidade de pessoas se deslocando nos horários de pico (até mesmo contribuindo para sua formação).

**Tabela 5: Frequência dos motivos de viagem identificados na revisão**

Motivo Viagem		
Casa-Trabalho	19	59%
Casa-Educação	3	9%
Casa-Lazer	2	6%
Encadeamento	2	6%
Evacuação	1	3%
N/I	5	16%

No entanto, como pode ser visto na Tabela 5, foram identificados estudos sobre diversos motivos de viagem, como deslocamentos para fins educacionais e recreacionais. Além disso, outros tipos específicos de viagem foram estudados, como o encadeamento de viagens (*trip-chaining*) e até mesmo movimentos em situação de emergência/evacuação.

Corroborar-se então a ideia de que o método de análise de escolha discreta pode ser aplicado nas mais diversas situações de interesse, variando apenas os atributos chaves que devem ser levantados pela aplicação do questionário de pesquisa.

### 3.3.4 Modelos Utilizados e Indicadores de Desempenho (*Goodness-of-Fit*)

Os principais modelos utilizados nos trabalhos revisados são: Multinomial Logit - MNL (35%), Mixed Multinomial Logit - ML (19%) e Nested Logit - NL (18%). Pode-se observar ainda que a maioria aplica mais de um modelo com o intuito de comparar o poder explicativo de cada um. Tipicamente, o MNL é adotado como a base de referência para essa análise comparativa.

Embora a quase totalidade dos trabalhos concentre a aplicação de modelos econométricos baseados na maximização da função utilidade, identificou-se um trabalho que propôs a utilização do método de Lógica Fuzzy para modelar o processo

de escolha modal. Entretanto, Ben-Akiva e Lerman (1985) insistem que se deve resistir à tentação de utilizar métodos que não sejam embasados em uma consolidada teoria microeconômica, como por exemplo, técnicas de *Data Mining* – ainda que apresentem desempenho superior quanto a predição correta dos modos de transporte selecionados – pois não teriam respaldo técnico que justifiquem sua extrapolação para projeções futuras. Além disso, os métodos de Inteligência Artificial não são parametrizados, impossibilitando obter informações sobre a contribuição individual de cada atributo.

Quanto aos indicadores de desempenho dos modelos (*Goodness-of-Fit*), os mais utilizados são a “Proporção de Log-Verossimilhança Ajustada” ( $\bar{\rho}^2$ ) e o “Percentual de Acertos” das previsões feitas. Alguns trabalhos apresentam resultados com demasiado baixo desempenho, provavelmente decorrentes do reduzido banco de dados utilizado. Assim, a presença de valores extremos pode resultar em distorções. Por isso adotou-se a mediana como referência para avaliar os indicadores de desempenho, apresentados na Tabela 6

**Tabela 6: Frequência dos modelos utilizados e seus indicadores de desempenho**

MODELOS			INDICADORES DE DESEMPENHO			
			$\bar{\rho}^2$ - Rô quadrado ajustado			% Correto
			Mediana	Mín	Máx	
MNL	24	35%	0,286	0,03	0,63	81%
Mixed MNL	13	19%	0,441	0,20	0,59	-
Nested Logit	12	18%	0,236	0,12	0,59	-
Binomial Logit	5	7%	0,287	0,14	0,48	78%
Outros	14	21%	-	-	-	-

Vale ressaltar que, conforme apresentado no item 2.4.4.3, o indicador de desempenho “Proporção de Log-Verossimilhança Ajustada”, denotado por  $\bar{\rho}^2$ , não deve ser confundido com o coeficiente de determinação,  $R^2$ . Diferentemente do segundo (e mais conhecido), o primeiro é mais disperso e representa ajuste adequado para valores superiores a 0,25 e excelentes para valores superiores a 0,4 (BEN-AKIVA; LERMAN, 1985; ORTÚZAR; WILLUMSEN, 2011).

Uma avaliação superficial desse indicador aponta o modelo Mixed Logit como o de melhor desempenho. No entanto, uma investigação mais detalhada permite verificar que a comparação do desempenho dos modelos é mais complexa. Por exemplo, observando os máximos valores obtidos, identifica-se que o Mixed Logit apresenta o mesmo desempenho do Nested Logit (0,59), ambos superados pelo modelo mais simples, o MNL (0,63).

Outro ponto interessante é que, na maior parte dos casos analisados, mesmo quando o ML apresenta o melhor desempenho para determinado banco de dados, seu resultado não é significativamente diferente dos demais modelos, como pode ser visto no trabalho de Bekhor e Shiftan (2009), artigo número 15 do quadro sumário contido no Anexo A, que também pode ser visto no exemplo ilustrativo da Figura 7.

Teoricamente, quanto mais flexível for o modelo, melhor é sua capacidade de representar o processo real de escolha, permitindo reproduzir a heterogeneidade da percepção entre distintos indivíduos, além de captar até mesmo atributos não observáveis (ou que o modelador tenha falhado em observar). Porém, na prática, não existe um modelo que por si só seja superior e apresente consistentemente resultados superiores aos demais.

De fato, Müller, Tscharktschiew e Haase (2008) obtém excelente desempenho de ajuste ( $\bar{\rho}^2 > 0,6$ ) com 81% de predições corretas, mesmo utilizando o simples modelo MNL. Enquanto Hess et al. (2007) justificam a utilização do modelo Nested Logit, uma vez que seria inviável aplicar o modelo Mixed Logit sobre um banco de dados de um caso real, composto por muitos atributos e observações.

Dessa forma, ressalta-se a responsabilidade do pesquisador em adotar os modelos que julgar mais adequados em função dos objetivos da pesquisa, do nível de detalhamento e dos recursos disponíveis (financeiros e, principalmente, de tempo).

### 3.3.5 Nível de Significância

Como pode ser visto na Tabela 7, não é possível identificar um padrão para a determinação do nível de significância adotado na avaliação dos coeficientes. A maioria dos trabalhos apresenta diferentes níveis para os coeficientes.

No entanto, é importante destacar que essa variação não torna uma variável mais ou menos significativa, apenas altera o rigor do critério para testar se o coeficiente estimado é significativamente diferente de zero, rejeitando então a hipótese nula.

**Tabela 7: Frequência dos níveis de significância identificados na revisão**

<b>Nível de Significância</b>		
95%	10	31%
90%	9	28%
99%	2	6%
N/I	11	34%

Vale ressaltar que os únicos dois trabalhos que adotaram nível de significância em 99%, o fizeram por causa do demasiado reduzido banco de dados utilizado. Logo, pode-se concluir que, caso o tamanho da amostra observada seja representativa, pode-se aceitar como significativos os coeficientes testados considerando nível de significância em 95%; ou até mesmo relaxar para o valor limite de 90%, caso o respectivo atributo seja de especial interesse para o estudo.

### 3.3.6 Análises

Embora cerca de 10% dos trabalhos não tenham apresentado claramente o resultado dos modelos de regressão, detalhando o valor dos coeficientes e sua respectiva significância (t-estatística), a grande maioria apresenta o quadro de resultados e analisa o sentido dos coeficientes (91%). Essa análise consiste na verificação da consistência das variáveis; isto é, verifica o sentido de influência da variável em relação à probabilidade de escolha de determinado modo de transporte, confrontando com o sentido esperado *à priori*.

Uma significativa parcela dos trabalhos (41%) analisa o efeito de políticas públicas, simulando o impacto na divisão modal decorrente da variação de determinados atributos, como tempo de viagem, custo ou redução de vagas de estacionamento. Tipicamente essa análise é realizada por meio da projeção agregada da demanda, aplicando o modelo estimado sobre o banco de dados, porém em diferentes cenários, com variações específicas dos atributos.

No entanto, é importante destacar que poucos artigos aprofundam sua análise além da verificação de consistência da calibração do modelo ou da simulação do efeito de políticas públicas. De fato, apenas 38% dos trabalhos analisam a elasticidade para avaliar a magnitude de contribuição dos diferentes atributos, o que permite identificar o incremento (ou redução) marginal da probabilidade de escolha de uma alternativa decorrente da variação marginal de uma variável explicativa.

Menor ainda é a quantidade de trabalhos que analisam a elasticidade cruzada, apenas 6%. Similar à elasticidade direta, essa análise consiste em avaliar o impacto da variação de uma determinada variável explicativa, porém sob o ponto de vista da probabilidade marginal de escolha dos demais modos (aos quais o atributo alterado não se aplica). Como por exemplo, a variação na propensão do uso do transporte público e da bicicleta em função do aumento em 50% do preço de estacionamento.

**Tabela 8: Frequência das análises identificadas na revisão**

Análises		
Coeficiente	29	91%
Efeito de Políticas Públicas	13	41%
Elasticidade	12	38%
Valor do Tempo	9	28%
Odds Ratio	4	13%
Não Identificado	3	9%
Distribuição de Probabilidade	2	6%
Elasticidade Cruzada	2	6%
Outros	2	6%

Assim, pode-se verificar que a maioria dos trabalhos concentra sua avaliação somente sobre o sentido dos coeficientes e sobre a simulação do efeito de diferentes políticas públicas. Embora a elasticidade direta/cruzada forneça importante conhecimento sobre a magnitude de influência dos atributos, essa análise ainda não é amplamente utilizada.

### 3.4 Considerações Finais

Por ser uma pesquisa exploratória, a revisão sistemática se configurou como uma etapa essencial do trabalho, o que permitiu identificar os fatores-chave utilizados com maior frequência na modelagem da divisão modal, fundamentando o desenvolvimento e aplicação dessa dissertação. Além disso, permitiu ainda identificar algumas lacunas referentes à análise dos modelos de escolha discreta. Uma delas, comentada a seguir, se destaca por motivar a expansão metodológica proposta nesta dissertação.

Dentre os 12 trabalhos científicos que analisaram a elasticidade direta (41%), pode-se observar que o fizeram de forma dispersa. Alguns avaliaram o efeito de políticas públicas, enquanto outros inferiram relação de causalidade ou avaliaram a elasticidade da demanda em um mercado específico. Além disso, apenas 6%

chegaram a considerar o efeito da elasticidade cruzada, avaliando o impacto de um atributo nos demais modos de transporte.

Assim, encerra-se a revisão bibliográfica destacando a oportunidade de ampliar a utilização dessas análises. A elasticidade direta e cruzada permite avaliar a magnitude da influência de cada atributo sobre a propensão de utilização de uma alternativa, possibilitando obter um ranqueamento das variáveis explicativas que exercem maior influência no processo de escolha modal. Dessa forma, respondendo a pergunta central deste estudo, apresenta-se em maior detalhe, no próximo capítulo, a contribuição metodológica proposta que permite apoiar a priorização de projetos e medidas mais eficazes para melhorar a mobilidade urbana e torná-la mais sustentável.

## 4 PROCEDIMENTO METODOLÓGICO

Após a contextualização teórica e revisão bibliográfica, apresenta-se o procedimento metodológico necessário para aplicar a técnica de análise de escolha discreta na modelagem da divisão modal. Esse capítulo inclui o detalhamento da obtenção e tratamento dos dados, assim como a especificação, calibração e os testes de verificação do desempenho dos modelos. As análises e simulações de impactos futuros são apresentadas no capítulo posterior, comentando e avaliando os resultados obtidos.

Atendendo a um dos objetivos específicos, utiliza-se um banco de dados obtido por meio de pesquisa de preferência revelada para exemplificar a aplicação da expansão metodológica proposta. Esse tipo de pesquisa apresenta apenas informação referente ao modo de transporte escolhido pelo entrevistado, não contendo dados sobre os atributos das demais alternativas. É importante destacar que, sem que haja pelo menos duas alternativas disponíveis, a decisão pelo modo de transporte não se configura como uma escolha, sendo o usuário identificado como cativo de determinada alternativa.

Para superar essa limitação imposta pela pesquisa de preferência revelada, adotam-se duas abordagens distintas. A primeira se baseia nos dados observados para estimar os valores medianos das alternativas não selecionadas, tendo o local de origem como vínculo de referência para garantir maior fidelidade possível aos valores estimados. Já a segunda, descarta completamente tais dados, estimando valores de distância e tempo para ambas as alternativas, baseando-se em um vasto sistema de dados georreferenciados atualizado, composto por elevada quantidade de informações sobre diferentes trajetos, dias e horários. Com informações de distância e tempo pode-se estimar, posteriormente, o custo referente aos deslocamentos de automóvel.

A Figura 10 apresenta o fluxograma das atividades necessárias para o desenvolvimento desta pesquisa, indicando sua seqüência de execução e destacando a etapa em que se procede a contribuição metodológica proposta (Atividade 4.5), resultante da expansão do cálculo da Elasticidade.

Posteriormente, apresenta-se a pesquisa utilizada para o levantamento de dados, as etapas de tratamento dos dados, calibração e testes de desempenho dos modelos, para cada uma das duas abordagens empregadas.

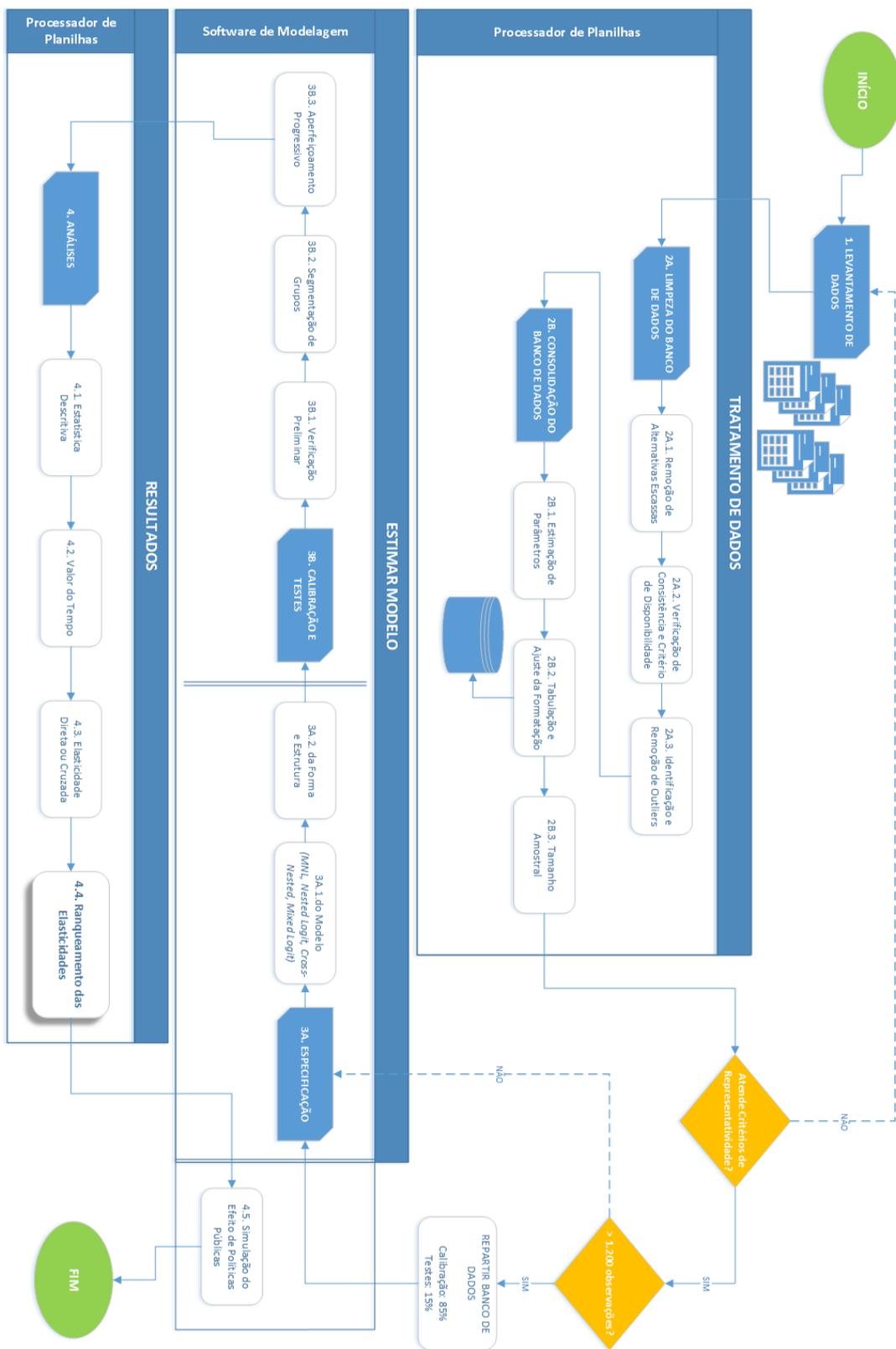


Figura 10: Fluxograma horizontal do procedimento metodológico proposto

#### 4.1 Levantamento de Dados - Pesquisa Multipropósito UFRJ

A primeira etapa para a modelagem da divisão modal consiste no levantamento de dados. Os dados utilizados nesse trabalho foram obtidos a partir da Pesquisa Multipropósito de transporte aplicada de forma *online* em 2015, no Campus da Cidade Universitária da Universidade Federal do Rio de Janeiro e disponibilizada abertamente (CARVALHO, 2016).

Embora essa pesquisa tenha sido aplicada inicialmente para analisar a acessibilidade à Cidade Universitária, a elaboração de seu questionário foi planejada com o objetivo de obter amplo conhecimento sobre os usuários do Campus Universitário. Além da caracterização socioeconômica do participante, podem-se obter dados sobre as características do modo de transporte e sobre o ambiente (origem e destino). Assim, foi possível obter algumas das principais variáveis explicativas identificadas na revisão bibliográfica sobre a modelagem da divisão modal.

O banco de dados resultante da pesquisa contou com a participação de 4.085 respondentes e ainda foi complementado pela análise de acessibilidade realizada por Carvalho (2016), que adotou procedimento similar ao proposto pelo ITDP para calcular o indicador de “Pessoas Próximas ao Transporte Público” (do inglês, *People Near Transit – PNT*).

##### 4.1.1 Elaboração do Questionário

O questionário é composto por 28 (vinte e oito) perguntas, divididas basicamente em três blocos de informações: 1) socioeconômicas, como idade, gênero e renda; 2) caracterização das viagens, como sua frequência, origem e motivo; e 3) sobre os modos de transporte, como custo, duração e distância de caminhada.

O questionário apresenta algumas perguntas de preenchimento obrigatório, mas é construído contendo uma sequência lógica. Assim, dependendo da resposta selecionada em determinada questão, o participante é direcionado diretamente para outra. Dessa forma, nem todos os participantes preencheram todas as 28 questões, embora a participação possa ser considerada como completa. O Anexo B apresenta a mensagem motivacional para estimular a participação da pesquisa, enviada por meio do sistema oficial da Universidade, SIGA, o que contribuiu para a elevada quantidade de respostas: 4.085 participações concluídas integralmente.

A Tabela 9 sumariza as principais variáveis explicativas identificadas na revisão bibliográfica que puderam ser obtidas pela aplicação da pesquisa, relacionando-as com as perguntas do questionário.

**Tabela 9: Principais variáveis explicativas obtidas na pesquisa de Carvalho (2016)**

#	Grupo de Características	Variável Explicativa	Questionário (Carvalho, 2016)
1	Indivíduo	Gênero	Pergunta 2
2	Indivíduo	Idade	Pergunta 3
3	Indivíduo	Renda Domiciliar	Pergunta 4
4	Indivíduo	Posse de Habilitação	Pergunta 5
5	Indivíduo	Quantidade de Carros no Domicílio	Pergunta 6
6	Modo	Custos	Pergunta 27
7	Modo	Tempo de Caminhada	Pergunta 28
8	Modo	Tempo de Viagem (total)	Perguntas 19 (ida) e 21 (volta)
9	Ambiente	Destino	Pergunta 13
10	Ambiente	Origem	Perguntas 7 a 10
11	Outros	Frequência de Deslocamentos	Pergunta 17

Assim, dentre as principais variáveis explicativas apresentadas na Tabela 4 e outras identificadas na revisão da literatura, o questionário aplicado permitiu observar 11 (onze) atributos importantes para a modelagem da divisão modal.

No entanto, conforme comentado anteriormente, o banco de dados utilizado nessa dissertação contou ainda com a complementação de um indicador *Proxy* para avaliar a acessibilidade de transporte público no local de origem dos deslocamentos informados pelos participantes, segundo calculado por Carvalho (2016).

#### 4.1.2 Indicador de Acessibilidade – *PNT*\*

Em função da baixa disponibilidade de dados oficiais sobre a acessibilidade de transporte, o Instituto de Políticas de Transporte e Desenvolvimento (da sigla em inglês, ITDP) propôs um indicador que funciona como um *Proxy* da acessibilidade ao transporte público. Destaca-se que esse atributo faz parte das características do ambiente e pode ser incorporado na modelagem da divisão modal.

O ITDP considera que uma pessoa está disposta a caminhar até um quilômetro para acessar um sistema de transporte de média e alta capacidade (com qualidade,

segurança e modicidade tarifária). Esse deslocamento máximo, tido como aceitável, é equivalente a uma caminhada que varie de 10 a 15 minutos.

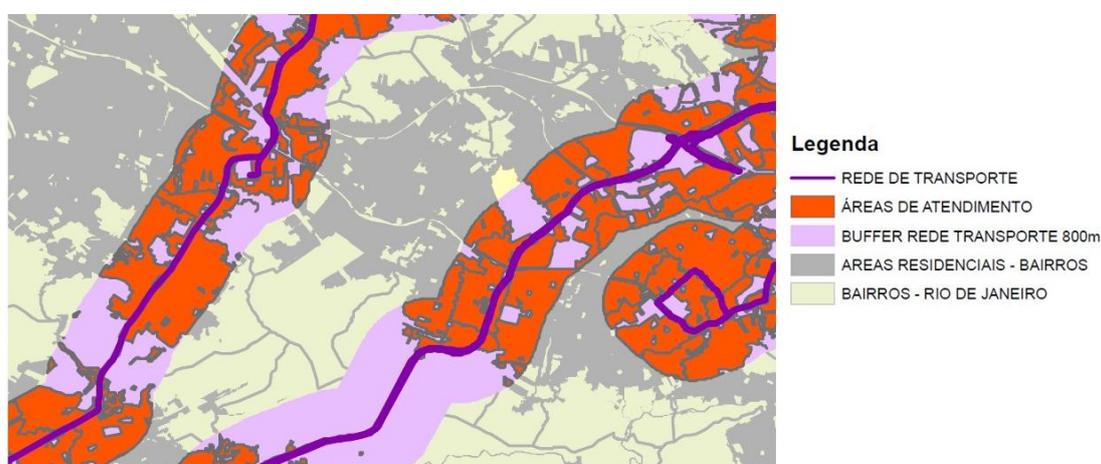
Assim, o indicador *PNT* aponta o percentual de pessoas que residem no entorno das estações de transportes de média e alta capacidade, delimitado por um raio de um quilômetro de distância a partir das estações, dividido pelo total da população da cidade ou município (ITDP, 2015).

Com base nesse procedimento proposto pelo ITDP, Carvalho (2016) realizou algumas alterações para calcular o indicador *PNT* dos bairros de origem dos respondentes do questionário aplicado:

- ❖ Para corrigir o fato de que os deslocamentos não ocorrem em uma linha reta, ao invés de considerar a distância total de um quilômetro, adotou 800 metros como o raio de deslocamento equivalente;
- ❖ Ao invés de limitar a área de cobertura como o entorno de estações de transporte de média e alta capacidade, considerou o entorno de toda a trajetória das linhas de ônibus municipais e intermunicipais que passam pela Cidade Universitária.

Ou seja, o indicador *PNT* foi calculado para cada bairro de origem, dado pelo percentual das pessoas que residem no entorno de interesse, delimitado por um raio de 800 metros ao longo das linhas de ônibus que chegam até a Cidade Universitária, dividido pelo total de moradores de cada bairro.

A Figura 11 ilustra o procedimento de análise espacial utilizado por Carvalho (2016) para obter o indicador *PNT*.



**Figura 11: Ilustração da análise espacial para cálculo do indicador *PNT* (CARVALHO, 2016)**

## 4.2 Tratamento do Banco de Dados

Para superar a limitação imposta pela pesquisa de preferência revelada, essa dissertação adota duas abordagens distintas para estimar valores referentes às alternativas não observadas:

Em função da reduzida quantidade de respostas selecionando exclusivamente as demais alternativas, o processo de escolha se resume basicamente a duas alternativas: 1) condução do automóvel ou 2) utilização do sistema de transporte público.

No entanto, a construção dos bancos de dados para as duas diferentes abordagens é feita de forma específica, detalhada a seguir.

### 4.2.1 Remoção de Participações Incompletas

Inicialmente a pesquisa utilizada contou com 4.085 entrevistas. No entanto, verificou-se que dezesseis participações estavam incompletas e precisavam ser removidas. Uma não informava a renda familiar; duas não informavam o tempo de caminhada; enquanto seis pessoas, mesmo afirmando possuir CNH e, portanto, sendo direcionadas à questão sobre a quantidade de veículos no domicílio, não responderam esta pergunta. Além disso, sete pessoas não informaram o modo de transporte utilizado para chegar à Cidade Universitária, sendo também eliminadas.

Assim, após essa verificação inicial que vale para a construção de ambos os bancos de dados, independente da abordagem de estimação adotada, o banco de dados contava com 4.069 observações completas.

### 4.2.2 Remoção de Alternativas Escassas

Determinada como premissa inicial, define-se que o modo de transporte selecionado para o trajeto de ida é mantido para a volta. Assim, descartam-se os campos referentes ao modo de transporte utilizado para sair da Cidade Universitária, considerando apenas as alternativas de chegada.

A Tabela 10 resume a frequência de observações em que as alternativas escassas foram selecionadas exclusivamente, assim como em combinação com demais modos.

**Tabela 10: Frequência de alternativas escassas**

	<b>Carona</b>	<b>Moto</b>	<b>Táxi</b>	<b>Bicicleta</b>	<b>Apé</b>
Exclusivamente selecionada	34	12	11	2	4
Presente em combinação	369	29	63	19	187

Dessa forma, removendo então as alternativas escassas, eliminam-se as 63 observações exclusivas referentes às opções apresentadas na tabela. Totalizando 4.006 observações.

Como o banco de dados #1 se baseia nos dados observados para estimar demais valores da alternativa não selecionada, fez-se necessário remover todas as 384 observações que continham a seleção de carona, moto ou táxi, uma vez que não se pode afirmar sobre qual destes o custo e tempo de viagem informado se referem. De forma similar, 202 observações tiveram que ser removidas, pois continham a seleção da opção automóvel misturada com as demais alternativas de transporte público. Assim, após as eliminações decorrentes das alternativas com reduzida frequência de seleção, o banco de dados totaliza 3.420 observações.

#### 4.2.3 Verificação de Consistência

Conforme apresentado anteriormente, um processo de escolha somente se configura como tal caso haja a disponibilidade de pelo menos duas alternativas. Dessa forma, essa etapa da limpeza de dados verifica se as observações contidas em ambos os bancos de dados são consistentes.

Admite-se que todos os indivíduos tem acesso ao sistema de transporte público. No entanto, para que a alternativa “condução de automóvel” seja disponível, exige-se que a pessoa tenha a Carteira Nacional de Habilitação, concomitante com a existência de pelo menos um veículo no domicílio.

Assim, aplicando o critério de disponibilidade para as observações de ambos os bancos de dados, removem-se 2.124 participações. Totalizando, após a verificação de consistência, 1.296 observações.

#### 4.2.4 Identificação e Remoção de *Outliers*

Essa etapa consiste em identificar e remover os valores que, de acordo com a distribuição dos dados, podem ser considerados como extremos/“além do aceitável”; ou seja, estão demasiado afastados do valor médio. A presença de *outliers* é comum ao tratar banco de dados com elevada quantidade de observações e podem indicar a significativa variabilidade dos dados, algum erro no procedimento experimental ou mesmo a existência de grupos/populações distintas.

A abordagem de estimação de valores do Banco de Dados #1 se baseia na mediana dos dados observados, que é uma medida estatística robusta, e, portanto, minimiza a influência de valores extremos. Mesmo assim, aplica-se o tratamento de *outliers*, tendo em vista que algumas respostas apresentavam valores considerados indevidos, possivelmente decorrentes do efeito de fadiga e perda de interesse ao preencher o questionário da pesquisa; ou representavam a realidade de um subgrupo específico.

Por exemplo, 2,5% (33 respondentes) dos entrevistados responderam gastar mais de 5 horas, em média, para seus deslocamentos de ida e volta até a Cidade Universitária. Dentre esses indivíduos, identificaram-se residentes de bairros fora da Região Metropolitana do Rio de Janeiro, com acesso mais dificultado ao Campus Universitário, assim como da Tijuca, que apresenta maior acessibilidade.

Para determinar os limites extremos superior e inferior, faz-se necessário realizar alguns passos intermediários. Em primeiro lugar, identifica-se o Terceiro e Primeiro Quartil, correspondentes à representação de 75% (Q3) e 25% (Q1) dos dados observados, respectivamente. A diferença entre esses patamares, representando então 50% “central” da distribuição dos dados, compõe o valor conhecido como Interquartil (IQ).

Então, os valores tidos como aceitáveis são distribuídos em distância de até uma vez e meia o valor do Interquartil, a partir do Primeiro e Terceiro quartis, de acordo com as seguintes equações:

$$IQ = Q3 - Q1 \quad (32)$$

$$\text{Limite Superior} = Q3 + 1,5 \times IQ \quad (33)$$

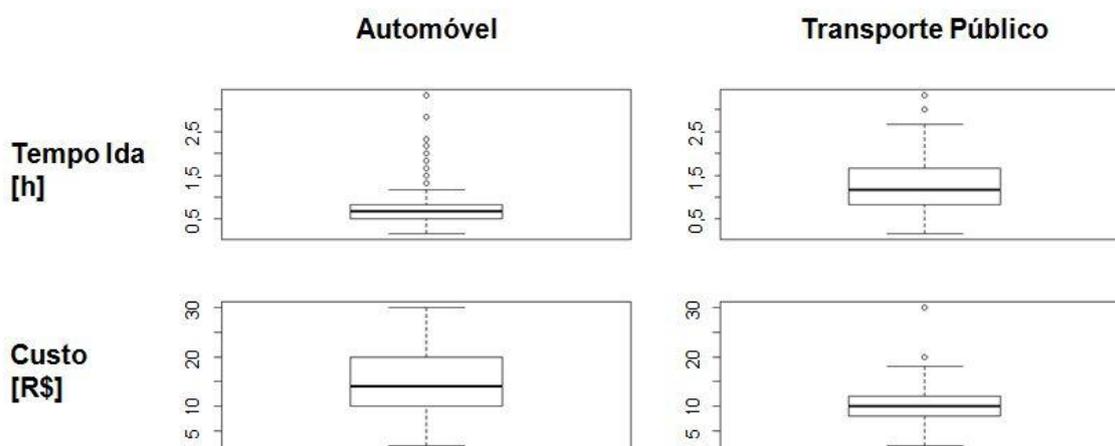
$$\text{Limite Inferior} = Q1 - 1,5 \times IQ \quad (34)$$

Aplicando o procedimento no Banco de Dados #1, para cada uma das duas alternativas possíveis, obtêm-se os limites superior e inferior para os atributos de tempo de viagem (ida) e custo, conforme apresentado na Tabela 11.

**Tabela 11: Limites Superior e Inferior de *Outlier* do Banco#1**

	Automóvel		Transporte Público	
	<i>Tempo - Ida [h]</i>	<i>Custo [R\$]</i>	<i>Tempo - Ida [h]</i>	<i>Custo [R\$]</i>
Limite Superior	1,325	35	2,93	18
Q3	0,83	20	1,67	12
IQ	0,33	10	0,84	4
Q1	0,5	10	0,83	8
Limite Inferior	0,005	-5	-0,43	2

A Figura 12 apresenta a representação gráfica dos valores extremos, identificando os limites extremos superior e inferior, além do Primeiro e Terceiro quartis. A barra horizontal representa o valor Interquartil. Tais gráficos são denominados de Box-Plot.



**Figura 12: Box-Plot dos parâmetros Tempo e Custo de viagem, por alternativa**

Aplicando os limites identificados para os parâmetros de tempo e custo de viagem, foram removidas 113 observações, conforme discriminado na Tabela 12.

**Tabela 12: Quantidade de observações removidas pela análise de Outlier**

	Automóvel		Transporte Público	
	<i>Tempo - Ida</i>	<i>Custo</i>	<i>Tempo - Ida</i>	<i>Custo</i>
Limite Superior	57	0	19	37
Limite Inferior	0	0	0	0

Após as remoções referentes aos valores identificados como *outliers*, o Banco de Dados #1 contém 1.183 observações. Já o Banco de Dados #2, por não ser afetado pela remoção de valores extremos, permanece totalizando 1.296 observações.

#### 4.2.5 Estimação de Parâmetros

Essa etapa consiste na estimação referente aos parâmetros das alternativas não selecionadas. Conforme mencionado anteriormente, o Banco de Dados #1 se baseia na mediana dos valores observados, tendo o local de origem como referência para garantir maior fidelidade aos valores a serem estimados. Dessa forma, caso determinada origem não possua observações suficiente sobre ambas as alternativas, a fórmula retorna mensagem de erro.

De fato, 114 observações tiveram que ser removidas, sendo 89 por não haver informação sobre viagens de automóvel em 46 das origens; e 25 referentes a viagens de transporte público em 12 origens. Assim, o Banco de Dados #1 totaliza 1.069 observações.

Já o Banco de Dados #2 conta com a base de dados do *Google Maps* para obter valores de distância e tempo de viagem para deslocamentos feitos de carro, sendo o custo estimado posteriormente. Para deslocamentos pelo transporte público, obtêm-se distância, tempo e custo. Ambas as alternativas de transporte são analisadas considerando o deslocamento do local de origem até a Cidade Universitária, em uma quarta-feira, com horário de chegada definido como 08 horas da manhã.

Para estimar o custo decorrente dos deslocamentos realizados de automóvel, verificaram-se duas abordagens: custo horário, conforme apresentado pelo Sistema de Custos Rodoviários, do DNIT (SICRO2); e custo por quilômetro percorrido, de acordo com o Inventário Nacional de Emissões Atmosféricas por veículos automotores rodoviários e preço médio praticado ao consumidor.

A primeira abordagem apresenta custos mais elevados, uma vez que leva em consideração para o cálculo os custos decorrentes da depreciação, impostos e seguro obrigatórios e manutenção, além do consumo de materiais para a operação veicular, como lubrificantes e combustível (DNIT, 2015).

Vale ressaltar que, de acordo com os critérios de disponibilidade estipulados na pesquisa, os indivíduos analisados já possuem, a priori, pelo menos um automóvel no

domicílio. Dessa forma, despreza-se o impacto dos custos fixos do veículo, sendo os custos variáveis (relacionados com a realização do deslocamento em si) a parcela que, de fato, exerce influência no processo de escolha modal. Dessa forma, utilizou-se a segunda abordagem para estimar os custos do automóvel.

A Tabela 13 apresenta o desempenho médio de veículos com motores movidos à gasolina e Flex (álcool e/ou gasolina), de acordo com a versão mais atualizada do Inventário Nacional de Emissões Atmosféricas por Veículos Automotores (MMA, 2014). Como pode ser visto, o desempenho médio considerado é de 11,363 quilômetros por litro de gasolina.

**Tabela 13: Desempenho médio de veículo automotor, por tipo de motor e ano - adaptado de MMA (2014)**

Ano	Tipo do Motor	
	Gasolina [km/l]	Flex Fuel [km/l]
2009	9,9	11,5
2010	10,4	12,2
2011	11,2	12,2
2012	11,3	12,2
Média	11,363	

O preço médio do litro da gasolina é obtido por meio da consulta ao portal da Agência Nacional de Petróleo, especificamente ao seu Sistema de Levantamento de Preços. Consultando os dados referentes ao preço da gasolina no município do Rio de Janeiro para os meses em que a pesquisa de acessibilidade à Cidade Universitária foi realizada (novembro e dezembro de 2015), obtém-se o preço médio de R\$3,789 por litro (ANP, 2016).

Assim, conhecendo o desempenho e preço médio da gasolina, estimou-se o custo médio do deslocamento do automóvel em função da distância estimada, sendo R\$ 0,334 por quilômetro percorrido.

Por outro lado, para corrigir o efeito do benefício tarifário concedido pelo Bilhete Único, referente à integração de ônibus, os custos do transporte público que somavam o valor de R\$7,60 foram reduzidos para R\$3,80. Embora o benefício tarifário se estenda também para a integração de trem com viagens de ônibus, o valor de R\$7,50 se confunde com os casos de ônibus e metrô, que não tem o mesmo benefício (Rio de Janeiro, 2009). Em virtude dessa maior dificuldade de identificação das viagens e pelo reduzido desconto tarifário (de R\$7,50 para R\$6,60 para viagens feitas com ônibus

intermunicipal e R\$6,50 para ônibus municipal), arbitrou-se não considerar esses descontos.

Utilizando a ferramenta API-Google Maps, automatizou-se o processo de estimação de distância e tempo de viagem para os 149 locais de origem dos deslocamentos. Vale ressaltar que, após a automatização do processo, verificaram-se manualmente as opções de transporte público, uma vez que a lógica de busca da ferramenta pela “melhor alternativa” é demasiada simples.

De fato, nove pares de deslocamentos foram alterados manualmente, considerando que apresentavam tempo de viagem ligeiramente superior (diferença inferior a 15 minutos), porém resultando em menor quantidade de transbordos e/ou menor custo; sendo eles: Centro, Cordovil, Coelho Neto, Ipanema, Leblon, Maracanã, Piedade, Vila da Penha e Vila Valqueire.

Referente aos deslocamentos realizados por transporte público, não foi possível obter para as seguintes origens: Maricá, Tanguá e “Fora da RMRJ”. Assim, fez-se necessário remover 32 observações. Logo, o Banco de Dados #2 totaliza 1.264 observações.

#### 4.2.6 Consolidação do Banco de Dados

Etapa final do tratamento dos dados que consiste no detalhamento de novas variáveis que foram criadas, com base nas observações originais, com o objetivo de complementar o processo de modelagem da divisão modal. Além disso, apresenta-se também a tabulação dos dados, essencial para compatibilizar sua utilização para a técnica de estimação de modelos.

Como o procedimento de modelagem não permite a presença de caracteres de texto (*string*) nos campos de dados, todas as informações de natureza não numérica foram categorizadas ou convertidas em valores arbitrados como representativos, descritos a seguir:

- ❖ Quantidade de Veículos: a resposta “4 ou mais” foi definida como 4;
- ❖ Relação/Função: as relações foram codificadas de 1 a 7, conforme discriminado na aba “Relação”;
- ❖ Renda Familiar: a resposta “mais de R\$30 mil” foi imputada como 35.000;

- ❖ Frequência: criou-se uma coluna computando a quantidade de vezes na semana que o indivíduo se desloca até o destino de interesse;
- ❖ Tempo de Viagem: a resposta “mais de 3 horas” foi convertida em 200 minutos. Posteriormente, converteram-se todos os valores para a unidade de hora;
- ❖ Custo: a resposta “mais que R\$20” foi definida como 30;
- ❖ Tempo de Caminhada: válido somente para o Banco #1, a resposta “menos de 5 minutos” foi definida como 2,5; enquanto “mais de 60 min” foi imputado como 70;
- ❖ Escolha Modal: criou-se essa variável para codificar a alternativa de fato selecionada pelo entrevistado, sendo 1 representando a “condução de automóvel”; e 2 equivalente ao uso do “sistema de transporte público”;
- ❖ Transformação em Binário: as respostas afirmativas foram transformadas em 1; enquanto as negativas, em 0. Isso se aplica aos parâmetros: CNH (se possui habilitação); Quinzenal e Mensal (referente à frequência de deslocamentos).

Além disso, com o objetivo de contribuir para enriquecer as informações do banco de dados a ser utilizado na modelagem da escolha modal, foram criadas variáveis específicas (com base na avaliação dos dados observados), sendo:

- ❖ Encadeamento de Viagem: assumindo valor 1 caso o indivíduo realizasse algum deslocamento intermediário, logo antes ou após chegar à Cidade Universitária;
- ❖ Destino CT: assumindo valor 1 caso o entrevistado se destinasse ao Centro de Tecnologia (que possui características diferentes dos demais centros de educação); e
- ❖ Frequência esporádica: assumindo valor 1 quando o indivíduo se destina à Cidade Universitária apenas uma ou duas vezes por semana.

Para a compatibilização final com o programa utilizado para a especificação do modelo de escolha discreta, converteu-se o banco de dados consolidado em formato de texto, com formatação específica MS-DOS. Além disso, é necessário garantir que os valores decimais atendam ao padrão norte americano, substituindo a vírgula por ponto (procedimento feito no próprio editor de texto, após conversão da planilha).

Dessa forma, ambos os bancos de dados foram devidamente consolidados. A Tabela 14 apresenta a quantidade de observações de cada Banco, assim como a distribuição das alternativas de transporte selecionadas.

**Tabela 14: Total de observações e divisão modal, por Banco de Dados**

	Total de Observações	Automóvel		Transp. Público	
		#	%	#	%
Banco #1	1.069	531	49,7%	538	50,3%
Banco#2	1.264	603	47,7%	661	52,3%

#### 4.2.7 Definição do Tamanho Amostral

O levantamento de dados consiste na obtenção de dados sobre determinada amostra da população de interesse, quando seria inviável financeiramente e/ou tecnicamente coletar dados sobre a população inteira.

O cálculo do tamanho amostral, então, consiste no exercício estatístico para determinar o número de observações que permita obter a maior quantidade de dados úteis e representativos da população de interesse, porém com o menor custo possível de coleta.

Embora exista uma formulação estatística para calcular o tamanho da amostra, esta depende da determinação de parâmetros de aceitabilidade, como a margem de erro tolerável e o nível de confiança da amostra.

Assim, o pesquisador se depara com o dilema da qualidade versus a viabilidade: se por um lado uma enorme quantidade de dados agrega maior informação e confiança à pesquisa, por outro, os recursos demandados para sua coleta podem inviabilizar sua execução. No entanto, uma reduzida quantidade de dados facilita a execução da pesquisa, mas resulta em um elevado grau de incerteza que compromete a qualidade e o mérito do próprio trabalho desenvolvido (ORTÚZAR; WILLUMSEN, 2011).

Cabe então, ao pesquisador, definir os parâmetros de aceitabilidade que permitam obter um tamanho amostral razoável. Quando não se conhece maiores detalhes sobre a população (como, por exemplo, a distribuição e variação de determinado parâmetro), adota-se o cálculo do tamanho amostral utilizando estimadores de proporção em amostras aleatórias simples, conforme a Equação 35.

$$N' = \frac{Z_{(1-\alpha/2)}^2 \cdot \hat{p} \cdot \hat{q}}{\varepsilon^2} \quad (35)$$

Onde,  $N'$  é o tamanho amostral preliminar a ser calculado;  $Z^2$  é o quadrado do valor da tabela Normal para um nível de confiança dado pelo índice  $(1 - \alpha/2)$ ;  $\hat{p}$  e  $\hat{q}$  são as proporções da amostra para determinada categoria, sendo  $\hat{q} = 1 - \hat{p}$  (por exemplo, a proporção da população que utiliza o automóvel como modo de transporte principal para se deslocar, ou a proporção que utiliza o sistema de transporte público); e  $\varepsilon^2$  é o quadrado da margem de erro tolerável.

Tipicamente, adota-se a margem de erro aceitável em  $\pm 5\%$  e o nível de confiança em 95% (que resulta em nível de significância  $\alpha = 2,5\%$ ). Para o caso em que a proporção da categoria de interesse seja desconhecida, adota-se uma abordagem conservadora que maximiza o valor do produto das proporções – e, conseqüentemente, o do tamanho da amostra –, arbitrando-as em  $\hat{p} = \hat{q} = 0,5$ . Isso significa dizer que, se forem feitas diversas observações, pode-se afirmar estatisticamente que a margem de erro arbitrada é satisfeita em 95% dos casos (nível de confiança).

Como a pesquisa multipropósito foi aplicada por meio de questionário *on-line*, não se define antecipadamente uma quantidade limite para encerrar o levantamento de dados. Ao invés disso, adota-se uma janela de tempo como o prazo para interrupção da pesquisa. Após o amplo tratamento aplicado aos dados para sua compatibilização com a técnica de modelagem de escolha discreta, o Banco de Dados #1 totaliza 1.039 observações, enquanto o Banco #2 possui 1.139 observações.

Dessa forma, adota-se o procedimento inverso ao apresentado para o cálculo do tamanho amostral. Partindo da quantidade de participações válidas obtidas para cada um dos bancos de dados, determinam-se os parâmetros de aceitabilidade que resultariam no tamanho amostral mais próximo possível à amostra obtida.

Porém, caso o tamanho total da população  $P$  seja conhecido, pode-se ainda aplicar um fator de correção que reduz o tamanho amostral preliminar calculado, conforme apresentado na Equação 36.

$$N = \frac{N'}{1 + \frac{N'}{P}} \quad (36)$$

De acordo com Carvalho (2016), a população total da Cidade Universitária é estimada em aproximadamente 40 mil pessoas. A Tabela 15 sumariza os principais parâmetros de aceitabilidade arbitrados, assim como o tamanho amostral preliminar e definitivo compatível com cada um dos bancos de dados consolidados.

**Tabela 15: Tamanho amostral em função de diferentes parâmetros de aceitabilidade**

	Nível de Confiança	$Z^{\alpha/2}$ (1- $\alpha/2$ )	Margem de Erro	Tamanho Preliminar	Tamanho Corrigido
Caso 1	95%	1,96	3%	1.067	1.039
Caso 2	95%	1,96	3%	1.067	1.039

Assim, o Banco de Dados #1, contendo 1.069 observações, pode ser considerado estatisticamente representativo da população de estudo para 95% dos casos, com margem de erro de três pontos percentuais, para mais ou para menos.

O Banco de Dados #2, com suas 1.264 respostas válidas, também pode ser considerado estatisticamente representativo para 95% dos casos, atendendo a mesma margem de erro de três pontos percentuais, para mais ou para menos.

### **4.3 Estimar o Modelo**

A terceira etapa da modelagem da divisão modal consiste em especificar a forma funcional de um modelo para representar o fenômeno de escolha, definindo como e quais atributos influenciam a decisão. Em função da maior facilidade de aplicação e interpretação, adotou-se a especificação de um modelo Multinomial Logit com parâmetros lineares.

As decorrentes etapas foram desenvolvidas com apoio do *software* gratuito e específico para a modelagem de escolhas discretas, BioGeme, conforme elaborado por Bierlaire (2003). As etapas da construção dos modelos são apresentadas a seguir em maiores detalhes.

#### **4.3.1 Especificação/Construção**

A construção dos modelos deve adotar a abordagem do aperfeiçoamento progressivo, conforme detalhado no Capítulo 2. Essa abordagem consiste em adicionar, gradativamente, novos parâmetros à especificação do modelo e permite avaliar o impacto e contribuição de cada coeficiente, além de identificar os parâmetros que resultam em melhoria significativa para o desempenho do modelo.

A Equação 37 apresenta a forma linear da contribuição dos atributos para a função utilidade de cada alternativa disponível.

$$U(j) = \beta_0 + \beta_1 \cdot X_1 + \dots + \beta_n \cdot X_n; \forall j \in (1; 2) \quad (37)$$

De acordo com a revisão bibliográfica, identificaram-se os fatores-chave a serem incorporados na modelagem da divisão modal. Por meio da aplicação do questionário utilizado no levantamento de dados, foi possível captar 12 dos parâmetros-chaves identificados na revisão. A Tabela 16 apresenta esses e outros atributos que foram acrescentados, de forma complementar, totalizando 15 parâmetros a serem utilizados para a construção dos modelos.

**Tabela 16: Variáveis explicativas, por grupo de características**

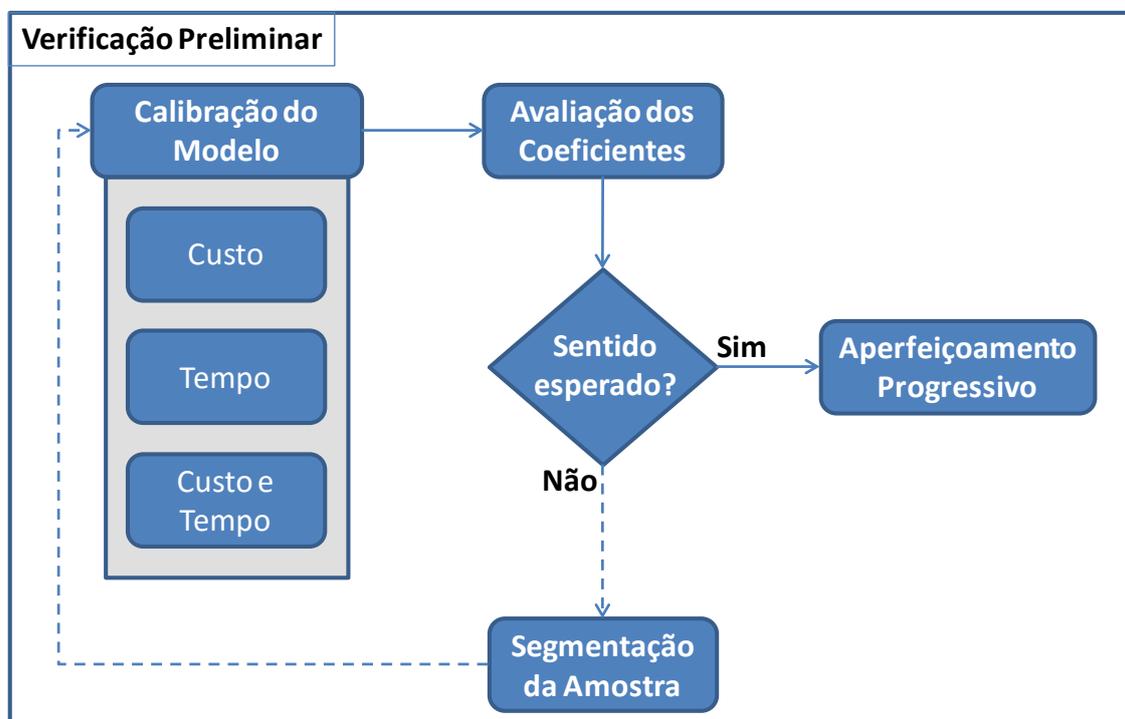
#	Grupo de Características	Variável Explicativa
1		Gênero
2		Idade
3	Indivíduo	Renda Domiciliar
4		Posse de Habilitação
5		Quantidade de Carros no Domicílio
6		Custos
7	Modo	Tempo de Caminhada
8		Tempo de Viagem
9		Destino
10	Ambiente	Origem
11		Acessibilidade Transporte Público (indicador PNT)
12	Outros	Frequência de Deslocamentos
13		Distância
14	Complementar	Encadeamento de Viagem
15		Transbordo

No entanto, antes de avançar na construção do modelo, deve-se realizar uma verificação preliminar de calibração e testes, avaliando a qualidade dos dados obtidos, o comportamento do modelo e seu desempenho.

#### 4.3.2 Calibração e Testes

A etapa de verificação preliminar consiste na avaliação informal sobre o sentido esperado de determinados coeficientes, além da validação da contribuição do coeficiente em si. Isto é, verificando se a hipótese nula pode ser rejeitada com nível de confiança de pelo menos 90%, assegurando que o coeficiente estimado seja significativo/estatisticamente diferente de zero.

A Figura 13 apresenta o procedimento do experimento realizado para avaliar preliminarmente se o modelo se comporta de forma adequada para, somente então, prosseguir com a etapa de aperfeiçoamento progressivo.



**Figura 13: Fluxograma da verificação preliminar**

Constrói-se a especificação de três modelos: 1) somente parâmetro custo; 2) somente parâmetro tempo; e 3) parâmetros custo e tempo. Para todos esses, utilizam-se coeficientes específicos para cada alternativa (1 e 2), permitindo obter melhor informação sobre seu sentido e nível de significância de contribuição. Depois, avaliam-se o sentido de contribuição dos coeficientes, verificando se estão de acordo com o esperado a priori (negativo). Como o modelo geral não se mostra adequado, procede-se à segmentação da amostra e retoma-se a avaliação dos coeficientes.

#### 4.3.2.1 Verificação Preliminar - Banco #1

A Tabela 17 sumariza os coeficientes e níveis de significância resultantes da calibração para os três modelos da verificação preliminar, considerando todas as observações do Banco #1.

**Tabela 17: Calibração dos modelos de verificação preliminar, Banco#1 - Total**

Coeficiente	BANCO #1 - TOTAL					
	Custo		Tempo		Custo + Tempo	
	Valor	p-valor	Valor	p-valor	Valor	p-valor
ASC_1	0,000	<i>fixo</i>	0,000	<i>fixo</i>	0,000	<i>fixo</i>
ASC_2	-0,892	0,00 ***	-0,811	0,00 ***	-1,140	0,00 ***
B1_CUSTO	-0,005	0,53	-	-	0,006	0,55
B1_TTIME1	-	-	-0,063	0,84	-0,041	0,91
B2_CUSTO	0,089	0 ***	-	-	0,056	0,03 **
B2_TTIME1	-	-	0,693	0 ***	0,620	0 ***

Obs.: \*, \*\*, \*\*\* representa nível de confiança de 90%, 95% e 99%, respectivamente

Como se pode observar no primeiro modelo, o coeficiente do custo do automóvel apresenta o sinal esperado (negativo), indicando que a utilidade percebida pelo usuário diminui na medida em que os custos aumentem. No entanto, tal coeficiente apresenta p-valor elevado e, conseqüentemente, não pode ser considerado estatisticamente diferente de zero. Já o custo do Transporte Público é significativo, com nível de confiança de 99%, porém apresenta sentido positivo, contrário ao esperado.

De forma similar, o segundo modelo apresenta o coeficiente do tempo de viagem do automóvel com o sentido esperado (negativo), mas não pode ser considerada significativa. Enquanto o tempo referente à Alternativa 2 é significativo, mas apresenta sentido contrário (positivo).

Já no terceiro modelo, contendo os atributos de custo e tempo, somente o coeficiente referente ao tempo de viagem do automóvel apresenta o sentido esperado, mas com p-valor elevado, não sendo estatisticamente diferente de zero. O coeficiente do custo da Alternativa 1 não pode ser considerado significativo, nem apresenta o sentido esperado. Ambos os coeficientes referentes ao transporte público são significativos, porém apresentam sentido de contribuição contrário ao esperado (positivo).

Para verificar se o modelo apresenta comportamento mais estável e adequado ao tratar grupos com comportamentos de escolha específicos, arbitrou-se a estratificação do banco de dados em função da relação do entrevistado com a Cidade Universitária. A Tabela 18 apresenta os coeficiente e níveis de significância resultantes da calibração para o terceiro modelo da verificação preliminar (custo e tempo de viagem), considerando a estratificação do Banco #1 em Alunos, composto por alunos da graduação e pós; Professores, incluindo professores e pesquisadores; e Outros,

considerando ex-alunos, funcionários administrativos e de empresas no Parque Tecnológico.

**Tabela 18: Calibração do modelo de verificação preliminar, Banco#1 – Estratificado**

<b>BANCO #1 - ESTRATIFICADO</b>						
<b>Coefficiente</b>	<b>Alunos:</b>		<b>Professores:</b>		<b>Outros:</b>	
	<b>Custo + Tempo</b>		<b>Custo + Tempo</b>		<b>Custo + Tempo</b>	
	Valor	<i>p-valor</i>	Valor	<i>p-valor</i>	Valor	<i>p-valor</i>
ASC_1	0,000	<i>fixo</i>	0,000	<i>fixo</i>	0,000	<i>fixo</i>
ASC_2	-0,916	0,00 ***	-2,940	0,14	0,268	0,82
B1_CUSTO	-0,009	0,43	0,108	0,12	0,029	0,49
B1_TTIME1	0,379	0,33	-1,010	0,61	-0,971	0,51
B2_CUSTO	0,058	0,04 **	0,388	0,07 *	-0,132	0,31
B2_TTIME1	0,683	0,00 ***	-2,130	0,09 *	0,242	0,75

Obs.: \*, \*\*, \*\*\* representa nível de confiança de 90%, 95% e 99%, respectivamente

De forma similar, os modelos estratificados também possuem comportamento inadequado. Embora seja possível identificar clara diferença na calibração dos coeficientes, indicando que o modelo de escolha possui comportamento significativamente diferente para os diferentes estratos, nenhum deles apresenta totalmente o sentido de contribuição esperado.

Considerando os professores e pesquisadores, apenas os coeficientes do tempo de viagem, de ambas as alternativas, apresentam o sentido esperado de contribuição (negativo), embora o do automóvel não seja significativo. Os coeficientes do custo de ambas as alternativas apresenta sinal positivo, contrário ao esperado.

O estrato Outros apresenta apenas os coeficientes do custo do automóvel e do tempo de viagem do transporte público com o sinal adequado (negativo). No entanto, possivelmente impactado pela menor quantidade de observações nesse estrato (73), nenhum deles pode ser considerado significativamente diferente de zero.

Já para o estrato Alunos, apenas o coeficiente do custo do automóvel possui o sentido esperado, porém com elevado *p-valor*, não pode ser considerado estatisticamente diferente de zero. Os demais coeficientes apresentam sinal contrário ao esperado, indicando que a utilidade percebida pelo indivíduo aumenta na medida em que a viagem seja mais demorada e mais cara - obviamente incoerente com a realidade.

Como se pode observar das conclusões inferidas pela verificação preliminar, nenhum destes modelos apresenta um comportamento de acordo com o esperado,

indicando que existe alguma característica peculiar sobre o comportamento de escolha modal da região de estudo, ou que a qualidade dos dados obtidos pelo questionário pode não ser adequada para o processo de modelagem. Para verificar essa segunda possibilidade, realiza-se a avaliação preliminar sobre o Banco de Dados #2, cujo processo de estimação dos parâmetros de custo e tempo descarta as respostas do questionário e se baseia em um banco de dados georreferenciado e atualizado (Google Maps).

#### 4.3.2.2 Verificação Preliminar - Banco #2

A Tabela 19 sumariza os coeficientes e níveis de significância resultantes da calibração para os três modelos da verificação preliminar, considerando todas as observações do Banco #2.

**Tabela 19: Calibração dos modelos de verificação preliminar, Banco#2 - Total**

Coeficiente	BANCO #2 - TOTAL					
	Custo		Tempo		Custo + Tempo	
	Valor	p-valor	Valor	p-valor	Valor	p-valor
ASC_1	0,000	fixo	0,000	fixo	0,000	fixo
ASC_2	-0,612	0,00 ***	-0,826	0,00 ***	-0,750	0,00 ***
B1_CUSTO	-0,088	0,00 ***	-	-	-0,056	0,35
B1_TTIME1	-	-	-3,530	0,00 ***	-3,000	0,01 ***
B2_CUSTO	0,020	0,06 *	-	-	0,039	0,00 ***
B2_TTIME1	-	-	-0,264	0,05 **	-0,579	0,00 ***

Obs.: \*, \*\*, \*\*\* representa nível de confiança de 90%, 95% e 99%, respectivamente

De acordo com o primeiro modelo, o coeficiente do custo do automóvel apresenta o sentido esperado de contribuição (negativo), além de ser significativo com nível de confiança de 99%. No entanto, embora o coeficiente do custo referente à Alternativa 2 também seja significativo (com nível de confiança de 90%), apresenta sentido contrário ao esperado.

Esse resultado indica que a utilidade percebida pelo usuário do transporte público aumenta na medida em que a tarifa fica mais cara. Embora essa interpretação seja incoerente com a realidade, tal avaliação deve ser revisada posteriormente, considerando um modelo mais avançado (com melhor desempenho), uma vez que a inclusão de atributos importantes para a modelagem podem afetar a magnitude e o sentido de contribuição de parâmetros pré-existentes, assim como sua significância (KOPPELMAN; BHAT, 2006).

Já no segundo modelo, que avalia apenas o tempo de viagem, ambos os coeficientes das Alternativas 1 e 2 são estatisticamente diferentes de zero, além de apresentarem o sentido esperado de contribuição (negativo). Ou seja, a utilidade percebida pelo usuário diminui na medida em que a viagem seja mais demorada. Pode-se observar ainda que a duração da viagem afeta os usuários de maneira distinta. De fato, a magnitude do coeficiente referente à Alternativa 1 (-3,530) é significativamente superior ao da Alternativa 2 (-0,264), indicando que a utilidade percebida pelos usuários do transporte público é menos impactada por esse atributo. Isso pode ser explicado, em parte, pela possibilidade de melhor aproveitamento do tempo de viagem, uma vez que o usuário do transporte público pode se dedicar a outras atividades ao invés da condução do veículo, como ler um livro, assistir um vídeo ou mesmo, dormir.

No terceiro modelo da verificação preliminar, que leva em consideração, concomitantemente, os atributos do custo e tempo, observa-se a manutenção dos resultados obtidos previamente. Todos os coeficientes, exceto o custo do transporte público, apresentam o sinal de acordo com o esperado (negativo), além de serem significativos com nível de confiança de 99%. Apenas o coeficiente do custo estimado do automóvel apresenta elevado p-valor (0,35), indicando que não pode ser considerado estatisticamente diferente de zero.

Porém, de acordo com Koppelman e Bhat (2006), se um atributo for importante para a modelagem (conforme identificado na revisão bibliográfica) e, além disso, apresentar o sentido de acordo com o esperado, não deve ser descartado do modelo. Nesse caso em que o atributo possui elevado p-valor (e, conseqüentemente, reduzido nível de significância), deve-se interpretar que o parâmetro exerce pouca ou nenhuma contribuição ao poder explicativo do modelo. Porém, ainda apresenta utilidade para a realização de análises posteriores, como por exemplo, a Disponibilidade em Gastar (do inglês, *Willingness to Pay*).

Embora a magnitude dos coeficientes do tempo de viagem do terceiro modelo tenha sofrido ligeira variação em relação ao segundo modelo, mantém-se a relação de menor impacto na utilidade percebida pelos usuários do transporte público, frente aos usuários do automóvel.

Conforme realizado anteriormente, arbitrou-se a estratificação do banco de dados em função da relação do entrevistado com a Cidade Universitária para verificar se o modelo apresenta resultado diferenciado ao tratar grupos com comportamentos específicos.

A Tabela 20 apresenta os coeficiente e níveis de significância resultantes da calibração para o terceiro modelo da verificação preliminar (custo e tempo de viagem), considerando a estratificação do Banco #2 em Alunos, composto por alunos da graduação e pós; Professores, incluindo professores e pesquisadores; e Outros, considerando ex-alunos, funcionários administrativos e de empresas no Parque Tecnológico.

**Tabela 20: Calibração do modelo de verificação preliminar, Banco#2 – Estratificado**

Coeficiente	BANCO #2 - ESTRATIFICADO					
	Alunos:		Professores:		Outros:	
	Valor	p-valor	Valor	p-valor	Valor	p-valor
ASC_1	0,000	fixo	0,000	fixo	0,000	fixo
ASC_2	-0,440	0,06 *	-0,589	0,66	-2,540	0,02 **
B1_CUSTO	-0,077	0,25	-0,700	0,13	0,477	0,09 *
B1_TTIME1	-2,700	0,04 **	15,400	0,07 *	-9,920	0,06 *
B2_CUSTO	0,028	0,03 **	0,034	0,57	-0,040	0,60
B2_TTIME1	-0,559	0,00 ***	-0,377	0,78	0,916	0,40

Obs.: \*, \*\*, \*\*\* representa nível de confiança de 90%, 95% e 99%, respectivamente

O estrato Alunos apresenta comportamento similar ao observado considerando todas as observações do Banco #2, com ligeiras variações na magnitude dos coeficientes. Isso pode ser explicado pelo fato de que as observações referentes aos alunos de graduação e pós representam cerca de 83% do total e, conseqüentemente, exercem significativo impacto na calibração dos coeficientes.

O estrato de Professores e Outros não apresentam o sentido esperado de contribuição, tanto para coeficientes relacionados ao custo como à duração da viagem. Enquanto o aumento do p-valor, observado nesses dois modelos, pode ser explicado pela menor quantidade de observações, que prejudica o processo de calibração.

No entanto, é importante destacar a diferença nos resultados obtidos nos modelos calibrados para os estratos Professores e Outros em relação ao de Alunos. O coeficiente do tempo do automóvel, por exemplo, difere em valores extremos, tanto no sentido de contribuição quanto em sua magnitude, assumindo o valor 15,4 para Professores e -9,92 para Outros. Já o tempo de viagem do transporte público apresenta sentido positivo para Outros.

Verifica-se, então, que existe uma significativa variação no comportamento de grupos específicos, influenciando a calibração dos modelos e, conseqüentemente, a posterior análise de resultados. Dessa forma, destaca-se a importância em identificar os grupos que devem ser modelados de forma separada, permitindo obter maior

conhecimento específico, além de mitigar a interferência de distintas características comportamentais.

Assim, encerra-se a verificação preliminar identificando o estrato dos Alunos como o grupo de observações que apresentou comportamento mais estável e coerente ao sentido de contribuição esperado dos coeficientes. Logo, aprofunda-se a modelagem somente sobre este estrato, adotando a abordagem do aperfeiçoamento progressivo.

#### 4.3.3 Aperfeiçoamento Progressivo

Conforme identificado pela etapa de verificação preliminar, referente ao Banco de Dados #2, o grupo de observações formado por alunos de graduação e pós apresenta comportamento mais estável e coerente com o esperado, podendo então prosseguir ao aprofundamento da modelagem para construir um modelo mais avançado e com adequado desempenho.

A partir do modelo utilizado na verificação preliminar, composto por atributos referente ao custo e tempo de viagem específicos para cada alternativa, acrescenta-se gradativamente variáveis explicativas. Verifica-se então se a inclusão do novo atributo apresenta aprimoramento estatisticamente significativo e se o sentido de contribuição corresponde ao esperado.

Dentre os quinze parâmetros identificados na Tabela 16, dois foram utilizados na verificação preliminar (tempo e custo de viagem) e dois foram descartados, uma vez que o processo de estimação do Banco de Dados #2 não permitiu sua obtenção (quantidade de transbordo e tempo de caminhada). Além disso, desconsiderou-se também a origem, que foi suplantada pelo indicador de acessibilidade; assim como a posse de habilitação para conduzir automóvel, tendo em vista que foi utilizado como critério de controle/disponibilidade para a construção do banco de dados.

Assim, testou-se a contribuição da inclusão de nove atributos, descritos a seguir, sendo: 1) Gênero; 2) Idade; 3) Renda Domiciliar; 4) Quantidade de veículos; 5) Encadeamento de viagens; 6) Destino; 7) Indicador de Acessibilidade; 8) Distância; e 9) Frequência dos deslocamentos.

**Tabela 21: Coeficientes e verificações estatísticas, modelos 1, 2 e 3**

BANCO #2 - ESTRATO ALUNOS						
Coeficiente	1M) Gênero		2M) Idade		3M) Renda	
	Valor	p-valor	Valor	p-valor	Valor	p-valor
ASC_1	0,000	fixo	0,000	fixo	0,000	fixo
ASC_2	-0,215	0,37	1,160	0,00 ***	1,960	0,00 ***
B1_CUSTO	-0,078	0,24	-0,070	0,03 **	-0,041	0,56
B2_CUSTO	0,026	0,05 **	0,030	0,03 **	0,016	0,24
B1_TTIME1	-2,860	0,03 **	-2,960	0,03 **	-2,790	0,04 **
B2_TTIME1	-0,575	0,00 ***	-0,559	0,00 ***	-0,350	0,09*
B0_HOMEM	-0,517	0,00 ***	-0,535	0,00 ***	-0,461	0,00 ***
B0_IDADE	-	-	-0,053	0,00 ***	-0,059	0,00 ***
B0_RENDA	-	-	-	-	-0,072	0,00 ***
<i>Obs.: *, **, *** representa nível de confiança de 90%, 95% e 99%, respectivamente</i>						
LL(inicial)	-709,46		-682,13		-658,34	
LL(final)	-682,13		-658,34		-626,91	
$\Delta K$	1		1		1	
$\chi^2$	✓ 54,66		✓ 47,59		✓ 62,85	
LL (cte)	-726,42		-726,42		-726,42	
LL(final)	-682,13		-658,34		-626,91	
Ktot - Kcte	5		6		7	
Rho <sup>2</sup>	0,06		0,09		0,14	
Rho <sup>2</sup> , ajust	0,05		0,09		0,13	

A Tabela 21 apresenta os coeficientes e níveis de significância dos atributos referentes aos modelos 1, 2 e 3 do aperfeiçoamento progressivo; além dos testes estatísticos de melhoria do modelo (Log-Verossimilhança) e seu desempenho (proporção de Log-Verossimilhança, Rho quadrado e ajustado).

A inclusão do atributo “Gênero” apresenta melhoria significativa ao poder explicativo do modelo, uma vez que o teste de Log-verossimilhança aponta um valor (54,66) superior a 3,84 (de acordo com distribuição Quadrado para nível de confiança de 95% e apenas 1 grau de liberdade, dado pela diferença da quantidade de parâmetros entre os modelos testados). Seu coeficiente também é significativo e apresenta sentido negativo de contribuição para a utilidade percebida pelo usuário do transporte público, indicando que os indivíduos do sexo masculino têm maior preferência pela utilização do automóvel. Assim, o gênero é incorporado ao modelo.

O atributo “Idade” também atende ao critério do teste de Log-Verossimilhança, provando estatisticamente possuir melhoria significativa ao poder explicativo do modelo. Seu coeficiente também é significativo e apresenta sentido negativo de contribuição, indicando que o aumento da idade resulta na redução da utilidade

percebida pelo usuário do transporte público. Esse sentido está de acordo com o esperado a priori, podendo ser justificado, na medida em que a pessoa envelhece, pela maior chance de possuir um automóvel disponível, assim como pelo maior nível de exigência relacionado ao conforto da viagem. Logo, o atributo Idade é incorporado ao modelo.

A inclusão do atributo “Renda Domiciliar” também apresenta melhoria estatisticamente significativa ao poder explicativo do modelo, sendo 62,85 muito superior ao valor crítico de teste, 3,84. Seu coeficiente é significativo e apresenta sentido negativo de contribuição, indicando que o aumento da renda domiciliar reduz a utilidade percebida pelo usuário referente à utilização do transporte público. Isso pode ser explicado pelo maior nível de exigência relacionado ao conforto e segurança da viagem, na medida em que há aumento na renda domiciliar; assim como a maior chance de possuir um automóvel disponível para realizar seus deslocamentos. Dessa forma, o atributo renda também é incorporado ao modelo.

De forma similar, a Tabela 22 apresenta os resultados da calibração e teste de desempenho referente aos modelos 4, 5 e 6 da etapa de aperfeiçoamento progressivo.

A inclusão do atributo “Quantidade de Veículos no Domicílio” resulta em aprimoramento significativo no poder explicativo do modelo, uma vez que o valor do teste de Log-Verossimilhança (67,47) é consideravelmente superior ao valor crítico (3,84). De fato, o indicador de desempenho do modelo 4M, Rho quadrado ajustado, salta de 0,13 para 0,17. Seu coeficiente é significativo e apresenta sentido negativo de contribuição, de acordo com o esperado, indicando que a utilidade percebida referente ao uso do transporte público varia de forma inversamente proporcional à quantidade de veículos. Tal comportamento pode ser explicado pelo aumento da probabilidade de disponibilidade do automóvel para realizar deslocamentos diários - na medida em que a quantidade de veículos no domicílio aumenta – facilitando a utilização do automóvel. Assim, este atributo também é incorporado ao modelo.

O aprimoramento do poder explicativo decorrente da inclusão do atributo “Encadeamento de Viagens” não pode ser estatisticamente considerado diferente de zero, uma vez que o valor do teste de Log-Verossimilhança (0,74) é inferior ao valor crítico. A princípio, espera-se que a ocorrência de encadeamento de viagens seja relacionada com o aumento da utilidade percebida pela utilização do automóvel – que garante maior flexibilidade e não impõe espera à realização do deslocamento. No entanto, além de não ser aprovado no teste do modelo, seu coeficiente apresenta sentido de contribuição positivo, ao contrário do esperado; e elevado p-valor,

indicando que não é significativamente diferente de zero. Por todos esses aspectos, este atributo é descartado, não sendo incorporado definitivamente ao modelo.

**Tabela 22: Coeficientes e verificações estatísticas, modelos 4, 5 e 6**

Coeficiente	BANCO #2 - ESTRATO ALUNOS					
	4M) Qtd. Veículos		5M) Encad. Viagem		6M) Destino (CT)	
	Valor	p-valor	Valor	p-valor	Valor	p-valor
ASC_1	0,000	fixo	0,000	fixo	0,000	fixo
ASC_2	3,250	0,00 ***	3,230	0,00 ***	3,560	0,00 ***
B1_CUSTO	-0,100	0,18	-0,104	0,16	-0,105	0,16
B2_CUSTO	0,025	0,08 *	0,025	0,09 *	0,022	0,14
B1_TTIME1	-2,210	0,12	-2,140	0,13	-2,330	0,10 *
B2_TTIME1	-0,438	0,04 **	-0,434	0,05 **	-0,456	0,04 **
B0_HOMEM	-0,424	0,00 ***	-0,431	0,00 ***	-0,264	0,07 *
B0_IDADE	-0,072	0,00 ***	-0,073	0,00 ***	-0,077	0,00 ***
B0_RENDA	-0,042	0,00 ***	-0,041	0,00 ***	-0,035	0,00 ***
B0_QTDVEIC	-0,812	0,00 ***	-0,813	0,00 ***	-0,838	0,00 ***
B0_TRIPCHAIN	-	-	0,132	0,39	-	-
B0_DESTINOCT	-	-	-	-	-0,663	0,00 ***
<i>Obs.: *, **, *** representa nível de confiança de 90%, 95% e 99%, respectivamente</i>						
LL(inicial)	-626,91		-593,18		-593,18	
LL(final)	-593,18		-592,81		-583,42	
$\Delta K$	1		1		1	
$\chi^2$	✓ 67,47		✗ 0,74		✓ 19,52	
LL (cte)	-726,42		-726,42		-726,42	
LL(final)	-593,18		-592,81		-583,42	
Ktot - Kcte	8		9		9	
Rho <sup>2</sup>	0,18		0,18		0,20	
Rho <sup>2</sup> , ajust	0,17		0,17		0,18	

Já o modelo 6M, que inclui o atributo “Destino”, apresenta aprimoramento significativo. Essa variável *dummy* verifica se o entrevistado se destina, especificamente, ao Centro de Tecnologia - CT. Como as características deste local se diferenciam dos demais centros de estudo (apresentando grande oferta de vagas de estacionamento, sem haver cobrança), acredita-se, a princípio, que influencie a escolha modal, aumentando a utilidade percebida referente à utilização do automóvel. Tal expectativa é corroborada pelo seu coeficiente, que é significativo com nível de confiança de 99% e apresenta sentido negativo de contribuição para a utilidade do transporte público, conforme esperado.

Embora o primeiro teste estatístico indique aprimoramento significativo, o indicador de desempenho sofre apenas ligeira melhoria, de 0,17 para 0,18. Por isso, tal atributo poderia ser posteriormente descartado, justificado pelo intuito de buscar

maior concisão do modelo. No entanto, a inclusão deste atributo promoveu ainda maior estabilidade ao modelo, alterando os demais coeficientes, como por exemplo, o tempo de viagem referente à Alternativa 1, que apresenta sentido esperado de contribuição e passou a ser significativa, com nível de confiança de 90%; e o custo da viagem referente à Alternativa 2, que não apresenta o sentido esperado de contribuição e passou a não ser significativa, mesmo com nível de confiança de 90%. Assim, o atributo do modelo 6M, que verifica se as viagens se destinam ao CT, deve ser incorporado definitivamente no modelo.

A Tabela 23 apresenta os resultados, conforme detalhados anteriormente, referentes aos últimos três modelos da etapa de aperfeiçoamento progressivo, 7, 8 e 9.

**Tabela 23: Coeficientes e verificações estatísticas, modelos 7, 8 e 9**

BANCO #2 - ESTRATO ALUNOS						
Coeficiente	7M) Acessibilidade		8M) Distância		9M) Frequência	
	Valor	p-valor	Valor	p-valor	Valor	p-valor
ASC_1	0,000	fixo	0,000	fixo	0,000	fixo
ASC_2	3,500	0,00 ***	3,570	0,00 ***	3,540	0,00 ***
B1_CUSTO	-0,110	0,15	-0,113	0,14	-0,103	0,17
B2_CUSTO	0,021	0,15	0,022	0,13	0,020	0,17
B1_TTIME1	-2,280	0,12	-2,210	0,13	-2,280	0,12
B2_TTIME1	-0,444	0,05 **	-0,497	0,03 **	-0,425	0,06 *
B0_HOMEM	-0,264	0,07 *	-0,265	0,07 *	-0,260	0,08 *
B0_IDADE	-0,077	0,00 ***	-0,077	0,00 ***	-0,073	0,00 ***
B0_RENDA	-0,035	0,00 ***	-0,034	0,00 ***	-0,034	0,00 ***
B0_QTDVEIC	-0,838	0,00 ***	-0,834	0,00 ***	-0,849	0,00 ***
B0_TRIPCHAIN	-	-	-	-	-	-
B0_DESTINOCT	-0,665	0,00 ***	-0,664	0,00 ***	-0,688	0,00 ***
B0_ACCESS	0,066	0,75	-	-	-	-
B0_DISTANCIA	-	-	0,014	0,55	-	-
B0_FREQUENC	-	-	-	-	-0,353	0,06 *
<i>Obs.: *, **, *** representa nível de confiança de 90%, 95% e 99%, respectivamente</i>						
LL(inicial)	-583,42		-583,42		-583,42	
LL(final)	-583,37		-583,24		-581,68	
$\Delta K$	1		1		1	
$\chi^2$	✗	0,11	✗	0,36	✗	3,48
LL (cte)	-726,42		-726,42		-726,42	
LL(final)	-583,37		-583,24		-581,68	
Ktot - Kcte	10		10		10	
Rho <sup>2</sup>	0,20		0,20		0,20	
Rho <sup>2</sup> , ajust	0,18		0,18		0,19	

A inclusão do atributo “Acessibilidade – Indicador PNT” não apresenta aprimoramento significativo ao poder explicativo do modelo, uma vez que o valor do teste de Log-Verossimilhança (0,11) é inferior ao valor crítico (3,84). A priori, acredita-se que o maior valor do indicador de acessibilidade, PNT, influencie diretamente aumentando a predisposição de utilização do transporte público. O sentido de contribuição do coeficiente é positivo, de acordo com a expectativa, no entanto seu elevado p-valor não permite rejeitar a hipótese nula do teste de significância, indicando que não pode ser considerado estatisticamente diferente de zero. Dessa forma, embora a revisão bibliográfica aponte esse atributo como importante para influenciar o comportamento de escolha modal, o parâmetro referente à acessibilidade deve ser descartado do modelo. Isso pode ser explicado por um comportamento peculiar dos indivíduos considerados na modelagem, indicando que não são sensíveis à maior facilidade de acesso ao sistema de transporte público para realizarem seus deslocamentos; ou ainda, pode significar que o indicador adotado não foi capaz de obter, de forma adequada, informação sobre a acessibilidade.

A inclusão do atributo “Distância”, de forma genérica para ambas as alternativas, também não contribui para aumentar o poder explicativo do modelo, uma vez que o valor do teste de Log-Verossimilhança (0,36) é inferior ao valor crítico (3,84). Além disso, seu coeficiente apresenta elevado p-valor (0,55), não podendo ser considerado estatisticamente diferente de zero. Embora a inclusão deste atributo possibilite realizar análises interessantes (principalmente para modelos que contenham alternativas motorizadas e não motorizadas), como a representação gráfica da distribuição da probabilidade acumulada de escolha modal em função da distância, esse parâmetro é descartado do modelo.

Já o modelo 9M apresenta o resultado da inclusão do último atributo a ser verificado, referente à “Frequência dos deslocamentos”. Foram testadas diversas variáveis *dummy*, como por exemplo, se as viagens são realizadas mais de três vezes por semana (esperando-se maior utilidade para o transporte público) ou se realizadas apenas no final de semana (esperando-se maior utilidade referente ao automóvel). No entanto, nenhuma das abordagens referentes à frequência do deslocamento apresentou aprimoramento significativo. Para facilitar a avaliação, apresenta-se apenas o resultado da inclusão da variável que indica a realização de duas ou apenas uma viagem por semana.

Como pode ser visto na coluna do modelo 9M, este atributo não apresenta aprimoramento significativo. Acredita-se, a priori, que a menor frequência da

realização de viagens aumentaria a utilidade percebida pelo uso do automóvel, uma vez que resultaria em menor parcela relativa do custo, além de não expor o indivíduo às dificuldades diárias relacionadas com a condução do veículo, como longos períodos de congestionamento, procura por vaga de estacionamento, constante manutenção veicular (abastecer, calibrar pneus, limpeza dos vidros etc.), dentre outros.

Embora seu coeficiente seja significativo (com nível de confiança de 90%) e o sentido negativo de contribuição esteja de acordo com o esperado, o indicador de desempenho,  $R^2$  quadrado ajustado, não apresenta incremento, o que reforça a conclusão do teste inicial. Dessa forma, o último atributo, referente à frequência, também não é incorporado ao modelo.

Assim, após analisar todos os atributos, encerra-se o aperfeiçoamento progressivo. Prossegue-se então à descrição formal do modelo e sua avaliação final.

#### 4.4 Modelo Estimado e Verificação de Desempenho

Assim, após concluir a etapa de aperfeiçoamento progressivo, identifica-se que o melhor modelo (mais estável, melhor desempenho e conciso) é o 6M, composto pelas seguintes equações da função utilidade das Alternativas 1 (automóvel) e 2 (transporte público):

$$U(1) = ASC_1 + B1\_CUSTO \times cost_1 + B1\_TTIME1 \times ttime_1$$

$$U(2) = ASC_2 + B2\_CUSTO \times cost_2 + B2\_TTIME1 \times ttime_2 + B0\_HOMEM \times d\_male \\ + B0\_IDADE \times age + B0\_RENDA \times income/mil + B0\_QTDVEIC \\ \times QtdVeic + B0\_DESTINOCT \times d\_ct$$

A Tabela 24 sumariza o significado dos diferentes atributos. Ressalta-se que a Alternativa 1 é adotada como base. Por esse motivo, sua função utilidade não apresenta a contribuição dos atributos genéricos, identificados pelo índice 0, como gênero, idade, renda etc.

**Tabela 24: Significado dos atributos do modelo e seus respectivos coeficientes**

<b>Coeficiente</b>	<b>Atributo</b>	<b>Significado</b>
ASC	-	Constante (representa a influência dos atributos não observados)
CUSTO	<i>cost</i>	Custo da viagem
TTIME	<i>ttime</i>	Tempo de viagem (ida)
HOMEM	<i>d_male</i>	Verifica se o gênero é masculino ( <i>Dummy</i> )
IDADE	<i>age</i>	Idade
RENDA	<i>income/mil</i>	Renda domiciliar (em milhar)
QTDVEIC	<i>QtdVeic</i>	Quantidade de veículos no domicílio
DESTINOCT	<i>d_ct</i>	Verifica se o destino é CT ( <i>Dummy</i> )

Nesse modelo, todos os coeficientes estão alinhados com o sentido de contribuição esperado; exceto o custo do transporte público. Vale destacar que algumas restrições foram relaxadas sobre essa variável, considerando os descontos tarifários provenientes do Bilhete Único somente para casos de integração entre ônibus. Esse coeficiente permanece com sentido contrário ao esperado, conforme observado anteriormente. No entanto, analisando então a versão mais avançada do modelo, de maior poder explicativo, tal atributo não é significativo para o nível mínimo de confiança de 90%. Dessa forma, deve ser interpretado como exercendo pouca ou nenhuma influência ao modelo. De fato, tal atributo só é mantido porque sua remoção promove instabilidade ao modelo, fazendo com que os coeficientes do tempo de viagem de ambas as alternativas deixem de ser significativos para o nível de confiança de 90%.

Ainda, o sinal positivo do coeficiente constante da Alternativa 2 (ASC2), que se refere à todos os atributos não captados pelo modelo, indica que existe uma maior pré disposição pela utilização do transporte público. Isto é, mantido todos os demais parâmetros idênticos para ambas as alternativas, o grupo formado por alunos da graduação e pós apresenta preferência pelo transporte público. Tal afirmação pode ser explicada pela maior receptividade a mudanças e quebra de paradigmas, característica do perfil mais jovem desse grupo; além de ser apoiada pela oportunidade de melhor aproveitamento do tempo durante a realização da viagem. De fato, a reduzida magnitude do coeficiente do tempo de viagem referente à Alternativa 2 corrobora tal interpretação.

Para testar o desempenho do modelo final, avalia-se seu poder explicativo sobre um banco de dados, verificando o percentual de predições corretas (conforme detalhado no item 2.4.4.1). Ressalta-se que tal verificação configura-se como um

interessante critério complementar para apoiar a seleção do modelo final, caso houvesse modelos com construções distintas, mas desempenho e comportamento similar.

A boa prática recomenda que esse processo de testes de um modelo seja feito sobre dados que não foram utilizados na etapa de calibração. De forma geral, realiza-se a repartição do banco de dados em duas partes: uma exclusiva para a etapa de calibração, correspondente a 85% do total; e outra exclusiva para a validação, com os 15% restantes. No entanto, para melhores resultados, admite-se que o banco de dados para calibração deve conter pelo menos mil observações; logo, para executar a repartição, o banco total deve ter pelo menos, 1.200 observações.

Para uma reduzida quantidade de dados (ou seja, inferior a 1.200 observações no total), é comum desprezar tal repartição, utilizando todas as informações disponíveis com o intuito de obter a melhor calibração possível (ORTÚZAR; WILLUMSEN, 2011). Como o banco de dados composto pelo grupo de alunos possui apenas 1.048 respostas, descarta-se a repartição do banco de dados.

Aplicando o modelo calibrado sobre o banco de dados calibração, realizam-se dois testes sobre seu desempenho de predição. Primeiro, conforme descrito no item 2.4.4.1, aplica-se a Equação 19 para calcular o Percentual de Acertos (*Hit Ratio*), admitindo-se que a alternativa que apresenta maior probabilidade de escolha será sempre escolhida e comparando-a com a alternativa de fato escolhida. Embora seja o indicador mais utilizado, conforme verificado na revisão bibliográfica, seu uso não é recomendado. De acordo com Train (2009), sua utilização se baseia em um princípio que fere o próprio conceito de se trabalhar com probabilidades; isto é, nem sempre a alternativa de maior probabilidade será escolhida. A Tabela 25 apresenta o resultado obtido por esse indicador, obtendo 70,7% de acertos.

**Tabela 25: Percentual de acertos: escolhas observadas versus projeção do modelo**

	<b>1</b>		<b>2</b>		<b>Correto Incorreto</b>	
<b>1</b>	22,1%	232	18,9%	198	70,7%	29,3%
<b>2</b>	10,4%	109	48,6%	509		

Segundo, adotando o procedimento apresentado no item 2.4.4.2, aplica-se a Equação 20 para calcular a Probabilidade Média de Predições Corretas (PMPC), computando apenas a média da probabilidade de escolha das alternativas de fato escolhidas. Tal procedimento resulta em um desempenho inferior, com probabilidade média de 62% de predições corretas.

**Tabela 26: Divisão modal real observada e estimada pelo modelo final, além de seu respectivo indicador de desempenho**

<b>Alternativa</b>	<b>Divisão Modal</b>		
	<b>Real *</b>	<b>Estimada (Hit Ratio)</b>	<b>Estimada* (PMPC)</b>
1: Automóvel	41%	33%	41%
2: Transporte Público	59%	67%	59%
Indicador de Desempenho	N/A	71%	62%

Obs.: \* indica variação a partir da quinta casa decimal

Concluindo a avaliação do modelo final obtido pela etapa de aperfeiçoamento progressivo, apresenta-se na Tabela 26 a comparação entre a divisão modal real observada da amostra e as estimadas pela aplicação do modelo no mesmo banco de dados, adotando dois diferentes procedimentos. Observa-se que o procedimento utilizado para calcular o Percentual de Acertos (*Hit Ratio*), embora indique melhor desempenho explicativo do modelo, implica na estimação de uma divisão modal significativamente diferente da real observada. Já o procedimento para o cálculo da PMPC apresenta uma estimativa agregada da divisão modal quase idêntica à real observada, comprovando a superioridade de utilização deste segundo indicador para avaliar o desempenho de predições corretas do modelo.

#### **4.5 Considerações Finais**

A primeira abordagem de estimação de parâmetros para o Banco #1 resulta em modelos com comportamento inadequado, apresentando coeficientes de tempo e custo de viagem com sentido contrário ao esperado a priori. Porém, essa limitação permitiu extrair importante observação sobre a construção de questionários. De fato, referente à pesquisas de Preferência Revelada, questiona-se a validade da resposta do entrevistado referente ao custo e tempo da viagem realizada, recomendando-se especial atenção ao utilizar tais dados no processo de modelagem.

Já a segunda abordagem de estimação de parâmetros das alternativas não selecionadas, resultando no Banco #2, apresenta significativo potencial de utilização para situações e casos reais. A ferramenta *API - Google Maps* permite automatizar o processo de estimação para diversos pares de Origem e Destino, viabilizando sua aplicação prática. Sugere-se ainda que seja aprofundada a utilização desta ferramenta de interação para desenvolvedores, ampliando o levantamento automático de dados

sobre as viagens, como por exemplo, o tempo de caminhada, a quantidade de transbordos e a própria identificação espacial de onde tais ocorrem, possibilitando posterior cruzamento de informações georreferenciadas para complementar o estudo sobre o comportamento de viagens.

Além disso, conforme destacado anteriormente, ressalta-se a importância de que, durante a elaboração de Planos de Mobilidade, a potencial diferença no comportamento de escolha modal de diferentes grupos seja levada em consideração. Essa abordagem segmentada de modelagem possibilita identificar como diferentes grupos são influenciados, permitindo formular políticas públicas orientadas especificamente para cada grupo - especialmente os identificados com maior potencial para a transferência modal.

Embora a conclusão da etapa de aperfeiçoamento progressivo não tenha sido capaz de obter um modelo com indicador de desempenho  $R^2$  quadrado ajustado superior a 0,25, tal procedimento metodológico contribui para atender a um dos objetivos específicos dessa dissertação: exemplificar o potencial da aplicação da modelagem da divisão modal. De fato, o próprio processo de calibração e verificação do desempenho decorrente da inclusão de diferentes atributos já permite obter interessante conhecimento sobre o comportamento de escolha modal do grupo focal estudado.

No entanto, é importante ressaltar que, como o modelo apresenta indicador de desempenho equivalente a 0,18, inferior ao patamar considerado como adequado, as análises posteriores realizadas a partir desse modelo não podem ser tomadas como válidas. Porém, essa pesquisa de dissertação não tem o objetivo de analisar e propor soluções específicas como se fosse um estudo de caso da Cidade Universitária. Na realidade, essa pesquisa apenas adota os dados disponíveis pela recente pesquisa de acessibilidade para exemplificar o potencial de aplicação da modelagem da divisão modal.

Assim, para atender ao objetivo específico da dissertação e exemplificar o potencial da aplicação da expansão metodológica proposta, prossegue-se a etapa de análises e avaliação dos resultados, apresentados no próximo capítulo.

## 5 RESULTADOS E ANÁLISES

Esse capítulo apresenta as análises realizadas a partir do modelo final, após o devido processo de aperfeiçoamento progressivo, de calibração e verificação de desempenho. Detalham-se os procedimentos necessários para sua realização, além de apontar seu potencial de inferência sobre o comportamento de viagens para melhor orientar a formulação de políticas públicas.

Destaca-se ainda a expansão metodológica proposta de ranqueamento dos atributos que exercem maior influência no processo de escolha modal, o que permite apoiar, de forma participativa e transparente, a etapa de priorização de investimentos, essencial para qualquer plano de mobilidade urbana.

Ao final, realizam-se algumas simulações para quantificar o impacto na divisão modal decorrente de variações específicas de diferentes atributos. Isto é, simulando o efeito de diferentes políticas públicas sobre a transferência modal desejada, atraindo usuários do automóvel para o transporte público.

### 5.1 Estatística Descritiva

A Tabela 27 apresenta um sumário descritivo dos atributos considerados na versão final do modelo de divisão modal, comparando o Banco Total com o estrato Alunos, para qual o modelo foi calibrado e validado.

**Tabela 27: Sumário dos atributos incluídos no modelo, nos bancos Total e Alunos**

Atributos	Banco Total				Banco Estratificado - Alunos				
	<i>Média</i>	<i>DP</i>	<i>Mín</i>	<i>Máx</i>	<i>Média</i>	<i>DP</i>	<i>Mín</i>	<i>Máx</i>	
Idade	29,6	11,5	18	81	26,7	8,7	18	68	
Renda [R\$]	9.923	7.899	1.000	35.000	9.268	7.707	1.000	35.000	
Quantidade de Veículos	1,5	0,8	1	4	1,6	0,8	1	4	
Custo [R\$]	Auto	6,18	3,14	0,33	25,55	6,31	3,26	0,33	25,55
	T.Púb.	8,4	6,1	0	32,59	8,7	6,3	0	32,59
Tempo [h]	Auto	0,4	0,1	0,07	1,07	0,4	0,1	0,07	1,07
	T.Púb.	1,7	0,7	0,17	6,97	1,7	0,7	0,17	6,97
Destino	CT	45%			41%				
	Não CT	55%			59%				
Gênero	Masculino	49%			48%				
	Feminino	51%			52%				
<b>Observações</b>	<b>1.264</b>			<b>1.048</b>					

Observa-se que, de acordo com o esperado, a média de idade do grupo referente aos alunos é inferior ao Banco Total. Embora o valor máximo seja de 68 anos, o menor valor médio, acompanhada do reduzido desvio padrão, indicam tal grupo com um perfil jovem.

De forma similar, a renda domiciliar dos alunos também apresenta valor médio e desvio padrão inferior ao do Banco Total, porém com ligeira redução. Tendo em vista que esse grupo é majoritariamente formado por alunos de graduação, esperava-se uma redução mais expressiva na renda, tendo em vista que por não serem ainda formados, não poderiam contribuir significativamente com a renda domiciliar. Essa reduzida diferença entre os grupos pode ser interpretada como um perfil de similar distribuição de renda domiciliar entre os diferentes grupos de usuários da Cidade Universitária.

Os demais atributos também apresentam reduzida diferença entre os valores médios, seu desvio padrão e amplitude dos dados. Destacam-se a ligeira maior quantidade de veículos no domicílio e custos relacionado ao automóvel para o grupo dos Alunos; enquanto o custo com o transporte público não apresenta variação.

Embora existam poucas observações sobre os demais indivíduos que se destinam à Cidade Universitária, tais diferenças, conforme evidenciadas pela estatística descritiva dos dados, corrobora a importância da segmentação do grupo de estudo durante a modelagem da divisão modal. Como o modelo foi calibrado e validado somente para o grupo formado por alunos da graduação e pós, as próximas análises se aplicam apenas a este segmento.

## **5.2 Análises**

As análises realizadas a partir do modelo final calibrado são objetivas e de fácil interpretação, contribuindo significativamente para aprofundar o conhecimento sobre o processo de escolha modal. De fato, as análises descritas nesse item, complementando-se entre si, apresentam o potencial para melhor orientar a formulação de políticas públicas. Portanto, configuram-se como interessante ferramenta para planejadores urbanos, de transporte e tomadores de decisão.

Destaca-se que o *software* BioGeme dispõe ainda de um aplicativo complementar, denominado BioSim, que auxilia significativamente no cálculo da elasticidade e na simulação e efeitos na divisão modal decorrente de variações

específicas em determinados atributos. A execução do BioSim segue a mesma estrutura do BioGeme, sendo necessário identificar o arquivo de modelo (".mod") e o banco de dados desejado (".dat"). Para obter o arquivo de modelo que contém os coeficientes já com seus valores devidamente calibrados, basta alterar um dos arquivos resultantes da calibração (feita pelo BioGeme), de formato ".res", renomeando manualmente sua extensão para o formato ".mod" (arquivo de modelo).

Ao ser executado, o programa aplica o modelo calibrado ao banco de dados desejado, listando para todas as observações a utilidade percebida de cada alternativa, por cada indivíduo, assim como suas respectivas probabilidades de escolha. Ao final desse processo, o programa gera um arquivo de resultado (".enu") que deve ser aberto em um processador de planilhas e concluída sua importação com o apoio do "assistente automático para importação de texto". Por último, deve-se atentar para indicar corretamente a formatação adotada para separador de milhares e de casas decimais, tendo em vista que o programa se baseia no padrão norte americano (vírgula para milhares; e ponto para decimais).

### 5.2.1 Valor do Tempo

Conforme apresentado no item 2.4.1, o Valor do Tempo é obtido pela relação do coeficiente do tempo de viagem sobre o de custo, conforme ilustrado na Equação 38.

$$VOT(j) = \frac{\beta_j \text{Tempo}}{\beta_j \text{Custo}} \quad (38)$$

Essa simples análise permite verificar a validade do modelo, por meio da comparação frente a resultados obtidos em estudos similares. Além disso, serve como um teste informal para avaliar o desempenho do modelo e verificar se o valor do tempo é coerente com o esperado.

Como o coeficiente do custo de viagem referente ao transporte público apresenta sinal contrário ao esperado, essa análise somente pode ser aplicada sobre a Alternativa 1.

Aplicando a Equação 38 sobre os coeficientes do tempo de viagem (-2,33) e custo (-0,105) do modelo final calibrado (6M), obtém-se um indicador Valor do Tempo equivalente a R\$22,19 por hora. Como o segmento de Alunos é majoritariamente formado por alunos de graduação (73%), esperava-se que esse valor fosse menor,

levemente superior à remuneração compatível com o trabalho de estágio remunerado (em torno de R\$12 por hora), tendo em vista que tais alunos não são formados.

No entanto, como destacado anteriormente, esse modelo não atende ao critério de desempenho, apresentando  $\bar{\rho}^2$  igual a 0,18, inferior ao patamar tido como adequado (0,25). Assim, embora este atributo provavelmente sofra influência da renda familiar, acredita-se que um modelo calibrado com uma maior e melhor base de dados, com  $\bar{\rho}^2$  superior a 0,25, resultaria em um VOT mais compatível com a realidade. Ressalta-se então o caráter ilustrativo dessa análise.

### 5.2.2 Elasticidade Direta e Cruzada

Conforme detalhado no item 2.1.4, a análise da elasticidade direta (e cruzada) apresenta significativo potencial analítico, permitindo obter maior conhecimento sobre o processo de escolha modal, embora seja aplicada em apenas 38% dos casos identificados na revisão bibliográfica. A elasticidade representa a variação percentual da probabilidade de escolha de uma alternativa, decorrente da variação em um ponto percentual de determinado atributo.

Para seu cálculo, adota-se o procedimento recomendado por (TRAIN, 2009), obtendo a média das diferenças percentuais da probabilidade de escolha individual, para cada alternativa. Assim, com o apoio do BioSim, simula-se o incremento sistemático de um, dois, cinco e dez pontos percentuais sobre os atributos de natureza contínua, computando a diferença resultante entre a probabilidade de escolha simulada e a original, para cada indivíduo e para cada incremento. Depois, calcula-se a média da variação de todos os indivíduos, para cada incremento, normalizando-as para um ponto percentual. Assim, obtém-se a elasticidade pela média dos quatro cenários de incremento simulados.

Para atributos de natureza discreta, como idade e quantidade de veículos no domicílio, adota-se um procedimento similar. Porém, ao invés de simular o efeito do incremento percentual, utiliza-se o incremento unitário. Ou seja, adicionam-se um, dois, cinco e dez anos aos valores originais do banco de dados, computando a diferença entre as probabilidades de escolha decorrentes de tal incremento.

Já para atributos representados por variáveis *dummy*, que assumem valor um ou zero, adota-se um procedimento diferenciado, uma vez que seria incoerente promover incremento de valores binários. Adota-se então a estratégia da

segmentação, calculando a média da probabilidade de escolha individual, para cada alternativa, condicionada aos critérios da própria variável *dummy*. Obtém-se a elasticidade, então, pela diferença das probabilidades de escolha entre os segmentos. Por exemplo, calcula-se a média da probabilidade de utilização da Alternativa 1, condicionado ao critério de gênero, obtendo posteriormente a diferença entre o sexo masculino e feminino (representando pela variável *dummy* como um e zero, respectivamente).

Por se tratar de um modelo com apenas duas alternativas (Binomial Logit), a Tabela 28 sumariza ambas as Elasticidades Direta e Cruzada. Vale ressaltar que, conforme destacado anteriormente, o coeficiente do custo referente ao transporte público apresenta sentido contrário ao esperado e, conseqüentemente, sua elasticidade mantém a mesma relação. Assim, para não dificultar a interpretação dessa análise, a elasticidade desse atributo é descartada.

**Tabela 28: Elasticidade Direta e Cruzada, por atributo, por alternativa**

Atributo	Elasticidade	
	Auto	T. Pub.
* Idade	1,473%	-1,473%
** Gênero	11,575%	-11,575%
Renda	0,062%	-0,062%
* QtdVeic	14,743%	-14,743%
** Destino(CT)	16,703%	-16,703%
Tempo_1	-0,165%	-
Custo_1	-0,116%	-
Tempo_2	-	-0,144%
Custo_2	-	<b>Descartada</b>

Obs.: Variável discreta (\*) e *dummy* (\*\*)

Vale ressaltar que a análise da elasticidade é normalizada pelo valor e unidade dos próprios atributos, sendo apresentada, de forma agregada, pela média das variações individuais. Logo, deve ser interpretada como o efeito do incremento em relação ao valor médio dos atributos, conforme apresentado na Tabela 27 (estatística descritiva); dado que os demais atributos permaneçam constantes (condição tipicamente denominada, em latim, de *ceteris paribus*).

De acordo com os resultados obtidos, cada ano acima da idade média (cerca de 26 anos) resultaria no aumento de 1,473 por cento na probabilidade de utilização do automóvel. Esse comportamento está alinhado à interpretação feita a partir dos coeficientes estimados do modelo final (6M), indicando uma tendência comportamental

das pessoas com mais idade em apresentar maior probabilidade de utilização do automóvel.

O gênero, por se tratar de um atributo de natureza binária, apresenta impacto de maior ordem de grandeza quando comparado à variação percentual de atributos contínuos. Observa-se que existe um comportamento notavelmente distinto em função do gênero do entrevistado: os indivíduos do sexo masculino apresentam probabilidade de escolha do automóvel 11,6% maior do que o sexo feminino. Esse coeficiente é significativo com nível de confiança de 90%.

Também de natureza binária, o atributo “Destino (CT)”, que verifica se o entrevistado se destina ao Centro de Tecnologia, apresenta significativa influência no processo de escolha modal. De acordo com lógica que motivou a criação da própria variável *dummy*, acredita-se que os indivíduos que se destinam ao CT possuem comportamento diferenciado. Tal comportamento seria influenciado por características específicas, distintas dos demais centros de ensino localizados na Cidade Universitária, como a maior quantidade de vagas de estacionamento, além de ser gratuito e possuir espaços reservados para professores e alunos. De fato, pela análise da elasticidade observa-se que as pessoas que se destinam ao CT apresentam acréscimo de 16,7 por cento na probabilidade de utilização do automóvel, frente aos demais centros de ensino do Fundão.

De forma similar, a quantidade de veículos no domicílio também se configura como atributo de natureza discreta e, portanto, apresenta impacto de maior ordem de grandeza. De acordo com o esperado a priori, acredita-se que a maior quantidade de veículos no domicílio está diretamente relacionada com a maior tendência de disponibilidade. Tal comportamento é corroborado pelo modelo, apresentando significativa influência na escolha modal: cada veículo adicional (além da média de 1,6 veículos por domicílio) representa um aumento de 14,75 por cento na probabilidade de utilização do automóvel; o segundo atributo mais expressivo.

Embora a elasticidade da renda possa parecer pouco expressiva, é importante lembrar que esse valor se refere à variação percentual de um atributo que foi incorporado no modelo na escala de milhar. Além disso, tal incremento se aplica ao valor médio da renda (cerca de R\$9.250 por domicílio). Ou seja, o incremento de cada ponto percentual na renda domiciliar (equivalente a aumento de R\$ 92,50) resultaria no acréscimo de 0,062 por cento na probabilidade de utilizar o automóvel.

Os atributos referentes às características da viagem (custo e duração), de ambas as alternativas, apresentam resultados de similar ordem de grandeza. No entanto, ressalta-se que sua variação deve ser aplicada sobre os valores médios da amostra, no qual o tempo de viagem do automóvel apresenta o menor valor e, conseqüentemente, maior impacto. De fato, observa-se que o incremento de dez por cento na duração da viagem de automóvel reduziria 1,65 por cento sua probabilidade de utilização; equivalente à variação de apenas 2,5 minutos.

Já o tempo de viagem do transporte público, com média de 1,7 horas (cerca de quatro vezes mais do que o automóvel), apresenta redução na probabilidade de sua utilização de 1,44 por cento, decorrente do incremento de 10 por cento da duração da viagem; equivalente à variação de, aproximadamente, 10 minutos. Ou seja, para a mesma variação de 2,5 minutos no tempo de viagem, a probabilidade de utilização da alternativa reduz em 1,65% e 0,35% para o automóvel e transporte público, respectivamente.

Essa sensibilidade à variação do tempo de viagem, cerca de quatro vezes menor para os usuários do transporte público, corrobora a interpretação inicial de que os indivíduos não respondem da mesma maneira à duração da viagem. Tal comportamento, conforme comentado anteriormente, pode ser explicado pela possibilidade de melhor aproveitamento do tempo ao longo do deslocamento.

Por último, o custo relacionado ao uso do automóvel, com média de R\$ 6,31 (para o trecho de ida), apresenta redução de 1,16 por cento na sua probabilidade de utilização, decorrente do incremento de dez pontos percentuais desse atributo; equivalente à cerca de 60 centavos. Isto é, para cada incremento de 50 centavos imposto ao custo do automóvel, reduzir-se-ia sua probabilidade de utilização em apenas um ponto percentual; indicando que esse atributo não exerce significativa influência no processo de escolha (considerando o comportamento médio dos indivíduos contidos na amostra).

Como pôde ser exemplificado pela análise desses resultados, a elasticidade (direta e cruzada) contribui significativamente para obter maior conhecimento sobre os atributos que influenciam o processo de escolha modal, importante para a melhor formulação de políticas públicas. No entanto, em função da característica da natureza dos próprios atributos (contínuas, discreta ou *dummy*), que apresentam elasticidades de diferentes ordens de grandeza, não é possível realizar análise comparativa de forma direta.

Assim, apresenta-se a seguir a expansão metodológica proposta para viabilizar a comparação entre atributos e, conseqüentemente, obter o ranqueamento por ordem de influência.

### 5.2.3 Contribuição Metodológica Proposta: Ranqueamento de Atributos

Conforme apresentado anteriormente, em função da distinta natureza dos atributos, não é possível comparar o resultado da elasticidade de forma direta. Apresenta-se, então, a expansão metodológica proposta que viabiliza tal comparação e permite obter o ranqueamento dos atributos que exercem maior influência na escolha modal.

O procedimento para obtenção da elasticidade para variáveis discretas simula o efeito do incremento unitário. De forma similar, para variáveis binárias avalia-se o efeito para os casos de valor igual a zero ou um. Logo, o incremento simulado pode ser interpretado como uma variação de cem por cento no atributo. Para que possam ser comparadas entre si, deve ser aplicado um fator de ajuste para considerar todos os atributos com a mesma variação percentual de incremento.

Como o conceito da elasticidade se baseia na variação percentual da probabilidade em função da variação percentual de um atributo, aplica-se, um fator de ajuste para que todas as elasticidades sejam apresentadas em relação ao incremento de um ponto percentual. Isto é, dividem-se por cem somente as elasticidades dos atributos de natureza discreta e *dummy*, possibilitando realizar a análise comparativa entre a magnitude da elasticidade de todos os atributos, mesmo que de forma virtual. Por último, para facilitar a posterior interpretação do resultado, registra-se o sentido da elasticidade, porém promove-se o ranqueamento sobre seus valores absolutos.

Esse ranqueamento permite identificar, de forma direta, as variáveis que, caso sejam alteradas (para mais ou para menos), promovem a maior variação na probabilidade de escolha de determinada alternativa, em detrimento de outras. Logo, políticas públicas que sejam orientadas especificamente para esses atributos tendem a obter resultados mais eficazes para a transferência modal desejada, como por exemplo, ampliar a utilização de transporte público ou de alternativas de transporte não motorizado.

A Tabela 29 apresenta o ranqueamento da elasticidade aplicado ao modelo calibrado para o segmento de alunos da Cidade Universitária.

**Tabela 29: Ranqueamento dos atributos por ordem de influência sobre a probabilidade de utilização do Transporte Público**

Atributo	Elasticidade		
	Auto	T. Pub.	Sentido
** Destino(CT)	0,167%	0,167%	↓
Tempo_1	-0,165%	0,165%	↑
* QtdVeic	0,147%	0,147%	↓
Tempo_2	0,144%	0,144%	↓
Custo_1	-0,116%	0,116%	↑
** Gênero	0,116%	0,116%	↓
Renda	0,062%	0,062%	↓
* Idade	0,015%	0,015%	↓

Obs.: Variável discreta (\*) e *dummy* (\*\*)

Como o modelo Binomial Logit considera apenas duas alternativas, considerou-se o ranqueamento apenas em relação à probabilidade de utilização do transporte público. No entanto, se houvessem outras alternativas (modos de transporte), seria possível adicionar lógicas secundárias, como critério de desempate, para o ranqueamento dos atributos.

Analisando o ranqueamento obtido, observa-se que o aumento do custo referente à utilização do automóvel exerce apenas moderada influência para estimular a transferência modal para o transporte público. Assim, medidas que venham impor uma cobrança adicional aos usuários do automóvel, como pedagiamento ou cobrança pelo estacionamento, além de serem impopulares e resultarem em desgaste político, promoveriam pouco resultado sobre seu uso. De fato, sua influência é similar ao gênero. Vale ressaltar que tal inferência se aplica a incrementos moderados no custo, uma vez que o modelo pode não ser representativo para aumentos de maior proporção.

Embora atributos como renda, gênero, idade e destino sejam características dos próprios indivíduos e, portanto, não possam ser diretamente influenciadas por meio de políticas públicas, seu ranqueamento permite identificar grupos de interesse. Isto é, grupos que apresentam comportamento específico e que devem ser estudados com maior atenção. Nesse caso, a renda e idade não são expressivas, enquanto o gênero apresenta moderada influência e o destino se destaca como o atributo de maior impacto.

Dessa forma, observa-se a necessidade de realizar uma campanha específica para influenciar o comportamento de escolha modal, tendo os indivíduos do sexo masculino como público alvo. Além disso, recomenda-se também que seja

aprofundada a análise sobre o comportamento dos indivíduos que se destinam ao Centro de Tecnologia, com o intuito de identificar claramente quais os fatores e características locais, e de seus usuários, que exercem expressiva influência para a utilização do automóvel.

Por outro lado, analisando as características do deslocamento, observa-se que a melhora do serviço do transporte público, de forma a reduzir seu tempo de viagem, exerce moderada influência. Já o aumento da impedância do automóvel, penalizando sua circulação e, conseqüentemente, aumentando seu tempo de viagem, exerceria impacto expressivo, sendo o atributo de segunda maior influência.

Vale ressaltar que esse impacto na escolha modal seria ampliado caso fosse adotada uma medida que afete, de forma simultânea, porém inversa, ambos os atributos do tempo de viagem. Como por exemplo, a implantação de corredor preferencial/exclusivo para o transporte público, por meio da segregação de faixa de rolamento já existente, similar ao presenciado no caso do BRT-Transoeste, no Rio de Janeiro, o que promoveria redução do tempo de viagem do ônibus e aumento da duração do deslocamento do automóvel.

Por último, destaca-se a significativa influência do atributo referente à quantidade de veículos no domicílio. Seu sentido de contribuição, representado por uma seta apontada para baixo, implica redução na probabilidade de utilização do transporte público na medida em que aumenta a taxa domiciliar de veículos. Tal comportamento pode ser explicado pela influência da maior chance de haver um veículo disponível para a realização de deslocamentos.

Assim, políticas públicas voltadas para restringir a quantidade de veículos por domicílio também exerceriam expressivo impacto para estimular a transferência modal do auto para o transporte público. Embora esse tipo de medida seja polêmica, países como Cingapura e China vem adotando-a com sucesso desde a década de 1980 (BARTER, 2005; OLSZEWSKI, 2007).

Como pôde ser exemplificado, a análise do ranqueamento se configura como uma ferramenta interessante tanto para o planejamento de transportes, como para o planejamento urbano, tendo em vista que podem ser incorporados na modelagem atributos referentes às características do ambiente construído. Como por exemplo, a existência de ciclovias, a qualidade das calçadas, ou até mesmo o parcelamento e uso do solo.

Além de sua contribuição para melhor orientar a formulação de políticas públicas mais eficazes, propõem-se ainda a aplicação dessa expansão metodológica para apoiar a priorização de investimentos, etapa importante e de presença obrigatória na elaboração de Planos de Mobilidade Urbana Sustentável. Uma vez concluído o desenvolvimento dos estudos e definidos os projetos e medidas recomendados para melhorar a qualidade e quantidade da circulação de pessoas e bens na cidade, deve-se realizar evento de audiência pública para aprovar, com participação popular, a priorização de investimentos e seus horizontes de planejamento.

Embora tal decisão seja tanto política quanto técnica, podendo ser influenciada por segmentos específicos da sociedade, melhor organizados, a utilização da expansão metodológica proposta contribui significativamente para esse processo, agregando transparência e respeitando a realidade e características locais.

Agrega transparência porque, ao se basear em um método quantitativo e de notório reconhecimento (modelagem da divisão modal por meio de modelos de escolha discreta), possui memória de cálculo que pode ser posteriormente auditada/verificada por eventuais interessados.

Respeita a realidade e características locais (do latim, *genius loci*) porque o processo de modelagem é construído a partir do levantamento de dados, consultando justamente a população do local de estudo. Assim, possibilita-se apoiar a priorização de investimentos que, de acordo com as características comportamentais da população, sejam mais eficazes para promover a transferência modal desejada. Vale destacar o devido cuidado para que, em aplicações práticas na elaboração de planos de mobilidade urbana, a amostra de observações coletadas seja suficientemente representativa dos diferentes estratos e segmentos da sociedade.

Uma vez concluída as interpretações a partir da análise da elasticidade, em função de variações percentuais de atributos aplicados aos valores médios, assim como do ranqueamento dos atributos por ordem de influência no processo de escolha, prossegue-se à etapa de projeção do efeito de políticas públicas, simulando a divisão modal para cenários específicos.

## 5.2.4 Projeção do Efeito de Políticas Públicas

Uma das principais vantagens da utilização dos modelos de escolha discreta é justamente permitir estimar a divisão modal frente à variação de atributos incorporados na modelagem.

Diferentemente da elasticidade, que avalia a variação da probabilidade de escolha modal a partir da variação percentual de atributos aplicada aos seus valores médios, esse item permite estimar a divisão modal por meio da simulação de cenários específicos, como por exemplo, a cobrança de pedágio urbano, a redução da oferta de vagas de estacionamento ou a adoção de tarifa zero para o transporte público.

Com o intuito de exemplificar o potencial dessa análise, estipulou-se a adoção de cinco cenários distintos, sendo: 1) cobrança de estacionamento, impondo acréscimo de R\$ 2 (a) e R\$5 (b) ao custo da utilização do automóvel; 2) melhoria do nível de serviço do transporte público, promovendo a redução de 15 minutos no tempo de viagem do transporte público para todos os deslocamentos; 3) afastar a oferta de vagas de estacionamento, fazendo com que, de forma geral, o tempo de viagem do automóvel aumente em 6 minutos; e 4) implantação de faixa preferencial para o transporte público nas vias de acesso à Cidade Universitária, promovendo de forma geral redução do tempo de viagem do transporte público em 30 minutos e, simultaneamente, acréscimo no deslocamento de carro em 15 minutos.

Adota-se procedimento similar ao utilizado para o cálculo da elasticidade, usando o BioSim para aplicar as variações desejadas nos atributos e, posteriormente, calculando a média da probabilidade de escolha individual, para cada alternativa.

A Tabela 30 sumariza a divisão modal simulada nos diferentes cenários.

**Tabela 30: Divisão modal original e simulada para cinco cenários**

Nome	Cenário			Divisão Modal			
	Custo	Tempo Auto	Tempo T.Pub.	Auto	T.Pub.	$\Delta$ Auto	$\Delta$ T.Pub.
Real	-	-	-	41%	59%	-	-
1a	↑ R\$ 2	-	-	37%	63%	-3,9%	3,9%
1b	↑ R\$ 5	-	-	32%	68%	-9,4%	9,4%
2	-	-	↓ 15 min	39%	61%	-2,1%	2,1%
3	-	↑ 6 min	-	37%	63%	-4,3%	4,3%
4	-	↑ 15 min	↓ 30 min	27%	73%	-14,1%	14,1%

Alinhado à magnitude de influência esperada pelo ranqueamento de atributos, os cenários que provem variação do tempo de viagem do automóvel apresentam mudança expressiva na divisão modal projetada. O Cenário 3, apenas pelo acréscimo de 6 minutos ao tempo de viagem do automóvel, estimula redução de 4,3 por cento em sua utilização. Esse desempenho é superior à cobrança de R\$2 pelo estacionamento (Cenário 1a) e à melhora do nível de serviço do transporte público, com redução de 15 minutos no seu tempo médio de viagem.

De acordo com a análise da elasticidade, que estimava uma redução de um ponto percentual na probabilidade de utilização do automóvel para cada acréscimo de 50 centavos em seu custo, a simulação do Cenário 1b (cobrança de R\$5 pelo estacionamento) permite projetar redução de 9,4% na utilização do automóvel.

A implantação de uma eventual faixa preferencial para o transporte público, com impacto simulado pelo Cenário 4, promoveria significativa transferência modal. Vale ressaltar que os valores simulados são adotados apenas com a finalidade de exemplificar o potencial de utilização do método. Logo, essa análise é especialmente interessante para a avaliação de medidas devidamente estudadas, no qual haja a melhor quantificação dos impactos esperados.

### **5.3 Considerações Finais**

Embora o banco de dados utilizado para o desenvolvimento dessa dissertação tenha sido obtido por meio de uma pesquisa multipropósito de transportes aplicada na Cidade Universitária do Rio de Janeiro, a contribuição metodológica proposta não é restrita a Campi Universitários, nem depende de características socioeconômicas ou da realidade do local de aplicação.

De fato, uma das principais vantagens do procedimento metodológico proposto é justamente sua alta replicabilidade, que pode ser aplicado em qualquer região, uma vez que as características locais e diferenças comportamentais são captadas pelo processo de modelagem da divisão modal, desenvolvido a partir de um levantamento de dados feito na própria região. Logo, sua aplicação independente das peculiaridades locais, como cultura, situação econômica, conscientização global sobre impacto da emissão de poluentes etc.

Tipicamente, a elaboração de planos de mobilidade urbana já requer a realização de pesquisas e levantamento de dados. Dessa forma, garantindo que esses

levantamentos incluam pelo menos algumas variáveis específicas, já seria suficiente para viabilizar o processo de modelagem da divisão modal.

O desenvolvimento dessa pesquisa se baseia na utilização de modelos de escolha discreta, técnica já amplamente consolidada para a modelagem da divisão modal. No entanto, a expansão metodológica proposta preenche uma lacuna identificada durante a revisão bibliográfica, viabilizando a comparação direta (mesmo que virtual) entre a elasticidade dos atributos, independente de sua natureza. Esse simples procedimento permite ranquear os atributos por ordem de influência no processo de escolha modal, facilitando a identificação dos parâmetros mais expressivos.

Além do ranqueamento melhor orientar a formulação de políticas públicas para resultados mais eficazes, sugere-se ainda uma nova aplicação para essa análise. Recomenda-se a sua aplicação para apoiar a etapa de priorização de investimentos e determinação do plano de ações, etapas obrigatórias na elaboração de planos de mobilidade urbana, agregando transparência ao processo e garantindo que a realidade e características locais (*genius loci*) sejam respeitadas e representadas, uma vez que a população do local de estudo é consultada na etapa de levantamento de dados.

## 6 CONCLUSÕES E RECOMENDAÇÕES FUTURAS

Este capítulo apresenta as principais conclusões extraídas ao final do desenvolvimento da pesquisa de dissertação. Indica também como se buscou atender aos objetivos específicos para que fosse possível atingir o objetivo geral e responder à questão central da pesquisa.

Para atender aos objetivos específicos, realizou-se a revisão da literatura sobre os modelos de escolha discreta para identificar as variáveis explicativas e os testes de desempenho utilizados com maior frequência para a modelagem da divisão modal. Esta etapa é detalhada no Capítulo 3, apresentando, no Anexo A, um quadro que discrimina os fatores-chave dos trinta e dois artigos revisados, por ordem cronológica e alfabética.

Segundo, a elaboração de um procedimento metodológico para modelar e posteriormente analisar, por meio de variáveis explicativas, o processo de escolha modal para viagens urbanas é detalhado ao longo do Capítulo 4, apresentando as etapas necessárias para o tratamento dos dados, a verificação preliminar para a identificação de segmentos com comportamentos diferenciados, e o aperfeiçoamento progressivo para a calibração do modelo. Apresenta também a etapa posterior de testes para avaliar seu desempenho.

Terceiro, a expansão metodológica proposta torna possível comparar diretamente os diversos atributos entre si, independente de sua natureza (contínua, discreta ou binária). Essa comparação permite hierarquizar os atributos por ordem de influência no processo de escolha modal, obtendo melhor conhecimento sobre o comportamento de viagens.

Quarto e último, aplica-se o procedimento metodológico proposto em um ensaio ilustrativo para exemplificar seu potencial na identificação de estratégias mais eficazes para promover a transferência modal desejada. O Capítulo 5 detalha a forma de cálculo e interpretação da elasticidade para atributos de distinta natureza, assim como a expansão proposta que permite compará-los (de forma virtual) e hierarquizar os atributos por ordem de influência no processo de escolha modal. Dentre os exemplos apresentados no item 5.2.3, foi possível identificar que, para o ensaio ilustrativo utilizado, medidas que impactem o custo de utilização do automóvel promoveriam apenas moderada influência na sua probabilidade de escolha, enquanto a duração da viagem do automóvel apresentaria influência expressiva. Logo, medidas

orientadas sobre o tempo de viagem seriam mais eficazes do que sobre o custo do automóvel. Além disso, a aplicação do método permite identificar grupos específicos que apresentam comportamento de elevada tendência à utilização do automóvel, como é o caso dos indivíduos do sexo masculino e dos entrevistados que se destinam ao Centro de Tecnologia.

O Objetivo Geral consiste em comprovar o potencial de aplicação da modelagem da divisão modal para definir medidas que promovam a transferência modal desejada de forma mais eficaz e, assim, permitam apoiar a priorização de investimentos e determinação do plano de ações em sistemas de transporte – etapas obrigatórias da elaboração de um Plano de Mobilidade Urbana Sustentável.

Assim, de acordo com o exposto, a aplicação do procedimento proposto – a partir de levantamento de dados consultando a população do local de estudo – permitiria hierarquizar os atributos por ordem de influência no processo de escolha modal, facilitando definir as medidas que promoveriam de forma mais eficaz a transferência modal desejada. Isto é, migrando usuários do automóvel para o transporte público ou para o transporte não motorizado, por exemplo. As análises decorrentes do procedimento apresentado permitem apoiar o desenvolvimento de planos de transporte em geral. No entanto, destaca-se seu potencial para apoiar, especificamente, as etapas de priorização de investimentos e determinação do plano de ações em um Plano de Mobilidade Urbana Sustentável, orientando a formulação de políticas públicas mais eficazes; motivação do desenvolvimento desta pesquisa.

Logo, a nova aplicação apresentada do procedimento proposto, na etapa de priorização de investimentos de um Plano de Mobilidade, visa responder à pergunta central da pesquisa, configurando-se como uma ferramenta para facilitar a priorização de medidas mais eficazes para melhorar a mobilidade urbana e torná-la mais sustentável.

Ressalta-se ainda que a utilização deste procedimento quantitativo agrega transparência ao processo, uma vez que a memória de cálculo pode ser posteriormente verificada; e respeita também a realidade e as características locais (*genius loci*), uma vez que a população local é consultada durante o levantamento de dados.

As análises realizadas ao longo do Capítulo 5 apoiam o planejamento de transportes. A avaliação dos resultados obtidos pelo cálculo da elasticidade permite obter conhecimento referente ao sentido e magnitude da influência dos atributos sobre

o processo de escolha modal de segmentos específicos da população, contribuindo para a avaliação e comparação de políticas públicas.

Ressalta-se que o modelo calibrado a partir dos atributos informados pelo entrevistado, Banco #1, não apresentou comportamento adequado. Na etapa de verificação preliminar, a calibração do modelo considerando os atributos custo e tempo resultava em coeficientes com sentido incoerente; isto é, contrário ao esperado. Dessa forma, questiona-se a validade e o compromisso com a realidade das respostas pelos os entrevistados em relação às características da viagem, especificamente os atributos de custo e duração da viagem.

De fato, somente foi possível dar prosseguimento à calibração do modelo ao descartar os dados informados e utilizar um procedimento alternativo para estimar o custo e tempo de viagem para deslocamentos realizados por automóvel e transporte público. Assim, recomenda-se que esses atributos não sejam incluídos em futuros questionários de preferência revelada; sendo capturados pela técnica de preferência declarada. E que, ao trabalhar com bancos de dados existentes – como é o caso desta dissertação –, seja tomado especial cuidado para verificar a coerência dos valores informados pelos entrevistados.

O procedimento alternativo para estimar informações sobre as características da viagem de ambos os modos de transporte, conforme mencionado anteriormente, contou com o apoio da ferramenta API do Google Maps. Dessa forma foi possível estimar os dados referentes ao custo e duração das viagens para determinados pares de origem e destino. Sobre a lógica de busca, especificaram-se as condições médias de circulação de uma quarta-feira (nem “pessimista”, nem “otimista”), com tempo planejado de chegada ao destino – Cidade Universitária – definido às oito horas da manhã.

Destaca-se então o potencial desta ferramenta para apoiar estudos referentes ao planejamento de transportes, uma vez que sua capacidade de automação de pesquisa e resumo dos resultados obtidos, aliada à flexibilidade de configuração dos parâmetros de busca e algoritmo de roteamento, viabiliza sua aplicação em casos reais e estudos mais complexos (elevada quantidade de pares de deslocamento).

Por último, mas não menos importante, ressalta-se a importância da segmentação da população durante a modelagem da divisão modal, reconhecendo que diferentes grupos podem apresentar comportamentos específicos. Além de obter conhecimento sobre os atributos e medidas que exerceriam maior influência sobre

diferentes grupos, essa abordagem segmentada permitiria ainda identificar grupos focais mais suscetíveis à transferência modal desejada, aumentando, por exemplo, a utilização do transporte público ou a de alternativas não motorizadas em detrimento do uso do automóvel.

Com vistas à replicabilidade na pesquisa, criou-se um repositório público – apresentado no Anexo C – compartilhando abertamente o banco de dados da pesquisa original, sua versão trabalhada para a modelagem da divisão modal e os arquivos referentes à construção dos modelos para serem executados no ambiente BioGeme. De acordo com a ética na pesquisa, as instruções para a adequada citação dos respectivos arquivos, caso sejam posteriormente referenciados, são apresentadas na página inicial do próprio repositório.

Entretanto, ao longo do desenvolvimento desta pesquisa de dissertação deparou-se com algumas dificuldades impostas ao se trabalhar com dados obtidos exclusivamente por uma pesquisa de preferência revelada. Além disso, obteve-se contato preliminar com modelos mais avançados e novos métodos para a modelagem do comportamento de viagens (*Travel Behavior*).

Apresentam-se a seguir algumas recomendações para o desenvolvimento de trabalhos futuros:

- ❖ Combinar as pesquisas do tipo Preferência Revelada e Preferência Declarada, para incorporar os benefícios de ambas, aprimorando a etapa de calibração do modelo;
- ❖ Estimar também, quando aplicados a estudos com mais de duas alternativas, os modelos Nested Logit e Mixed Logit, comparando posteriormente o desempenho obtido por cada um;
- ❖ Incorporar técnicas de “Clusterização” para identificar segmentos específicos da população que apresentem comportamento diferenciado e, portanto, devem ser modelados de forma segregada;
- ❖ Desenvolver um procedimento que facilite a informação do CEP pelos entrevistados, uma vez que esse dado possibilita obter atributos com maior precisão, ao invés de considerar apenas o centroide dos bairros;
- ❖ Aprofundar a utilização da ferramenta *API* do *Google Maps* para que seja possível obter dados complementares, como o tempo de caminhada e a quantidade de transbordo. Além de identificar, de forma georreferenciada, os trechos em que ocorrem tais movimentos, permitindo o posterior

cruzamento de informações espaciais para estudar o comportamento de viagens;

- ❖ Adotar integralmente, referente ao transporte público, o efeito dos descontos tarifários da política em vigor, considerando todas as suas modalidades de integração;

Por fim, visando obter simultaneamente informações sobre os aspectos que influenciam positiva e negativamente a transferência modal, recomenda-se a utilização da técnica *Best-Worst Scalling*, enriquecendo o estudo sobre o comportamento de viagens.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

AMOH-GYIMAH, R.; AIDOO, E. N. **Mode of transport to work by government employees in the Kumasi metropolis, Ghana.** *Journal of Transport Geography*, v. 31, p. 35–43, 2013.

ANP. **Sistema de Levantamento de Preços - Síntese dos preços praticados: Custo médio da gasolina ao consumidor.** Disponível em: <[http://www.anp.gov.br/preco/prc/Resumo\\_Mensal\\_Municipio.asp](http://www.anp.gov.br/preco/prc/Resumo_Mensal_Municipio.asp)>. Acesso em: 28 mar. 2016.

BANISTER, D. **The sustainable mobility paradigm.** *Transport Policy*, v. 15, n. 2, p. 73–80, 2008.

BANISTER, D.; BERECHMAN, Y. **Transport investment and the promotion of economic growth.** *Journal of Transport Geography*, v. 9, n. 3, p. 209–218, 2001.

BARTER, P. A. **A vehicle quota integrated with road usage pricing: A mechanism to complete the phase-out of high fixed vehicle taxes in Singapore.** *Transport Policy*, v. 12, n. 6, p. 525–536, 2005.

BEKHOR, S.; SHIFTAN, Y. **Specification and Estimation of Mode Choice Model Capturing Similarity between Mixed Auto and Transit Alternatives.** *Journal of Choice Modelling*, v. 3, n. 2, p. 29–49, 2009.

BEN-AKIVA, M. et al. **Modeling methods for discrete choice analysis.** *Marketing Letters*, v. 8, n. 3, p. 273–286, 1997.

BEN-AKIVA, M.; LERMAN, R. **Discrete Choice Analysis: theory and application to Travel Demand.** 1985.

BHAT, C. R. **Accommodating flexible substitution patterns in multi-dimensional choice modeling: formulation and application to travel mode and departure time choice.** *Transportation Research Part B: Methodological*, v. 32, n. 7, p. 455–466, 1998.

BHAT, C. R. **A multi-level cross-classified model for discrete response variables.** *Transportation Research Part B: Methodological*, v. 34, n. 7, p. 567–582, 2000.

BHAT, C. R.; SARDESAI, R. **The impact of stop-making and travel time reliability on commute mode choice.** *Transportation Research Part B: Methodological*, v. 40, n. 9, p. 709–730, 2006.

BIERLAIRE, M. **BIOGEME: A Free Package for The Estimation of Discrete Choice Models.** *Swiss Transport Research Conference*, p. 1–27, 2003.

BRASIL. **Lei 12.587/2012.** 2012.

BRASIL. **PL. 7.898/2014.** 2014.

BRASIL. **PlanMob: Caderno de referência para a elaboração de Plano de Mobilidade Urbana.** 2 ed. 2015.

CALDAS, M.; BLACK, I. G. **Formulating a methodology fro modelling revealed preference discrete choice data - The Selectively Replicated Logit Estimation.** *Transportation Research Part B: Methodological*, v. 31, n. 6, p. 463–472, 1997.

CARSON, R. T. et al. **Benefit Transfer of Environmental and Resource Values.** 1. ed. Dordrecht: Springer Netherlands, 2015.

CARVALHO, C. H. R. **Emissões relativas de poluentes do transporte motorizado de passageiros nos grandes centros urbanos brasileiros**. 2011.

CARVALHO, G. S. D. **Caracterização e análise da demanda por transporte em um Campus Universitário: o caso da UFRJ**. 2016.

CASCETTA, E. **Transportation Systems Analysis: Models and Applications**. 2. ed. Springer, 2009.

CERVERO, R. **Built environments and mode choice: toward a normative framework**. *Transportation Research Part D: Transport and Environment*, v. 7, n. 4, p. 265–284, 2002.

CIRILLO, C.; AXHAUSEN, K. W. **Evidence on the distribution of values of travel time savings from a six-week diary**. *Transportation Research Part A: Policy and Practice*, v. 40, n. 5, p. 444–457, 2006.

COLVILE, R. N. et al. **UCL Discovery - The transport sector as a source of air pollution**. *Atmospheric Environment*, v. 35, n. 9, 2001.

DAVE, S. M.; RAYKUNDALIYA, D. P.; SHAH, S. N. **Modeling Trip Attributes and Feasibility Study of co-ordinated Bus for School Trips of Children**. *Procedia - Social and Behavioral Sciences*, v. 104, p. 650–659, 2013.

DE JONG, G. et al. **A model for time of day and mode choice using error components logit**. *Transportation Research Part E: Logistics and Transportation Review*, v. 39, n. 3, p. 245–268, 2003.

DE PALMA, A.; ROCHAT, D. **Mode choices for trips to work in Geneva: an empirical analysis**. *Journal of Transport Geography*, v. 8, n. 1, p. 43–51, 2000.

DNIT. **SICRO 2 - Sistema de Custos Rodoviários**. 2015. Disponível em: <<http://www.dnit.gov.br/custos-e-pagamentos/sicro/sudeste/rio-de-janeiro/2015/setembro/rio-de-janeiro-setembro-2015>>.

DOMENCICH, T.; MCFADDEN, D. **Urban Travel Demand - A Behavioral Analysis**. Oxford: North-Holland Publishing Company Limited, 1975.

ECONOMETRIC-SOFTWARE. **Why NLOGIT is best for me?** Disponível em: <[http://www.limdep.com/features/best\\_choice.php](http://www.limdep.com/features/best_choice.php)>.

EL-BANY, M. E.-S. et al. **Policy sensitive mode choice analysis of Port-Said City, Egypt**. *Alexandria Engineering Journal*, v. 53, n. 4, p. 891–901, 2014.

ENAU, C.; GERBER, P. **Beliefs about energy, a factor in daily ecological mobility?** *Journal of Transport Geography*, v. 41, p. 154–162, 2014.

ERMAGUN, A.; SAMIMI, A. **Promoting active transportation modes in school trips**. *Transport Policy*, v. 37, p. 203–211, 2015.

FAIZ, A. **Automotive emissions in developing countries-relative implications for global warming, acidification and urban air quality**. *Transportation Research Part A*, v. 27, n. 3, p. 167–186, 1993.

FIRJAN. **Os custos da (i)mobilidade nas Regiões Metropolitanas do Brasil**. 2014.

FISHER, R. A.; YATES, F. **Statistical tables for Biological, Agricultural and Medical Research**. 6. ed. Londres: Longman Group United Kingdom, 1995.

FLÜGEL, S. et al. **Methodological challenges in modeling the choice of mode for a new travel alternative using binary stated choice data – The case of high speed rail in Norway**. Transportation Research Part A: Policy and Practice, v. 78, p. 438–451, 2015.

FONSECA, J. J. S. **Metodologia da pesquisa científica**. 2002.

GERHARDT, T.; SILVEIRA, D. **Métodos de Pesquisa**. 1 ed. Porto Alegre. 2009.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. 2007.

HABIB, K. N.; WEISS, A. **Evolution of latent modal captivity and mode choice patterns for commuting trips: A longitudinal analysis using repeated cross-sectional datasets**. Transportation Research Part A: Policy and Practice, v. 66, p. 39–51, 2014.

HAMMADOU, H.; MAHIEUX, A. **Transportation Demand Management in a Deprived Territory: A Case Study in the North of France**. Transportation Research Procedia, v. 4, p. 300–311, 2014.

HEALTH EFFECTS INSTITUTE. **Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects**. 2010.

HENSHER, D. A. **Accounting for scale heterogeneity within and between pooled data sources**. Transportation Research Part A: Policy and Practice, v. 46, n. 3, p. 480–486, 2012.

HENSHER, D. A.; ROSE, J. M.; GREENE, W. H. **Applied choice analysis**. 2 Ed Cambridge University Press, 2015.

HESS, S.; DALY, A.; ROHR, C.; HYMAN, G. **On the development of time period and mode choice models for use in large scale modeling forecasting systems**. Transportation Research Part A: Policy and Practice, v. 41, n. 9, p. 802–826, 2007.

HESS, S.; FOWLER, M; ADLER, T.; BAHREINIAN, A. **A joint model for vehicle type and fuel type choice: evidence from a cross-nested logit study**. 12th World Conference for Transportation Research, p. 1–17, 2010.

HESS, S.; BIERLAIRE, M.; POLAK, J. W. **Estimation of value of travel-time savings using mixed logit models**. Transportation Research Part A: Policy and Practice, v. 39, n. 2-3, p. 221–236, 2005.

IBGE. **Pesquisa Nacional de Amostra por Domicílio**. 2012

IBGE. **População estimada em 2015**. 2015

ITDP. **Putting People Near Transit (and Transit Near People)**. 2015. Disponível em: <<https://www.itdp.org/putting-people-near-transit-and-transit-near-people/>>.

KOPPELMAN, F. S.; BHAT, C. **A Self Instructing Course in Mode Choice Modeling : Multinomial and Nested Logit Models**. United States Department of Transportation. 2006.

KOPPELMAN, F. S.; WEN, C.-H. **The paired combinatorial logit model: properties, estimation and application**. Transportation Research Part B: Methodological, v. 34, n. 2, p. 75–89, 2000.

KRYGSMAN, S.; ARENTZE, T.; TIMMERMANS, H. **Capturing tour mode and activity choice interdependencies: A co-evolutionary logit modeling approach**. Transportation Research Part A: Policy and Practice, v. 41, n. 10, p. 913–933, 2007.

KRZYŻANOWSKI, M.; KUNA-DIBBERT, B.; SCHNEIDER, J. **Health effects of transport-related air pollution**. WHO Regional Office Europe, 2005.

MCADAM, K.; STEER, P.; PERROTTA, K. **Using continuous sampling to examine the distribution of traffic related air pollution in proximity to a major road**. Atmospheric Environment, v. 45, n. 12, 2011.

MCFADDEN, D. **Conditional logit analysis of qualitative choice behavior**. 1974.

MCFADDEN, D. **Regression-based specification tests for the multinomial logit model**. Journal of Econometrics, v. 34, n. 1, p. 63–82, 1987.

MCFADDEN, D.; TYE, W. B.; TRAIN, K. **An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model**. 1977.

MMA. **Inventário Nacional de Emissões Atmosféricas por Veículos Automotores Rodoviários**. 2014.

MÜLLER, S.; TSCHARAKTSCHIEW, S.; HAASE, K. **Travel-to-school mode choice modelling and patterns of school choice in urban areas**. Journal of Transport Geography, v. 16, n. 5, p. 342–357, 2008.

OLSZEWSKI, P. S. **Singapore motorization restraint and its implications on travel behavior and urban sustainability**. Transportation, v. 34, n. 3, p. 319–335, 2007.

ONU. **Rio +20: Fatos sobre as cidades**. Rio de Janeiro, 2012. Disponível em: <<http://www.onu.org.br/rio20/cidades.pdf>>.

ONU. **2014 Revision of World Urbanization Prospects**. 2014.

ONU. **17 Objetivos de Desenvolvimento Sustentável**. 2015.

ORTÚZAR, J. DE D.; WILLUMSEN, L. G. **Modelling Transport**. 4 ed. 2011.

PAI, M. et al. **Systematic reviews and meta-analyses: an illustrated, step-by-step guide**. The National Medical Journal of India, v. 17, n. 2, p. 86–95, 2003.

PALMA, A. DE; ROCHAT, D. **Mode choices for trips to work in Geneva: an empirical analysis**. Journal of Transport Geography, v. 8, n. 1, p. 43–51, 2000.

PAULSSEN, M. et al. **Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice**. Transportation, v. 41, n. 4, p. 873–888, 2014.

PULUGURTA, S.; ARUN, A.; ERRAMPALLI, M. **Use of Artificial Intelligence for Mode Choice Analysis and Comparison with Traditional Multinomial Logit Model**. Procedia - Social and Behavioral Sciences, v. 104, p. 583–592, 2013.

REVELT, D.; TRAIN, K. **Mixed logit with repeated choices: households' choices of appliance efficiency level**. Review of Economics and Statistics, v. 80, n. 4, p. 547–657, 1998.

RIO DE JANEIRO. **Lei Estadual 5.628/09 - Bilhete Único RJ**. 2009.

ROMÁN, C.; MARTÍN, J. C.; ESPINO, R. et al. **Valuation of travel time savings for intercity travel: The Madrid-Barcelona corridor**. Transport Policy, v. 36, p. 105–117, 2014.

SADRI, A.; UKKUSURI, S. **Analysis of hurricane evacuee mode choice behavior**. Transportation Research part C: Emerging Technologies, v. 48, p. 37–46, 2014.

SANTOS, G.; MAOH, H.; POTOGLOU, D.; VON BRUNN, T. **Factors influencing modal split of commuting journeys in medium-size European cities.** *Journal of Transport Geography*, v. 30, p. 127–137, 2013.

SCHWANEN, T.; MOKHTARIAN, P. L. **What affects commute mode choice: neighborhood physical structure or preferences toward neighborhoods?** *Journal of Transport Geography*, v. 13, n. 1, p. 83–99, 2005.

SHAHIN, S.; HÜSEYİN, T. O.; KEMAL, Ö. S. **Evaluating Transportation Preferences for Special Events: A Case Study for a Megacity, Istanbul.** *Procedia - Social and Behavioral Sciences*, v. 111, p. 98–106, 2014.

SU, Q.; ZHOU, L. **Parking management, financial subsidies to alternatives to drive alone and commute mode choices in Seattle.** *Regional Science and Urban Economics*, v. 42, n. 1-2, p. 88–97, 2012.

SUN, Q.; ZHU, L.; CHEN, B. **A Dynamic Generalized Cost Based Logit Model for Passenger Corridors.** *Journal of Transportation Systems Engineering and Information Technology*, v. 13, n. 4, p. 15–21, 2013.

TRAIN, K. **Discrete Choice Methods with Simulation.** 2. ed. Springer Netherlands, 2009.

VAN CAN, V. **Estimation of travel mode choice for domestic tourists to Nha Trang using the multinomial probit model.** *Transportation Research Part A: Policy and Practice*, v. 49, p. 149–159, 2013.

VAN, H. T.; CHOOCHARUKUL, K.; FUJII, S. **The effect of attitudes toward cars and public transportation on behavioral intention in commuting mode choice - A comparison across six Asian countries.** *Transportation Research Part A: Policy and Practice*, v. 69, p. 36–44, 2014.

VAN WEE, B.; ANNEMA, J. A.; BANISTER, D. **The Transport System and Transport Policy: an Introduction.** Londres: Edward Elgar Publishing, 2013.

## ANEXO A

Quadro dos fatores-chave identificados na revisão bibliográfica, discriminado por artigo.

#	Autor / Ano	Variáveis		Motivo Viagem	Modelo	Goodness-of-Fit			Análise
		Dependente	Independente			Pseudo-R <sup>2</sup>	% Correto	Nível Sig.	
1	Caldas e Black (1997)	Trem	Tempo de viagem (veículo)	Casa-Trabalho	SRLE (Selective Replicated Logit Estimation)	0,139	-	-	Coeficientes
		Ônibus	Tempo de viagem (espera) Custos ou Tarifa		Nested Logit	0,123			
2	Bhat (1998)	Carro	Horário de viagem Possui emprego (dummy) Idade Gênero	Casa-Trabalho	MNL				Coeficientes
		Carona	# carros por membro família Renda domiciliar Destino		Mixed MNL - T	-	-	-	
		Transporte Coletivo	Custos ou Tarifa Tempo de viagem (total) Tempo de espera / distancia da viagem		Mixed MNL - M Mixed MNL - MT				
3	Bhat (2000)	Carro Carona	Tempo de viagem (total) Tempo de viagem (espera) Custos ou Tarifa	Casa-Trabalho	MNL MGMNL	-	-	-	-
		Transporte Coletivo	# carros no domicílio Renda domiciliar		MCMNL				
4	Palma e Roachat (2000)	Carro	Tempo de viagem (total) Custos ou Tarifa	Casa-Trabalho	Nested Logit	0,525	-	90%	Elasticidade
		Carona	# anos utilizando a mesma rota Atravessa fronteira (dummy) Gênero						
		Ônibus	# pessoas no domicílio Presença de Jovem (dummy) Presença de Idoso (dummy)						
		VLT	Emprego (categorias) Educação (categorias)						
		Trem	Baixa renda (dummy) Alta renda (dummy)						

#	Autor / Ano	Variáveis		Motivo Viagem	Modelo	Goodness-of-Fit			Análise
		Dependente	Independente			Pseudo-R <sup>2</sup>	% Correto	Nível Sig.	
5	Koppelman e Hua Wen (2000)	Avião	Frequência	Casa-Trabalho	MNL	-	-	-	Coeficientes
		Trem	Custos ou Tarifa		Nested Logit				
		Carro	Tempo de viagem (veículo)		Paired Combinatorial Logit				Elasticidade Cruzada
			Tempo de viagem (espera)						Valor do Tempo
6	Cervero (2002)	Carro	Densidade (Origem)	Casa-Trabalho	Binomial Logit (Carro)	0,114	-	95%	Coeficientes
		Carona	Densidade (Destino)		Binomial Logit (Transporte Coletivo)	0,305			
			Diversidade (Origem)						
		Diversidade (Destino)							
		Transporte Coletivo	Caminhabilidade (Origem)		MNL	0,113			Elasticidade
			Caminhabilidade (Destino)						
			Tempo de viagem (total)						
			Custos ou Tarifa						
			Possui automóvel (dummy)						
			Possui carteira de habilitação (dummy)						
7	de Jong et al. (2003)	Carro	Tempo de viagem (total)	Casa-Trabalho	Mixed MNL	0,313	-	90%	Valor do Tempo
		Trem	Custos ou Tarifa		MNL	-			
			Horário de viagem (realizada - desejado)						Elasticidade
8	Hess, Bierlaire e Polak (2005)	Carro	Custos ou Tarifa	N/I	MNL	0,286	-	95%	Coeficientes
		Trem	Frequência		Mixed MNL	0,323			
		Avião	Tempo de viagem (total)						
9	Schwanen e Mokhtarian (2005)	Carro	# carros no domicílio	Casa-Trabalho	MNL	0,259	-	90%	Coeficientes
		Ônibus	Renda domiciliar						
			Idade						
			Gênero						
	Personalidade (categorias)								
	Trem	Estilo de Vida (categorias)			Nested Logit	0,268			
		Uso do Solo (origem)							
		TNM	# pessoas no domicílio						
			Emprego (categorias)						
10	Bhat e Sardesai (2006)	Carro	# carros no domicílio	Encadeamento de Viagens	MNL (PR)	0,116	-	95%	Coeficientes
		Carona	Renda domiciliar		MNL (PD)	0,028			
			Renda individual / Renda domiciliar						
			Tempo de viagem (total)						
			Custos ou Tarifa						
	Ônibus	Tempo de viagem (confiabilidade)		MNL (PR+PD)	0,047			Valor da Confiabilidade	
		Distância (total)							
			Realiza paradas no deslocamento para o trabalho (dummy)						
		Trem	Realiza paradas no meio do dia (dummy)		Mixed MNL	0,204			Efeito de Políticas Públicas
			Presença de mercado na estação de trem (binário)						

#	Autor / Ano	Variáveis			Modelo	Goodness-of-Fit			Análise
		Dependente	Independente	Motivo Viagem		Pseudo-R <sup>2</sup>	% Correto	Nível Sig.	
11	Cirillo e Axhausen (2006)	Carro	Local de moradia Idade	Encadeamento de Viagens	MNL	-	-	95%	Coeficientes
		Carona	Estado Civil (dummy) Possui automóvel (dummy)		Mixed MNL				
		Transporte Coletivo	Possui vale transportes (dummy) Tempo de viagem (total) # atividades secundárias		Mixed MNL - Panel Data				Valor do Tempo
		TNM	Custos ou Tarifa Presença de criança <6anos (dummy)						
12	Hess et al. (2007)	Carro	Tempo de viagem (total)	Casa-Trabalho	MNL	0,3309	-	-	Coeficientes Elasticidade
		Transporte Coletivo	Custos ou Tarifa		Nested Logit	0,339			Efeito de Políticas Públicas
13	Krygsman, Arentze e Timmermans (2007)	Carro	Proporção Tempo de viagem (TC/Carro) Tempos (Acesso + Saída) / Tempo viagem TC	Casa-Trabalho	MNL	0,44	-	-	Coeficientes
		Carona	Distância (total) # carros no domicílio / # trabalhadores						
		Trem	# pessoas no domicílio Gênero						
		Transporte Coletivo	Renda domiciliar Presença de criança <6anos (dummy) Forma urbana (dummy)						
14	Muller, Tsharaktschiew e Haase (2008)	A pé	Idade	Casa-Educação	MNL	0,63	81%	99%	Coeficientes
		Bicicleta	Distância (total)						
15	Bekhor e Shiftan (2009)	Transporte Coletivo	Possui automóvel (dummy)	Casa-Trabalho	Nested Logit	0,268	-	95%	Distribuição P(modos) x Distância
		Carro	Clima (Dummy: tempo bom ou ruim)						
		Carro	Tempo de viagem (total)		MNL	0,261			
		Carona	Preço do estacionamento Tempo de viagem (caminhada) # carros no domicílio Custos ou Tarifa		Cross-Nested Logit	0,268			
	Trem (Kiss and Ride, Park and Ride)	Tempo de viagem (veículo)	Cross-Nested Mixed Logit	0,269	Distribuição P(modos) x Variável				
	Ônibus	Tempo de viagem (espera)							

#	Autor / Ano	Variáveis		Motivo Viagem	Modelo	Goodness-of-Fit			Análise
		Dependente	Independente			Pseudo-R <sup>2</sup>	% Correto	Nível Sig.	
16	Hensher (2012)	Carro	Tempo de viagem (veículo) Frequência	Casa-Trabalho	Mixed MNL	0,5881			Coeficientes
		Carona	Existência de estação (Origem) Existência de estação (Destino)						
		Ônibus	Custos ou Tarifa Preço da gasolina		GMXL1 (Generalised mixed logit)	0,5889	-	-	Valor do Tempo
		VLT	Preço do estacionamento Tempo de viagem (variabilidade) Preço do pedágio Renda individual		GMXL2 (Generalised mixed logit)	0,5931			
17	Su e Zhou (2012)	Carro	Gerenciamento de estacionamento Subsídio financeiro (estímulo)	Casa-Trabalho	Nested Logit	0,502		90%	Coeficientes
		Carona	Estratégia adotadas pela Empresa (7 variáveis dummy para avaliar se a empresa adota as estratégias recomendadas)						
		Transporte Coletivo	Amenidades e características do solo no local de trabalho (variáveis dummy para avaliar se tem calçada, bicicletário)						
			Cargo (dummy) Horário de trabalho (se flexível ou não)						
18	Amoh-Gyimah e Aidoo (2013)	Transporte Individual (Carro + moto)	Distância (total)	Casa-Trabalho	Conditional Logit			90%	Elasticidade Cruzada
			Custos ou Tarifa Tempo de viagem (total)						
		Transporte Coletivo (Ônibus + taxi)	Densidade Residências (população) Densidade Trabalho (Empregos) Característica da empresa						
18	Amoh-Gyimah e Aidoo (2013)	TNM		Casa-Trabalho	Conditional Logit			90%	Efeito de Políticas Públicas
		Transporte Individual (Carro + moto)	Gênero Idade						
		Transporte Coletivo (Ônibus + taxi)	Educação (categorias) Estado Civil (dummy) Renda domiciliar # pessoas no domicílio Distância (total) Tempo de viagem (total) Tempo de viagem (veículo)						

#	Autor / Ano	Variáveis			Motivo Viagem	Modelo	Goodness-of-Fit			Análise
		Dependente	Independente				Pseudo-R <sup>2</sup>	% Correto	Nível Sig.	
19	Dave, Raykundaliya e Shah (2013)	Rickshaw	Gênero Idade Renda domiciliar Possui automóvel (dummy)	Casa-Educação	Binomial Logit	-	67%	90%	Coeficientes	
		Van	# membros do domicílio empregados Distância (caminhada) Tempo de viagem (espera) Tempo de viagem (veículo) Custos ou Tarifa							Efeito de Políticas Públicas
		Ônibus	Horário de saída (adiantado ou atrasado) Distância (total)							
20	Pulugurta, Arun e Errampalli (2013)	Carro	Tempo de viagem (veículo)	Casa-Trabalho	MNL	0,025	40%	-	Coeficientes	
		Motocicleta	Tempo de viagem (espera)						Efeito de Políticas Públicas (fuzzy)	
		Ônibus	Custos ou Tarifa							
		Rickshaw	Conforto		Fuzzy	N/A	69%			
21	Sun, Zhu e Chen (2013)	Carro	Tempo de viagem (total)	N/I	Dynamic Generalized Cost	-	-	-	-	
		Trem	Custos ou Tarifa							
		Avião	Custo de construção							
22	Santos et al. (2013)	Carro	Extensão malha ciclável # Ônibus / 1 mil habitantes Taxa de motorização % população idosa	Casa-Trabalho	MNL	0,331	-	95%	Coeficientes	
		Motocicleta	PIB per capita % residências com menores de idade							
		Transporte Coletivo	# Moto / 1 mil habitantes Preço da Gasolina População						Efeito de Políticas Públicas	
		TNM	# dias de chuva Densidade Estudante Custos ou Tarifa							
23	El-Bany et al. (2014)	Carro	Tempo de viagem (veículo) Tempo de Viagem (espera)	N/I	MNL	-	-	95%	Coeficientes	
		Táxi	Custos ou Tarifa Tempo de viagem (caminhada)						Efeito de Políticas Públicas	
		BRT	Necessita de transbordo (dummy) Renda individual							

#	Autor / Ano	Variáveis		Motivo Viagem	Modelo	Goodness-of-Fit			Análise
		Dependente	Independente			Pseudo-R <sup>2</sup>	% Correto	Nível Sig.	
24	Enaux e Gerber (2014)	Carro	Gênero Idade	Casa-Trabalho	Binomial Logit	0,480	88,5%	90%	Coeficientes
		Trem	Educação (categorias) Presença de crianças no domicílio (dummy) Renda domiciliar						
		Ônibus	Horário de trabalho (categorias) Estrutura urbana (Origem) Estrutura urbana (Destino) Oferta de estacionamento (categorias)						Efeito de Políticas Públicas
		Outro	Acessibilidade TC (categorias) Percepção Carro (categorias) Percepção Meio Ambiente						
25	Ermagun e Samimi (2014)	Carro	# pessoas no domicílio Renda domiciliar	Casa-Educação	MNL	0,33		99%	Coeficientes
		Ônibus Fretado	Possui automóvel (dummy) Idade						
		Transporte Coletivo	Gênero Distância (caminhada) Custos ou Tarifa Segurança						Efeito de Políticas Públicas
TNM	Tempo de viagem (total) Conforto	Nested Logit	0,38						
26	Habib e Weiss (2014)	Carro	Idade Gênero	Casa-Trabalho	Parametrized MNL	-	-	95%	Coeficientes
		Carona	Ocupação Profissional (categorias) # pessoas no domicílio						
		Transporte Coletivo (acesso a pé)	Local de moradia Possui automóvel (dummy) Horário de viagem						Valor do Tempo
		Transporte Coletivo (park and ride)	Origem Destino Distância (total) Tempo de viagem (total) Custos ou Tarifa						
27	Hammadou e Mahieux (2014)	Carro	Idade Gênero	N/I	MNL	0,541	83%	90%	Coeficientes
		Carona	Custos ou Tarifa Tempo de viagem (veículo) Tempo de viagem (espera)						
		Ônibus	Emprego						Efeito de Políticas Públicas
		Bicicleta	Composição familiar do domicílio Renda domiciliar						
		A pé	Frequência Uso do Solo (origem)						

#	Autor / Ano	Variáveis		Motivo Viagem	Modelo	Goodness-of-Fit			Análise		
		Dependente	Independente			Pseudo-R <sup>2</sup>	% Correto	Nível Sig.			
28	Roman et al. (2014)	Avião	Custos ou Tarifa Tempo de viagem (veículo) Tempo de viagem (espera) Frequência	N/I	MNL	-	-	-	Coeficientes		
		TAV	Custos (acesso) Renda domiciliar Idade Gênero							Mixed MNL	Elasticidade Cruzada
		Ônibus	# pessoas no domicílio Renda per capita domiciliar / tempo livre						Valor do Tempo		
		Ônibus Fretado	Gênero Estado Civil (dummy) Renda domiciliar Residência								
29	Sadri et al. (2014)	Táxi	# pessoas no domicílio	Evacuação	Nested Logit	0,236	-	95%	Coeficientes		
		Ônibus	# pessoas >65anos no domicílio								
		Carona	Destino Experiência prévia em evacuação (Dummy)								
30	Shahin, Huseyin e Kemal (2014)	Carro	Idade Gênero Renda domiciliar Possui automóvel (dummy)	Casa-Lazer	Binomial Logit	0,214	-	95%	Coeficientes		
		Transporte Coletivo	Tempo de viagem (total) Custo da viagem / Renda diária Possui ingresso da temporada (dummy)								
31	Tan Van, Choocharukul e Fujii (2014)	Carro	Aspectos simbólicos (identidade pessoal)	Casa-Trabalho	MNL	0,131	-	90%	Coeficientes		
		Transporte Coletivo	Aspectos instrumentais (vantagens) Aspectos afetivos (emoção)								
		A pé	Aspectos de ordenamento (ambiente)								
32	Flugel et al. (2015)	Bicicleta	Gênero Idade	Casa-Lazer	MNL	0,195	-	-	Coeficientes		
		Motocicleta	Possui carteira de habilitação (dummy) Possui automóvel (dummy)								
		Carro	Custos ou Tarifa							Heteroscedastic Logit	0,548
		Avião	Tempo de viagem (veículo)							Nested Logit	0,593
		Trem	Tempo de viagem (acesso)							Cross-Nested Logit	0,447
	Ônibus	Tempo de viagem (espera)									
	TAV	Frequência									

## ANEXO B

Mensagem de motivação para participar da pesquisa, distribuída pelo sistema oficial da Universidade, SIGA (CARVALHO, 2016).

**UFRJ** Sistema Integrado de Gestão Acadêmica

Prezado(a) aluno(a)

O PROGRAMA DE ENGENHARIA DE TRANSPORTES (PET) da COPPE/UFRJ em parceria com o Laboratório RESET (REDE DE ESTUDOS DE ENGENHARIA E SOCIOECONÔMICOS EM TRANSPORTES) está desenvolvendo uma pesquisa com o intuito de compreender os padrões de mobilidade da Comunidade Universitária presente no campus da Ilha do Fundão.

Os dados coletados nessa pesquisa permitirão criar análises sobre como a população da Cidade Universitária da UFRJ se desloca para o campus e futuramente nortear políticas públicas com foco na melhoria do transporte público.

Para isso precisamos da sua contribuição respondendo este questionário. Isso pode ser feito em cerca de 5 minutos.

PARA PESSOAS QUE SÃO DE UNIDADES NÃO LOCALIZADAS NA ILHA DO FUNDÃO, PEDIMOS QUE RESPONDA SOMENTE SE FOREM PARA A CIDADE UNIVERSITÁRIA AO MENOS UMA VEZ AO MÊS.  
PEDIMOS, TAMBÉM, QUE SÓ RESPONDA UMA ÚNICA VEZ!

LINK DO FORMULÁRIO:

<http://www.surveymonkey.com/s/3/2461652/MOBILIDADE-CIDADE-UNIVERSITARIA-UFRJ>

Agradecemos desde já a sua colaboração!

Sistema Integrado de Gestão Acadêmica  
Superintendência Geral de Ensino  
Universidade Federal do Rio de Janeiro

Esta mensagem foi gerada através do SIGA-UFRJ.  
Por nossa política de privacidade, não divulgamos e-mails de alunos aos usuários do sistema. Como há a necessidade de envio de comunicados, foi criado um serviço no sistema para esse fim. Assim, esta mensagem foi enviada sem que seu e-mail fosse revelado ao remetente abaixo identificado. Você pode respondê-la (dar "reply") ao remetente como um e-mail comum, porém nessa resposta seu e-mail poderá ser facilmente obtido pelo remetente.

## **ANEXO C**

Endereço do repositório público para disponibilização aberta de dados utilizados nesta pesquisa de dissertação:

[www.github.com/marcellovictorino/Pesquisa\\_Cid\\_Uni](https://www.github.com/marcellovictorino/Pesquisa_Cid_Uni)