

UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE ENGENHARIA DE SÃO CARLOS

MARINA URANO DE CARVALHO CALDAS

Estratégias para calibração de modelo Logit Multinomial por meio de  
Algoritmos de Aprendizagem de Máquinas: uma abordagem associada a grandes  
conjuntos de alternativas

São Carlos

2021



MARINA URANO DE CARVALHO CALDAS

Estratégias para calibração de modelo Logit Multinomial por meio de Algoritmos de Aprendizagem de Máquinas: uma abordagem associada a grandes conjuntos de alternativas

**Programa de Pós-graduação em  
Engenharia de Transportes da EESC-USP**  
**Exemplar definitivo (corrigido). O exemplar original está disponível na CPG da EESC-USP**

São Carlos, 14/08/2021

**RESOLUÇÃO CoPGr N ° 6018, DE 13 DE OUTUBRO DE 2011, artigo 5°**

Dissertação apresentada à Escola de Engenharia de São Carlos, como parte dos requisitos para obtenção do título de Mestre em Ciências.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Cira Souza Pitombo

São Carlos

2021

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da EESC/USP com os dados inseridos pelo(a) autor(a).

C145e Caldas, Marina Urano de Carvalho  
Estratégias para calibração de modelo Logit Multinomial por meio de Algoritmos de Aprendizagem de Máquinas: uma abordagem associada a grandes conjuntos de alternativas / Marina Urano de Carvalho Caldas; orientadora Cira Souza Pitombo. São Carlos, 2021.

Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia de Transportes e Área de Concentração em Planejamento e Operação de Sistemas de Transporte -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2021.

1. Algoritmos Genéticos. 2. Algoritmo CART. 3. Logit Multinomial. 4. Aprendizagem de Máquinas. 5. Distribuição de viagens. 6. Grandes conjuntos de alternativas. I. Título.

## FOLHA DE JULGAMENTO

Candidata: Engenheira **MARINA URANO DE CARVALHO CALDAS**.

Título da dissertação: “Estratégias para calibração de modelo *Logit Multinomial* por meio de Algoritmos de Aprendizagem de Máquinas: uma abordagem associada a grandes conjuntos de alternativas”.

Data da defesa: 01/06/2021.

### **Comissão Julgadora**

#### **Resultado**

Profa. Associada **Cira Souza Pitombo** Aprovada  
**(Orientadora)**

(Escola de Engenharia de São Carlos – EESC/USP)

Profa. Dra. **Ana Margarita Larrañaga Uriarte** Aprovada  
(Universidade Federal do Rio Grande do Sul/UFRGS)

Prof. Dr. **Luis Henrique Magalhães Costa** Aprovada  
(Universidade Estadual Vale do Acaraú/UVA)

Coordenador do Programa de Pós-Graduação em Engenharia de Transportes:

Prof. Associado **Adalberto Leandro Faxina**

Presidente da Comissão de Pós-Graduação:

Prof. Titular **Murilo Araujo Romero**



## AGRADECIMENTOS

Aos meus pais, pelo amor, apoio e investimento aplicado para o desenvolvimento de uma formação com excelência.

À minha orientadora e amiga, Cira Souza Pitombo, pela confiança depositada e pelo suporte e aprendizado fornecidos durante toda a minha jornada acadêmica.

À minha irmã e melhor amiga, Mariana Urano, pelo companheirismo, lealdade e apoio fornecidos desde minha mudança para São Carlos.

À minha Belinha, pela lealdade, amor e companheirismo.

Aos fiéis amigos da minha cidade natal (Fortaleza/CE), Luana, Marina, Orlando, Rebecca, Odara, Pedro e Picho, pelo apoio e amizade, mesmo com a longa distância.

Aos grandes amigos e colegas da cidade de São Carlos, Ana Paula, Diego, Gabrieli, Helena, Ingrid, Isadora, Luis Otávio, Luiza Azolin, Luiza Rijo, Maiara, Mariane, Marina, Renan e tantos outros aqui não mencionados, pela amizade e pela certeza de que o período distante da minha cidade natal me proporcionou maturidade e um imensurável aprendizado.

Ao grande amigo Kleber Damasceno, pelo apoio irrestrito, pelas reflexões e por todas as vezes que me acompanhou tarde da noite após as aulas.

Aos colegas e professores do Departamento de Engenharia de Transportes da Escola de Engenharia de São Carlos, por toda ajuda e conhecimento fornecidos para o desenvolvimento deste trabalho.

Aos Professores Luis Henrique Costa, Antônio Néelson e Lucas Assirati, por todas as contribuições para a execução deste trabalho.

À Escola de Engenharia de São Carlos, por todo o suporte acadêmico e infraestrutura fornecidos.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela bolsa de estudos de mestrado e pelo auxílio para realização desta pesquisa.





## RESUMO

CALDAS, M. U. C. **Estratégias para calibração de modelo Logit Multinomial por meio de Algoritmos de Aprendizagem de Máquinas: uma abordagem associada a grandes conjuntos de alternativas.** 2021. 142 p. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2021.

Os modelos de escolha discreta, em especial o Logit Multinomial, surgem como alternativa aos modelos agregados tradicionais quando se deseja suprir a necessidade de considerar características individuais dos viajantes, além de variáveis agregadas sociodemográficas e de viagens, para previsão dos destinos dos deslocamentos. Apesar disso, há importantes restrições operacionais referentes às técnicas tradicionais de calibração desses modelos que impõem obstáculos ao seu uso em situações com muitas alternativas e parâmetros a serem estimados, como no caso da escolha de destinos urbanos. Diante do uso cada vez mais frequente e das contribuições registradas na literatura frente às ferramentas tradicionalmente usadas, os algoritmos de Aprendizagem de Máquinas surgem como artifício possível para viabilizar a calibração de modelos complexos. Assim, este trabalho tem como objetivo a aplicação de algoritmos de Aprendizagem de Máquinas na calibração de modelo Logit Multinomial para estimativa de escolha de destinos urbanos. Para isso, duas aplicações foram analisadas: (1) a proposição de uma estratégia, utilizando algoritmo CART (*Classification And Regression Tree*), para redução da quantidade de parâmetros em casos com grandes conjuntos de alternativas; e (2) o uso de Algoritmos Genéticos como ferramenta de calibração do modelo Logit Multinomial. Para ambos os casos, utilizou-se os resultados da calibração tradicional para validação e comparação. O município escolhido para este estudo foi Santa Maria (RS) e os dados utilizados são oriundos de pesquisa domiciliar e do Censo 2010. Os resultados evidenciam que a aplicação da estratégia proposta com algoritmo CART mostra-se viável, uma vez que não causou prejuízo à qualidade geral do modelo. Além disso, a aplicação de Algoritmos Genéticos para calibração do modelo Logit Multinomial em casos com muitas alternativas e/ou parâmetros relacionados ao indivíduo mostrou resultados superiores à calibração tradicional, com a ressalva da necessidade de um número alto de simulações, haja vista a grande quantidade de parâmetros a serem estimados.

Palavras-chave: Algoritmos Genéticos. Algoritmo CART. Logit Multinomial. Aprendizagem de Máquinas. Distribuição de viagens. Grandes conjuntos de alternativas.



## ABSTRACT

CALDAS, M. U. C. **Strategies for calibrating the Logit Multinomial model using Machine Learning Algorithms: an approach associated with large choice sets.** 2021. 142 p. Thesis (Master) – São Carlos School of Engineering, University of São Paulo, São Carlos, 2021.

Discrete choice models, especially Logit Multinomial, have emerged as an alternative to traditional aggregate models considering the importance of including individual characteristics of commuters, as well as aggregated sociodemographic and travel variables, to predict trip destinations. Despite this, there are important operational restrictions regarding traditional calibration techniques of these models that pose obstacles in terms of using them in situations with many alternatives and parameters to be estimated, as in the case of choosing urban destinations. Considering the ever-increasing use and contributions found in the literature compared to traditionally used tools, Machine Learning algorithms have emerged as a possible solution to enable the calibration of complex models. Thus, this study aims to apply Machine Learning algorithms to the calibration of the Logit Multinomial model to estimate urban destinations choices. To do this, two applications were analyzed: (1) proposing a strategy, using the CART (Classification and Regression Tree) algorithm, to reduce the number of parameters in cases with large choice sets; and (2) using Genetic Algorithms as a calibration tool for the Logit Multinomial model. For both cases, the results of the traditional calibration were used for validation and comparison. The municipality chosen for this study was Santa Maria (RS) and the data used are from OD household survey and the 2010 Census. The results show that the application of the proposed strategy using the CART algorithm proves to be viable as it did not affect the overall quality of the model. In addition, the application of Genetic Algorithms to calibrate the Logit Multinomial model in cases with many alternatives and/or parameters related to the individual showed better results than the traditional calibration. However, a high number of simulations was needed considering the large number of parameters to be estimated.

Keywords: Genetic Algorithms. CART algorithm. Logit Multinomial. Machine Learning. Trip distribution. Large choice sets.



## LISTA DE FIGURAS

### 2. REVISÃO DA LITERATURA E CONCEITUALIZAÇÃO FERRAMENTAL

Figura 2.1: Algoritmo de divisão CART .....	35
Figura 2.2: Configuração básica de um algoritmo genético .....	36

### 3. MATERIAIS E MÉTODO

Figura 3.1: Localização e dados populacionais do distrito Sede no município de Santa Maria .....	40
Figura 3.2: Disposição dos bairros de Santa Maria (RS) .....	40
Figura 3.3: Caracterização da escolha dos destinos por bairro.....	44
Figura 3.4: Bairros e “macrobairros” da cidade de Santa Maria (RS, Brasil), caracterizadas segundo (a) densidade populacional e (b) renda familiar média. ....	46
Figura 3.5: (a) Distribuição das escolhas de destinos por “macrobairro” e (b) linhas de desejo que representam os pares origem-destino mais frequentes .....	48
Figura 3.6: Método proposto para calibração de modelo Logit para distribuição de viagens urbanas.....	50
Figura 3.7: Método proposto para estratégia proposta de redução da quantidade de parâmetros do modelo Logit Multinomial para grandes quantidades de alternativas .....	51
Figura 3.8: Procedimento metodológico proposto para redução da quantidade de parâmetros .....	52
Figura 3.9: Exemplo do procedimento adotado com Árvore de Decisão para redução da quantidade de parâmetros a serem estimados.....	53

### 4. RESULTADOS ASSOCIADOS À ESTRATÉGIA PARA REDUÇÃO DA QUANTIDADE DE PARÂMETROS A SEREM ESTIMADOS EM MODELOS DE ESCOLHA DISCRETA PARA CASOS COM GRANDES CONJUNTOS DE ALTERNATIVAS

Figura 4.1: CART para a variável independente “Motivo Estudo” e variável dependente “macrobairros”.....	62
---------------------------------------------------------------------------------------------------------	----

Figura 4.2: Variáveis associadas a parâmetros significativos (em azul) e não significativos (em cinza) das funções utilidade de cada “macrobairro”, a partir de diferentes porcentagens adotadas .....	65
Figura 4.3: Variáveis associadas a parâmetros significativos (em azul) e não significativos (em cinza) das funções utilidade de cada “macrobairro” após as calibrações sem o procedimento proposto, segundo níveis de significância de 5% e 10%. .....	67
Figura 4.4: Detalhamento das calibrações para o modelo sem o procedimento e com procedimento (porcentagens adotadas de 10%, 50% e 80%) .....	69
Figura 4.5: Quantidade de parâmetros excluídos conforme aumento da porcentagem adotada para o procedimento proposto.....	70
Figura 4.6: Valores de $\rho^2$ e $\rho^2$ ajustado para cada calibração executada .....	71
Figura 4.7: Valores de log-verossimilhança finais para cada calibração.....	72
Figura 4.8: Valores do Critério de Informação de Akaike para cada calibração.....	72

## **5. MODELO LOGIT MULTINOMIAL PARA DISTRIBUIÇÃO DE VIAGENS URBANAS: CALIBRAÇÕES CLÁSSICA E POR ALGORITMOS GENÉTICOS**

Figura 5.1: Disposição dos bairros de Santa Maria (RS).....	77
Figura 5.2: Variáveis associadas a parâmetros significativos (em azul) e não significativos (em cinza) das funções utilidade de cada bairro, a partir da porcentagem de variação adotada de 20%. .....	79
Figura 5.3: Variáveis associadas a parâmetros significativos (em azul) e não significativos (em cinza) das funções utilidade de cada bairro na calibração final.....	82
Figura 5.4: Código com a função objetivo do Algoritmo Genético.....	87
Figura 5.5: Valores da função objetivo ao longo das gerações.....	87
Figura 5.6: Histograma dos valores de função objetivo para as melhores soluções no cenário livre .....	89
Figura 5.7: Histograma das taxas de acertos para as melhores soluções no cenário livre .....	89
Figura 5.8: Gráfico de dispersão das taxas de acertos e valores de função objetivo no cenário livre .....	89
Figura 5.9: Histograma dos valores de função objetivo para as melhores soluções no cenário	

semi-restrito .....	90
Figura 5.10: Histograma das taxas de acertos para as melhores soluções no cenário semi-restrito .....	90
Figura 5.11: Gráfico de dispersão das taxas de acertos e valores de função objetivo no cenário semi-restrito .....	91
Figura 5.12: Histograma dos valores de função objetivo para as melhores soluções no cenário restrito .....	92
Figura 5.13: Histograma das taxas de acertos para as melhores soluções no cenário restrito..	92
Figura 5.14: Gráfico de dispersão das taxas de acertos e valores de função objetivo no cenário restrito .....	92
Figura 5.15: Histograma das distâncias de viagem observadas e estimadas .....	93





## LISTA DE TABELAS

### **2. REVISÃO DA LITERATURA E CONCEITUALIZAÇÃO FERRAMENTAL**

Tabela 2.1: Síntese dos problemas de pesquisa, hipóteses e contribuições..... 38

### **3. MATERIAIS E MÉTODO**

Tabela 3.1: Identificação dos bairros..... 41

Tabela 3.2: Variáveis utilizadas nos modelos ..... 47

Tabela 3.3: Caracterização da variável dependente..... 47

Tabela 3.4: Exemplo de banco de dados com variáveis relativas ao indivíduo e às alternativas  
..... 55

Tabela 3.5: Síntese dos dados e aplicativos utilizados em cada etapa metodológica..... 57

### **4. RESULTADOS ASSOCIADOS À ESTRATÉGIA PARA REDUÇÃO DA QUANTIDADE DE PARÂMETROS A SEREM ESTIMADOS EM MODELOS DE ESCOLHA DISCRETA PARA CASOS COM GRANDES CONJUNTOS DE ALTERNATIVAS**

Tabela 4.1: Taxas de acertos para cada modelo, obtidas a partir da amostra de validação..... 73

Tabela 4.2: Resultados dos testes estatísticos para cada comparação ..... 74

### **5. MODELO LOGIT MULTINOMIAL PARA DISTRIBUIÇÃO DE VIAGENS URBANAS: CALIBRAÇÕES CLÁSSICA E POR ALGORITMOS GENÉTICOS**

Tabela 5.1: Funções utilidade para cada bairro de destino..... 83

Tabela 5.2: Parâmetros considerados para a calibração por Algoritmos Genéticos..... 86

Tabela 5.3: Desempenho dos diferentes modelos quanto à distribuição de distâncias ..... 94

Tabela 5.4: Comparativo entre os sinais dos parâmetros das variáveis analisadas para cada calibração..... 96



## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>23</b>
1.1 MOTIVAÇÃO E JUSTIFICATIVA DA PESQUISA .....	23
1.2 OBJETIVOS.....	25
1.2.1 Objetivos específicos.....	25
1.3 HIPÓTESES DE PESQUISA.....	25
1.4 ORGANIZAÇÃO DO DOCUMENTO .....	26
<b>2. REVISÃO DA LITERATURA E CONCEITUALIZAÇÃO FERRAMENTAL .....</b>	<b>27</b>
2.1 MODELOS DE DISTRIBUIÇÃO DE VIAGENS .....	27
2.1.1 Modelos clássicos agregados.....	27
2.1.2 Modelos de escolha discreta.....	28
2.1.2.1 <i>Modelo Logit Multinomial</i> .....	29
2.2 ALGORITMOS DE APRENDIZAGEM DE MÁQUINAS .....	32
2.2.1 Árvore de Decisão .....	33
2.2.2 Algoritmos Genéticos.....	35
<b>3. MATERIAIS E MÉTODO .....</b>	<b>39</b>
3.1 DADOS .....	39
3.1.1 Pesquisa Domiciliar.....	43
3.1.2 Tratamento dos dados.....	44
3.2 MÉTODO.....	49
3.2.1 Método para estratégia de redução da quantidade de parâmetros .....	50
3.2.1.1 <i>Detalhamento da estratégia proposta</i> .....	51
3.2.1.2 <i>Restrições metodológicas</i> .....	55
3.3 APLICATIVOS UTILIZADOS .....	56
<b>4. RESULTADOS ASSOCIADOS À ESTRATÉGIA PARA REDUÇÃO DA QUANTIDADE DE PARÂMETROS A SEREM ESTIMADOS EM MODELOS DE</b>	

<b>ESCOLHA DISCRETA PARA CASOS COM GRANDES CONJUNTOS DE ALTERNATIVAS .....</b>	<b>59</b>
4.1 FORMULAÇÃO DAS FUNÇÕES UTILIDADE.....	59
4.2 APLICAÇÃO DO PROCEDIMENTO PROPOSTO .....	61
4.3 COMPARAÇÕES E ANÁLISES DE DESEMPENHO.....	70
<b>5. MODELO LOGIT MULTINOMIAL PARA DISTRIBUIÇÃO DE VIAGENS URBANAS: CALIBRAÇÕES CLÁSSICA E POR ALGORITMOS GENÉTICOS...77</b>	
5.1 CALIBRAÇÃO CLÁSSICA .....	77
5.1.1 Aplicação da estratégia proposta.....	78
5.1.2 Formulação das funções utilidade.....	80
5.1.3 Resultados da calibração.....	81
5.1.4 Validação .....	85
5.2 CALIBRAÇÃO POR ALGORITMOS GENÉTICOS .....	85
5.2.1 Livre.....	88
5.2.2 Semi-restrito.....	90
5.2.3 Restrito.....	91
5.3 COMPARAÇÕES ENTRE AS ABORDAGENS .....	92
<b>6. CONSIDERAÇÕES FINAIS .....</b>	<b>99</b>
6.1 SÍNTESE DO TRABALHO E HIPÓTESES DE PESQUISA .....	99
6.1 CONCLUSÕES .....	101
6.1 SUGESTÕES PARA TRABALHOS FUTUROS .....	102
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>103</b>
<b>APÊNDICE A – Calibração sem o procedimento por “macro Bairros” para nível de significância de 5%.....</b>	<b>111</b>
<b>APÊNDICE B – Calibração sem o procedimento por “macro Bairros” para nível de significância de 10%.....</b>	<b>115</b>
<b>APÊNDICE C – Detalhamento das calibrações .....</b>	<b>119</b>
<b>APÊNDICE D – Calibração clássica por bairros.....</b>	<b>123</b>

<b>APÊNDICE E – Calibração por algoritmos genéticos que gerou o menor valor para a função objetivo no cenário livre .....</b>	<b>129</b>
<b>APÊNDICE F – Calibração por algoritmos genéticos que gerou o menor valor para a função objetivo no cenário semi-restrito .....</b>	<b>135</b>
<b>APÊNDICE G – Resultado da calibração por algoritmos genéticos com o menor valor para a função objetivo no cenário restrito.....</b>	<b>139</b>
<b>ANEXO A – Formulários da Pesquisa Domiciliar .....</b>	<b>141</b>



## 1 INTRODUÇÃO

*Este capítulo introduz aspectos relevantes desta pesquisa, abordando e contextualizando o problema e a motivação que a norteiam e expondo seus objetivos gerais e específicos. Além disso, ao final deste capítulo, apresenta-se uma breve descrição da organização geral deste texto.*

### 1.1 MOTIVAÇÃO E JUSTIFICATIVA DA PESQUISA

O planejamento urbano é essencial para a otimização dos sistemas de transportes e para os estudos de viabilidade de políticas públicas relacionadas à mobilidade. Nesse contexto, o estudo da demanda atual e futura, considerando seu objetivo de compreender os deslocamentos de pessoas e mercadorias e os seus determinantes (Domecich e McFadden, 1975; Kawamoto, 2004; Ortúzar e Willumsen, 2011), constitui um importante elemento para que esse planejamento seja feito de maneira adequada.

Ao se compreender que um modelo deve refletir o que ocorre na realidade, considerando os elementos e relações relevantes existentes (Novaes, 1986), fica clara a natureza individual com que se deve analisar a demanda por transportes, ou seja, o quanto características socioeconômicas desagregadas, demográficas e de uso do solo são essenciais para a modelagem em transportes.

Assim, muitos modelos desagregados, como o modelo Logit Multinomial, o qual considera que a probabilidade de o indivíduo fazer uma escolha pode ser determinada a partir de características socioeconômicas e da atratividade de uma opção frente às outras (Ben-Akiva e Lerman, 1985; Domecich e McFadden, 1975), têm sido amplamente utilizados para diversas finalidades, como escolha de rota e de horário de partida (Ben-Akiva e Bierlaire, 2003), previsão de comportamento de pedestres (Antonini et al., 2006), além de ser amplamente empregado a problemas relativos à escolha modal (Bhat, 1995; Bhat, 1998; Ben-Akiva e Bierlaire, 1999; Koppellhuber et al., 2017; Madan e Groenhout, 1987; Ding e Zhang, 2016; Lindner et al., 2017).

Sobre o estudo da etapa de distribuição de viagens, muitas técnicas foram utilizadas ao longo dos anos. Desde modelos clássicos agregados (Casey, 1955; Schneider, 1959; Evans 1970; Wilson, 1970; Evans e Kirby, 1974; Williams, 1976) até modelos de comportamento individual desagregado (Fotheringham, 1983; Ben-Akiva e Lerman, 1985; Bhat et al., 1998; Novaes, 1986), muito se tem discutido sobre as vantagens e as restrições de cada método,

sempre visando à melhoria das análises diante da importância do entendimento da dinâmica das viagens urbanas.

Ao realizar análises desagregadas sobre a escolha de destinos, no entanto, a grande quantidade de alternativas a serem consideradas em alguns casos, como no estudo da escolha de destinos urbanos, constitui um obstáculo importante, uma vez que as técnicas tradicionais de calibração de modelos desagregados – comumente utilizadas em estudos de escolha modal – não são, no geral, aplicáveis quando a quantidade de parâmetros a serem estimados é demasiadamente grande.

Diante disso, os algoritmos de Aprendizagem de Máquinas (*Machine Learning*) surgem como soluções inovadoras às problemáticas envolvendo as técnicas de calibração tradicionais. Sobre isso, diversos trabalhos destacam o seu uso como alternativa viável para muitas áreas do conhecimento (Goldberg e Holland, 1988), e, no âmbito de planejamento de transportes, a aplicação dessas ferramentas tem aumentado gradualmente, à medida em que se observam relevantes contribuições frente às limitações já conhecidas das técnicas tradicionalmente usadas. Nesse sentido, observa-se na literatura aplicações desses algoritmos para as etapas de escolha modal (Zhong e Lingras, 2008) e de alocação de tráfego (Sung e Rakha, 2009), por exemplo. Para a distribuição de viagens, etapa de interesse desse trabalho, importantes avanços têm surgido, como o uso de lógica fuzzy e de Algoritmos Genéticos para estimativa de distribuição de viagens entre cidades industriais e pontos turísticos (Kalić e Teodorović, 2003), a proposição de modelos de análise agregada para distribuição de viagens utilizando Algoritmos Genéticos (Goel et al., 2012), ou mesmo a elaboração de modelos desagregados para previsão de destinos em viagens intermunicipais (Souza et al., 2018).

Como contribuição aos trabalhos recentes que apontaram melhorias significativas na eficiência da modelagem para diversos fins ao utilizarem algoritmos de Aprendizagem de Máquinas e diante da complexidade na calibração de modelos Logit Multinomiais para problemas com grandes conjuntos de alternativas, o presente estudo destina-se à proposição de estratégias que facilitem, ou mesmo viabilizem, a calibração de modelos Logit em casos com grandes conjuntos de alternativas, utilizando como estudo de caso um problema de escolha de destinos urbanos. Essas estratégias consistem na proposição de um procedimento metodológico que utiliza algoritmo CART (*Classification And Regression Tree*) para redução da quantidade de parâmetros a serem estimados em modelos de escolha discreta desagregados e na calibração do modelo Logit Multinomial por meio de Algoritmos Genéticos, a qual foi posteriormente comparada ao modelo obtido por meio da calibração clássica.



O capítulo seguinte fornece melhor embasamento teórico ao estudo proposto ao realizar uma revisão da literatura sobre a evolução dos modelos de distribuição de viagens, com seus avanços e restrições, e sobre as contribuições que os principais algoritmos de Aprendizagem de Máquinas podem oferecer para melhorar a modelagem de demanda por transportes. Além disso, o próximo capítulo discorre acerca dos problemas envolvendo modelos de escolha discreta com grandes conjuntos de alternativas.

## 1.2 OBJETIVOS

Esta pesquisa visa à proposição de estratégias para calibração de modelo Logit Multinomial envolvendo grandes conjuntos de alternativas a partir de algoritmos de Aprendizagem e Máquinas e de uma amostra desagregada para estimativa de escolha de destinos urbanos.

### 1.2.1 Objetivos específicos

1. Propor uma estratégia para redução de parâmetros para casos com grandes conjuntos de alternativas no modelo Logit Multinomial utilizando algoritmo CART;
2. Avaliar o desempenho dos Algoritmos Genéticos frente às técnicas tradicionais de modelagem desagregada (Logit Multinomial) para o problema de estimativa de destinos urbanos;
3. Disponibilizar uma biblioteca aberta para uso acadêmico, contendo o código utilizado para a calibração por Algoritmos Genéticos, assim como os arquivos de entrada utilizados (banco de dados e limites estabelecidos para os parâmetros).

## 1.3 HIPÓTESES DE PESQUISA

O presente estudo está baseado em três hipóteses de pesquisa principais:

1. É viável a aplicação de Algoritmos Genéticos para modelagens desagregadas de distribuição de viagens urbanas;
2. Quando comparado ao modelo obtido pela calibração tradicional do modelo Logit Multinomial, o modelo obtido pela calibração por Algoritmos Genéticos é eficaz para análises desagregadas de distribuição de viagens urbanas;

3. O algoritmo CART pode ser utilizado para simplificar a calibração de modelos Logit Multinomiais em casos com grande conjunto de alternativas e/ou muitas variáveis explicativas associadas aos indivíduos.

#### 1.4 ORGANIZAÇÃO DO DOCUMENTO

Além deste capítulo, o texto constitui-se de outras cinco partes. No capítulo a seguir, apresenta-se uma revisão da literatura, que aborda as diferentes técnicas e os problemas que as envolvem e que fornecem sustentação às lacunas nas quais esta pesquisa propõe-se a preencher. Adicionalmente, são apresentadas as definições e conceitos das ferramentas abordadas neste trabalho. No Capítulo 3, são detalhados os dados e o processo de amostragem e expostos o procedimento metodológico proposto e os aplicativos utilizados. O Capítulo 4, por sua vez, traz os resultados associados à estratégia proposta para redução da quantidade de parâmetros em modelos de escolha discreta. No Capítulo 5, apresenta-se os resultados referentes às calibrações clássica e por meio de Algoritmos Genéticos do modelo Logit Multinomial para uma abordagem desagregada de distribuição de viagens urbanas. Por fim, o Capítulo 6 discorre sobre as principais conclusões e considerações diante dos resultados obtidos com a estratégia proposta para redução da quantidade de parâmetros em modelos de escolha discreta e com a comparação entre os modelos calibrados tradicionalmente e por meio de Algoritmos Genéticos.

## 2 REVISÃO DA LITERATURA E CONCEITUALIZAÇÃO FERRAMENTAL

*Neste capítulo, será exibida a revisão da literatura associada às lacunas de pesquisa. Além disso, apresenta-se os conceitos e aplicações relacionados aos principais modelos e técnicas abordadas neste trabalho: modelo Logit Multinomial, algoritmo CART (Classification And Regression Tree) e Algoritmos Genéticos.*

### 2.1 MODELOS DE DISTRIBUIÇÃO DE VIAGENS

Dentre os modelos de previsão de demanda atual e futura, o mais difundido e aplicado é o modelo Quatro Etapas, caracterizado por uma sequência definida por: i) geração de viagens; ii) distribuição de viagens; iii) escolha modal e iv) alocação de tráfego. Nesse contexto, a distribuição de viagens, em destaque nesse trabalho, surge como a etapa em que é realizada a caracterização da demanda através da estimação da Matriz Origem-Destino para data futura.

Muito embora se apliquem tradicionalmente os modelos clássicos que consideram variáveis agregadas, como os de fator de crescimento, o gravitacional e o de oportunidades intervenientes (Novaes, 1986), muitos trabalhos destacam os efeitos positivos do emprego de modelos desagregados, como os modelos de escolha discreta, para previsão de demanda (Chow et al., 2005; Mishra et al., 2013), uma vez que possibilitam a análise de características individuais que podem influenciar na escolha dos destinos.

A seguir, apresenta-se uma contextualização histórica e conceitual acerca dos modelos clássicos agregados utilizados na etapa de distribuição de viagens, assim como suas restrições, e dos modelos desagregados de escolha discreta, com ênfase para o modelo Logit Multinomial, e as problemáticas decorrentes do seu uso em casos com grandes conjuntos de alternativas. Exibem-se, ainda, conceitos e aplicações de técnicas que empregam Inteligência Artificial, mais especificamente algoritmos de Aprendizagem de Máquinas.

#### 2.1.1 Modelos clássicos agregados

Os modelos clássicos podem ser divididos em modelos de fator de crescimento (aqui representado pelo método de Fratar), gravitacionais e de oportunidades intervenientes (Cascetta et al., 2007).

O modelo Fratar (Evans, 1970; Williams, 1976) é o mais conhecido e aplicado modelo da família de modelos baseados em fator de crescimento. Esse tipo de procedimento é

considerado de fácil compreensão e aplicação, uma vez que só precisa de uma matriz presente e de fatores de crescimento estimados. Há, no entanto, consideráveis limitações associadas ao modelo Fratar: a necessidade de pesquisas acerca das zonas de origem e de destino, a dependência da acurácia da matriz presente e o fato de desconsiderar possíveis alterações na configuração dos transportes da cidade em análise, seja pela inclusão de novos modos de transporte, seja por modificações na região (novas zonas, por exemplo) ou por intervenções no sistema de transportes (como tarifas e políticas de mobilidade) (Cascetta et al., 2007; Ortúzar e Willumsen, 2011).

Inicialmente aplicado na década de 1950 por Casey (1955) e diversas vezes aprimorado (Lowry, 1964; Wilson, 1967; Evans e Kirby, 1974; Mikkonen e Luoma, 1999), o modelo gravitacional é, provavelmente, o mais empregado para a etapa de distribuição de viagens. A sua maior contribuição frente aos modelos de fator de crescimento é que nele são considerados os efeitos de impedância, como distância e custo. Além disso, sua aplicação não requer uma matriz origem-destino inicial (Ortúzar e Willumsen, 2011).

Com relação às limitações do modelo gravitacional, a que mais se discute é sua natureza agregada, ou seja, o fato de não considerar que as escolhas dos destinos são realizadas individualmente (Wilson, 1970).

Concebido por Stouffer (1940) e adaptado por Schneider (1959), o modelo de oportunidades intervenientes baseia-se na premissa de que a probabilidade de uma viagem se destinar a determinado local é proporcional à quantidade de oportunidades oferecidas.

Apesar de possuir uma base teórica mais consistente do que o modelo gravitacional (Wilson, 1967), que faz uma analogia simples à lei gravitacional de Newton, e das muitas adaptações (Kühlkamp e Ulysséa Neto, 2003; Kühlkamp, 2003; Lemos et al., 2017), o modelo de oportunidades intervenientes é mais indicado para o estudo de viagens em que o conjunto de oportunidades seja decisivo para a escolha do destino, como em viagens com motivos de compra. Além disso, o modelo não é muito usual, provavelmente pelo seu entendimento difícil, pela pouca vantagem frente ao modelo gravitacional e pela falta de aplicativos adequados (Ortúzar e Willumsen, 2011).

### 2.1.2 Modelos de Escolha Discreta

Os modelos de escolha discreta surgiram no contexto em que se discutia que a probabilidade de escolha de um indivíduo é função de suas características individuais e

socioeconômicas, além das características das alternativas (Ben-Akiva e Lerman, 1985). Amplamente utilizados por pesquisadores de várias áreas, como econometria e análise de demanda por transportes (Bierlaire e Feterison, 2009), os modelos de escolha discreta possibilitam uma análise desagregada dos processos de decisão.

O modelo de escolha discreta mais simples e usado na prática é o Modelo Logit Multinomial (Ortúzar e Willumsen, 2011), descrito na subseção a seguir. Apesar disso, vale destacar o desenvolvimento de outros modelos nas últimas décadas, que, embora mais sofisticados, apresentam formulações complexas e mais difíceis de serem aplicadas quando comparadas ao modelo Logit Multinomial (Bierlaire, 2003). É o caso do modelo Logit hierárquico, que considera a existência de correlação entre as alternativas (Bierlaire, 2003; Ortúzar e Willumsen, 2011), e do modelo Logit misto, o qual divide a componente aleatória da função utilidade em duas partes, sendo uma caracterizada pela existência de correlação e a outra considerando a independência das alternativas (Ortúzar e Willumsen, 2011; Train, 2009). Destaca-se, ainda, a existência do modelo probit, no qual assume-se distribuição normal para a componente aleatória da função utilidade (Ortúzar e Willumsen, 2011; Train, 2009).

#### 2.1.2.1 Modelo Logit Multinomial

O modelo Logit consiste num modelo de escolha discreta baseado na probabilidade de o indivíduo fazer uma escolha a partir de características socioeconômicas e da atratividade da alternativa frente às outras (Ben-Akiva e Lerman, 1985; Domesich e McFadden, 1975). Por atratividade pode-se entender como a utilidade das diferentes alternativas disponíveis no conjunto de escolha.

A função utilidade, por sua vez, resulta de uma combinação de variáveis explicativas que a maximize. Essas variáveis são fatores observáveis, explicados a partir de características individuais e das alternativas, e subjetivos, que são aqueles de origem aleatória (Ben-Akiva e Lerman, 1985). A utilidade da alternativa  $i$  para um indivíduo  $j$ , representada por  $U_{ij}$ , é então determinada por uma parcela mensurável ( $V_{ij}$ ), composta pelos atributos relativos às alternativas e ao indivíduo, e pela parcela aleatória, representada por  $\varepsilon_{ij}$ .

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad (2.1)$$

No caso do modelo Logit Multinomial, assume-se que os resíduos das funções utilidade são identicamente e independentemente distribuídos de acordo com uma distribuição de Gumbel. De modo geral, a utilidade apresenta a configuração a seguir e os coeficientes são comumente estimados a partir da maximização da verossimilhança.

$$V_{ij} = a_i + b_i \cdot x_{1ij} + c_i \cdot x_{2ij} \quad (2.2)$$

Onde:

$V_{ij}$ : utilidade da alternativa  $i$  para o indivíduo  $j$ ;

$x_{1ij}$  e  $x_{2ij}$ : variáveis explicativas, relativas à alternativa  $i$  e ao indivíduo  $j$ ;

$a_i$ ,  $b_i$  e  $c_i$ : coeficientes a serem estimados.

Definidas as utilidades para cada opção e indivíduo, a probabilidade de uma alternativa ser escolhida é definida por:

$$P_{ij} = \frac{e^{V_{ij}}}{\sum_{k=1}^n e^{V_{kj}}} \quad (2.3)$$

Onde:

$P_{ij}$ : probabilidade da alternativa  $i$  ser escolhida pelo indivíduo  $j$ ;

$n$ : quantidade de alternativas disponíveis para o indivíduo  $j$ .

Embora apresente restrições, como a propriedade da independência das alternativas irrelevantes (IIA), segundo a qual a proporção de probabilidade de escolha de uma alternativa sobre outra não é afetada pela inclusão ou retirada de novas alternativas (Luce e Suppes, 1965) – o que representa uma limitação sempre que houver alternativas correlacionadas (Train, 2009) –, o modelo Logit Multinomial baseia-se no comportamento individual e essa característica é sua maior vantagem em relação aos modelos clássicos agregados de distribuição de viagens (Ortúzar e Willumsen, 2011).

Nesse sentido, apesar de comumente aplicado para modelagem da etapa de escolha modal, o modelo Logit tem sido cada vez mais utilizado para análises desagregadas de distribuição de viagens. Mishra et al. (2013) e Chow et al. (2005), por exemplo, utilizaram dados da cidade de Maryland (EUA) e do estado da Flórida (EUA), respectivamente, para comparar o desempenho do modelo gravitacional com modelos de escolha discreta para a etapa de distribuição de viagens. Os resultados mostraram-se melhores para o modelo de escolha

discreta e ilustraram a importância de um modelo que inclua variáveis socioeconômicas dos indivíduos em sua análise.

Há de se destacar, ainda, que os modelos de escolha discreta, a partir da interpretação dos parâmetros das funções utilidade calibradas, possibilitam uma compreensão dos fatores que influenciam na escolha dos destinos pelos indivíduos, não se limitando à obtenção da matriz origem-destino. Esse é um aspecto importante, uma vez que viabiliza a elaboração de estratégias para a solução de problemas de mobilidade urbana e para o planejamento de novos empreendimentos.

Ainda que o modelo Logit se mostre teoricamente consistente e bem-sucedido em suas aplicações, o sucesso nas previsões para pequenas quantidades de alternativas, como no caso da escolha do modo de transporte, levou à aplicação dos modelos de escolha discreta em situações com grande número de alternativas (Bordley, 2013), como ocorre em muitos casos de escolhas de destinos.

No entanto, apesar das vantagens em relação aos modelos agregados, as técnicas tradicionais de calibração dos modelos de escolha discreta, em especial o modelo Logit Multinomial, apresentam duas restrições práticas importantes à sua aplicação: (1) a coleta dos dados necessários associados aos indivíduos pode ser difícil (Bordley, 2013); (2) as ferramentas computacionais à disposição não conseguem calibrar a grande quantidade de parâmetros necessários associados a um grande conjunto de alternativas e/ou muitas variáveis explicativas (Chen et al., 2005; Keane and Wasi, 2012; Winston, 1983), ou, quando o fazem, demandam longo intervalo de tempo e configurações computacionais (como memória e processador) robustas e caras (Lemp e Kockelman, 2012). Sobre isso, Genkin et al. (2007) destacam o fato de que as calibrações tradicionais envolvem operações matemáticas com inversão de matrizes, o que gera grande demanda por recursos computacionais, limitando, portanto, a quantidade de variáveis nos modelos de previsão.

Isso posto, muitos trabalhos abordam tal problemática, e algumas estratégias foram propostas com o intuito de amenizá-la. McFadden (1978a), por exemplo, sugeriu que, nos modelos Logit, as estimações fossem realizadas a partir de um conjunto do total de alternativas, gerado aleatoriamente. Apesar de muitos pesquisadores adotarem o método proposto por McFadden (Hansen, 1987; Friedman et al., 1992; Bowman e Ben-Akiva, 2001; Keane e Wasi, 2012), Chen et al. (2005) ressaltaram que, apesar de reduzir o esforço computacional, as estimativas oriundas de conjuntos de escolhas aleatoriamente sorteadas desconsideram informações úteis, e, portanto, podem ser estatisticamente menos eficientes. Além disso,

Nerella e Bhat (2004), a partir da avaliação de desempenho (capacidade de estimar a função log-verossimilhança com precisão e capacidade de replicar as probabilidades de escolha individuais) de modelos com diferentes tamanhos de amostras de alternativas, recomendam que o conjunto de alternativas gerado aleatoriamente contenha, no mínimo, 1/8 do conjunto completo de alternativas, o que poderia ser um problema quando o conjunto original possui milhares de opções.

Outra estratégia utilizada é a utilização da amostragem por importância para gerar conjuntos de alternativas (Ben-Akiva e Lerman, 1985; Frejinger et al., 2009). Com ela, as alternativas com maior probabilidade de maximizarem a utilidade são selecionadas com mais frequência. Assim, mais informações são fornecidas ao modelo, tornando a estimativa mais eficiente (Ben-Akiva e Lerman, 1985). Lemp e Kockelman (2012), no entanto, ressaltam que nem sempre a identificação das alternativas mais prováveis pode ser feita de maneira intuitiva, e que, mesmo quando acontece, não se pode garantir que a estratégia forneça resultados eficientes. Dessa forma, até o presente momento, não foi encontrada na literatura proposição de estratégias que não sejam baseadas na redução do conjunto de alternativas.

## 2.2 ALGORITMOS DE APRENDIZAGEM DE MÁQUINAS

Algoritmos de Aprendizagem de Máquinas (*Machine Learning*) correspondem a um conjunto de métodos indutivos de reconhecimento de padrões que consideram o conjunto de dados para fazer previsões e classificações, sem a necessidade de instruções programadas (Michie et al., 1994). O sistema de aprendizado dessas técnicas pode ser de três tipos: supervisionado, não supervisionado e por reforço (Russel e Norvig, 2002).

Para o aprendizado supervisionado, são fornecidos dados de entrada e de saída, de forma que a máquina encontre uma regra geral para esses dados. No aprendizado não supervisionado, por sua vez, os dados de saída não são fornecidos, permitindo ao computador livre reconhecimento de padrões. Finalmente, no aprendizado por reforço, o programa deve exercer determinado objetivo em que o sistema indica o desempenho de uma ação aleatória (Russel e Norvig, 2002).

No contexto de demanda por transportes, muitos algoritmos de Aprendizagem de Máquinas têm se mostrado eficientes frente às técnicas tradicionais (Pulugurta et al., 2013). Segundo Xie et al. (2003), as técnicas supervisionadas oferecem várias vantagens em relação aos modelos tradicionais (como o Logit): não precisam especificar nenhuma estrutura



específica, são capazes de modelar relações mais complexas do comportamento humano e podem obter padrões de subgrupos homogêneos, e não de toda a população. Essas vantagens são também evidenciadas no trabalho de Paredes et al. (2017), no qual foram comparados modelos de Aprendizagem de Máquinas e de escolha discreta para previsão de demanda.

Os tópicos a seguir realizam uma breve discussão sobre algumas das principais técnicas de Aprendizagem de Máquinas supervisionadas utilizadas na análise de demanda por transportes, com aplicações recentes na área, sobretudo no que diz respeito à etapa de distribuição de viagens.

### 2.2.1 Árvore de Decisão

Árvore de Decisão é um modelo estatístico supervisionado para previsão e classificação de dados, para o caso de variável dependente numérica ou categórica, respectivamente. Essa técnica, não paramétrica, consiste em divisões sucessivas a partir de regras de indução definidas para formação de conjuntos homogêneos de dados, segundo variável dependente (Breiman et al., 1984). As condições de divisão são representadas por nós e a estrutura hierárquica final é denominada árvore. Os principais algoritmos de Árvores de Decisão são CHAID - *Chi-Square Automatic Interaction Detection* (Kass, 1980), CART (Breiman et al., 1984) e C4.5 (Quinlan, 1983).

Esse método tem sido amplamente aplicado na área de planejamento de transportes. Pitombo et al. (2011) utilizaram o algoritmo de Árvore de Decisão para analisar as relações entre variáveis socioeconômicas, de uso do solo, participação em atividades e padrões de viagens encadeadas. Xie et al. (2003) aplicaram algoritmos de Árvore de Decisão, juntamente com Redes Neurais Artificiais, para estimar a etapa de escolha modal e os compararam ao modelo Logit. Os autores constataram a melhor performance das técnicas de Aprendizagem de Máquinas utilizadas em relação ao modelo tradicional.

No contexto de distribuição de viagens, foco deste trabalho, Pitombo e Guimarães (2016) e Pitombo et al. (2017) empregaram a técnica de Árvore de Decisão para a modelagem desta etapa, para o caso de viagens intermunicipais. Como resultado, eles observaram que, em comparação ao modelo tradicional gravitacional, a Árvore de Decisão mostrou-se mais eficiente. Ainda, os autores destacaram as vantagens do uso da Árvore de Decisão, como sua fácil aplicabilidade, ausência de suposições matemáticas rígidas e a falta de restrições aos tipos de variáveis de entrada.

Ainda que demonstre bom desempenho frente aos modelos convencionais, esse método apresenta restrições ao seu uso: seu caráter não paramétrico, que impede que seja testada a significância dos parâmetros estimados, e o fato de avaliar uma variável por vez (Shmueli et al., 1996).

- Algoritmo CART

Este estudo aplicou o algoritmo CART (*Classification And Regression Tree*), o qual foi desenvolvido por Breiman et al. (1984) e é baseado na realização de uma sequência de divisões binárias do conjunto de dados inicial (nó raiz) até que a máxima homogeneidade, segundo valores de variável dependente, seja alcançada dentro dos nós filhos terminais. Assim, a partição do banco de dados é efetuada de forma a minimizar a impureza dos nós filhos. Essa impureza é representada pelo índice Gini (Equação 2.4) para o caso de árvore de classificação.

$$G(t) = 1 - \sum_{i=1}^n p^2 \left( \frac{i}{t} \right) \quad (2.4)$$

Onde:

n: quantidade de categorias da variável dependente;

p(i/t): probabilidade da categoria i se formar no nó t.

Para determinar o melhor critério de divisão (melhor variável explicativa e valor de corte), calcula-se, a partir dos valores do índice Gini para cada nó, o valor do aprimoramento para cada possível divisão. Para uma divisão ser considerada a mais adequada, seu valor de aprimoramento deve ser o maior dentre os de todas as divisões possíveis e é calculado a partir da Equação (2.5).

$$\text{Aprimoramento} = G(t_{pai}) - p_{Esq} \cdot G(t_{Esq}) - p_{Dir} \cdot G(t_{Dir}) \quad (2.5)$$

Onde:

p<sub>Esq</sub>: proporção de casos da categoria i da variável dependente no nó filho da esquerda;

p<sub>Dir</sub>: proporção de casos da categoria i da variável dependente no nó filho da direita;

t<sub>Esq</sub>: nó filho da esquerda;

t<sub>Dir</sub>: nó filho da direita.

A Figura 2.1 esquematiza a divisão binária realizada pelo algoritmo CART, em que  $t_{Pai}$  é denominado nó pai,  $x_j$  é uma variável explicativa  $j$ ,  $x_j^R$  é o melhor valor de divisão da variável  $x_j$ ,  $P$  é a probabilidade de ocorrência de cada categoria de variável dependente (para problemas de classificação) em cada nó, e  $t_{Esq}$  e  $t_{Dir}$  são os dois nós filhos gerados na primeira divisão binária.

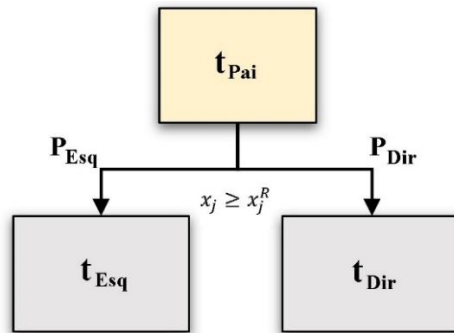


Figura 2.1: Algoritmo de divisão CART (adaptado de Breiman et al., 1984).

### 2.2.2 Algoritmos Genéticos

Introduzido por John Holland (1975), esse método considera as premissas biológicas básicas da genética, evolução e seleção natural ao fazer uma analogia com o fato de que os indivíduos herdam características dos seus genitores, que seus genes podem sofrer mutações que poderão ser herdadas pelas próximas gerações e que os melhores indivíduos geneticamente serão os que possuem mais chances de sobreviver e gerar descendentes (seleção natural). Nesse sentido, os modelos de Algoritmos Genéticos focam na busca da melhor solução frente a um conjunto de soluções possíveis, a partir de objetivos bem delimitados (função *fitness* ou função objetivo).

A Figura 2.2 ilustra o funcionamento básico de um Algoritmo Genético. A população inicial é composta por um conjunto de indivíduos gerados aleatoriamente e que correspondem a possíveis soluções para o problema analisado. Cada iteração do algoritmo é chamada de geração e é constituída de quatro etapas fundamentais: cálculo das aptidões, seleção, cruzamento e mutação (Goldberg, 1989).

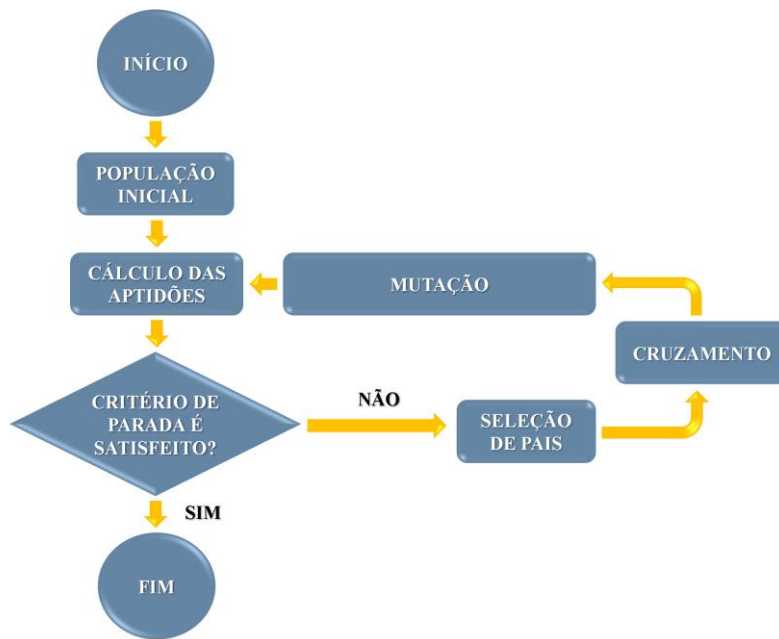


Figura 2.2: Configuração básica de um algoritmo genético. Adaptado de (Negnevitsky, 2005).

O cálculo das aptidões é realizado a partir de uma função objetivo, que otimiza uma medida que varia conforme especificação do problema (Mitchell, 1997). Os valores das aptidões são usados para ordenar os indivíduos mais aptos (melhores soluções) de uma população.

Na etapa de seleção, as melhores soluções são selecionadas e utilizadas para gerar uma nova população. Essa seleção é realizada de modo que as soluções com as maiores aptidões têm maiores probabilidades de serem selecionados e de serem usados nos cruzamentos para gerar a população da nova geração (Goldberg, 1989).

Na etapa de cruzamento (*crossover*), por sua vez, os indivíduos – ou soluções – selecionados são embaralhados e particionados de modo que cada novo indivíduo possua características dos indivíduos selecionados que sofreram cruzamento (Goldberg, 1989). Essa etapa é muito importante, pois garante que, ao longo das iterações, melhores soluções sejam geradas.

Por fim, a etapa de mutação proporciona maior diversidade nas soluções e impede que o algoritmo convirja rapidamente a mínimos locais ao promover alterações em um ou mais indivíduos dentro de uma população (Carvalho et al., 1999). O parâmetro para o processo de mutação é determinado em código e determina a probabilidade de um indivíduo ter um de seus genes alterado aleatoriamente. Essas alterações podem ser benéficas ou prejudiciais ao processo de seleção, de forma que, se benéficas, gerarão indivíduos com maiores valores de aptidão, enquanto, se prejudiciais, promoverão redução nesses valores.

Vale ressaltar, ainda, uma estratégia nomeada *elitismo*, na qual o melhor indivíduo de uma geração é replicado na geração seguinte, de modo que a melhor solução encontrada até o momento (maior valor de aptidão) não se perca devido ao caráter estocástico do método (Michie et al., 1994).

Em comparação aos outros métodos, os Algoritmos Genéticos destacam-se por: trabalhar com a estimação dos parâmetros conjuntamente, e não individualmente; utilizar uma função objetivo bem delimitada, não derivada de um conhecimento auxiliar; lidar bem com grandes bancos de dados, poder ser aplicado a qualquer tipo de otimização, possuir flexibilidade quanto à quantidade de parâmetros e ser um método probabilístico, sem regras determinísticas (Goldberg, 1989).

No que diz respeito à sua aplicação, Hu e Zhang (2008) empregaram dados agregados da cidade de Xian para prever a distribuição espacial em áreas urbanas utilizando Algoritmos Genéticos. Cunha et al. (2009), por sua vez, usaram Algoritmos Genéticos para calibrar modelos de performance veicular usados por simuladores de tráfego. Rocha (2019) e Rocha et al. (2021) utilizaram Algoritmos Genéticos para otimização da modelagem espacial de demanda por transportes.

Para o caso específico da etapa de distribuição de viagens, é crescente a quantidade de trabalhos que fazem uso de Algoritmos Genéticos. Goel et al. (2012) obtiveram resultados satisfatórios ao utilizar essa técnica para propor um modelo de análise agregada de distribuição de viagens na área urbana de Deli (Índia). Kalić e Teodorović (2003) aplicaram algoritmos de Aprendizagem de Máquinas (dentre eles Algoritmos Genéticos) para estimar a distribuição de viagens entre cidades industriais e pontos turísticos de uma determinada região, obtendo também bons resultados. Sung e Rakha (2009), por sua vez, compararam o modelo gravitacional com o obtido por algoritmos de Aprendizagem de Máquinas para a etapa de distribuição de viagens de uma situação hipotética e concluíram que este último teve melhor desempenho.

Souza et al. (2018) empregaram algoritmos de Aprendizagem de Máquinas (Algoritmos Genéticos e Árvore de Decisão) para modelar a etapa de distribuição de viagens intermunicipais considerando dados desagregados de 11 cidades do estado da Bahia (Brasil). A aplicação desses métodos ocorreu pela tentativa de investigar as vantagens de utilização de algoritmos de Aprendizagem de Máquinas frente ao modelo gravitacional. Como resultado, os autores constataram a eficácia desse tipo de método para prever distribuição de viagens intermunicipais, mas não o compararam, assim como os autores acima mencionados, ao método tradicional de

calibração do modelo Logit, nem realizaram a pesquisa utilizando dados de viagens urbanas, como propõe o presente trabalho.

Destaca-se, ainda, o trabalho de Zhong e Lingras (2008). Os autores empregaram, assim como este trabalho, Algoritmos Genéticos para calibrar os parâmetros do modelo Logit Multinomial. Esse trabalho, porém, utiliza o método para a etapa de escolha modal, diferentemente da presente pesquisa, que foca na aplicação de Algoritmos Genéticos para a etapa de distribuição de viagens.

Em suma, destacam-se as seguintes lacunas de pesquisa, verificadas por meio da revisão da literatura, relativas a este trabalho:

- Ausência de estudos, até o presente momento, que utilizem Algoritmos Genéticos em análises desagregadas de distribuição de viagens urbanas (Lacuna 1);
- Ausência de estudos que comparem o modelo obtido por Algoritmos Genéticos com o modelo Logit Multinomial tradicional para análises desagregadas de distribuição de viagens (Lacuna 2);
- As ferramentas computacionais à disposição não conseguem calibrar modelos Logit Multinomiais tradicionais em casos com grande conjunto de alternativas e/ou muitas variáveis explicativas associadas ao indivíduo (Lacuna 3).

A Tabela 2.1, a seguir, traz uma síntese dos problemas de pesquisa, associados às contribuições e hipóteses apresentadas neste trabalho.

Tabela 2.1: Síntese dos problemas de pesquisa, hipóteses e contribuições.

Problema/Lacuna de Pesquisa	Hipótese	Contribuição
(1) Ausência de estudos que utilizem Algoritmos Genéticos para análises desagregadas de distribuição de viagens urbanas;	É viável a aplicação de Algoritmos Genéticos para modelagens desagregadas de distribuição de viagens urbanas;	Uso inédito, até o presente momento, de Algoritmos Genéticos para uma análise desagregada de distribuição de viagens urbanas;
(2) Ausência de estudos que comparem o modelo Logit Multinomial tradicional com o modelo obtido por Algoritmos Genéticos para análises desagregadas de distribuição de viagens urbanas;	Quando comparado ao modelo obtido pela calibração tradicional do modelo Logit Multinomial, o modelo obtido pela calibração por Algoritmos Genéticos é eficaz para análises desagregadas de distribuição de viagens urbanas;	Validação do uso de Algoritmos Genéticos para problemas de distribuição de viagens urbanas por meio da calibração do modelo Logit Multinomial tradicional;
(3) Dificuldade operacional em calibrar modelos Logit Multinomiais tradicionais em casos com grande conjunto de alternativas e/ou muitas variáveis explicativas.	O algoritmo CART pode ser utilizado para simplificar a calibração de modelos Logit Multinomiais em casos com grande conjunto de alternativas e/ou muitas variáveis explicativas.	Elaboração de uma estratégia, utilizando o algoritmo CART, para redução da quantidade de parâmetros a serem estimados em modelos Logit Multinomiais para casos com grande conjunto de alternativas.

### 3 MATERIAIS E MÉTODO

*Neste capítulo, detalham-se os dados empregados nesta pesquisa - obtidos por meio de pesquisa domiciliar -, seu processo de amostragem, o tratamento realizado e os aplicativos utilizados. Além disso, apresenta-se o procedimento metodológico com as etapas executadas para realização deste estudo.*

#### 3.1 DADOS

Os dados utilizados são referentes ao município de Santa Maria (RS), mais especificamente ao distrito urbano (distrito Sede), no qual residem e trabalham 95% da população do município. Esses dados foram retirados do Plano Diretor de Mobilidade de Santa Maria (IPLAN, 2013). Os dados sobre os destinos possíveis, por sua vez, são do inventário urbano, levantado também para a elaboração do Plano de Mobilidade Urbana. Alguns dos dados agregados sociodemográficos das origens têm como fonte o Censo do IBGE (Instituto Brasileiro de Geografia e Estatística) (IBGE, 2010). Vale ressaltar que os destinos urbanos se referem ao dia inteiro anterior à entrevista. A não seleção de uma faixa horária específica ocorreu para evitar a obtenção de uma matriz O/D com muitos valores nulos.

De acordo com dados do Censo do IBGE de 2010, o distrito em questão concentrava, naquele ano, 246.465 habitantes e era formado por 41 bairros. A Figura 3.1 ilustra a distribuição populacional do município de Santa Maria e a localização do distrito Sede.

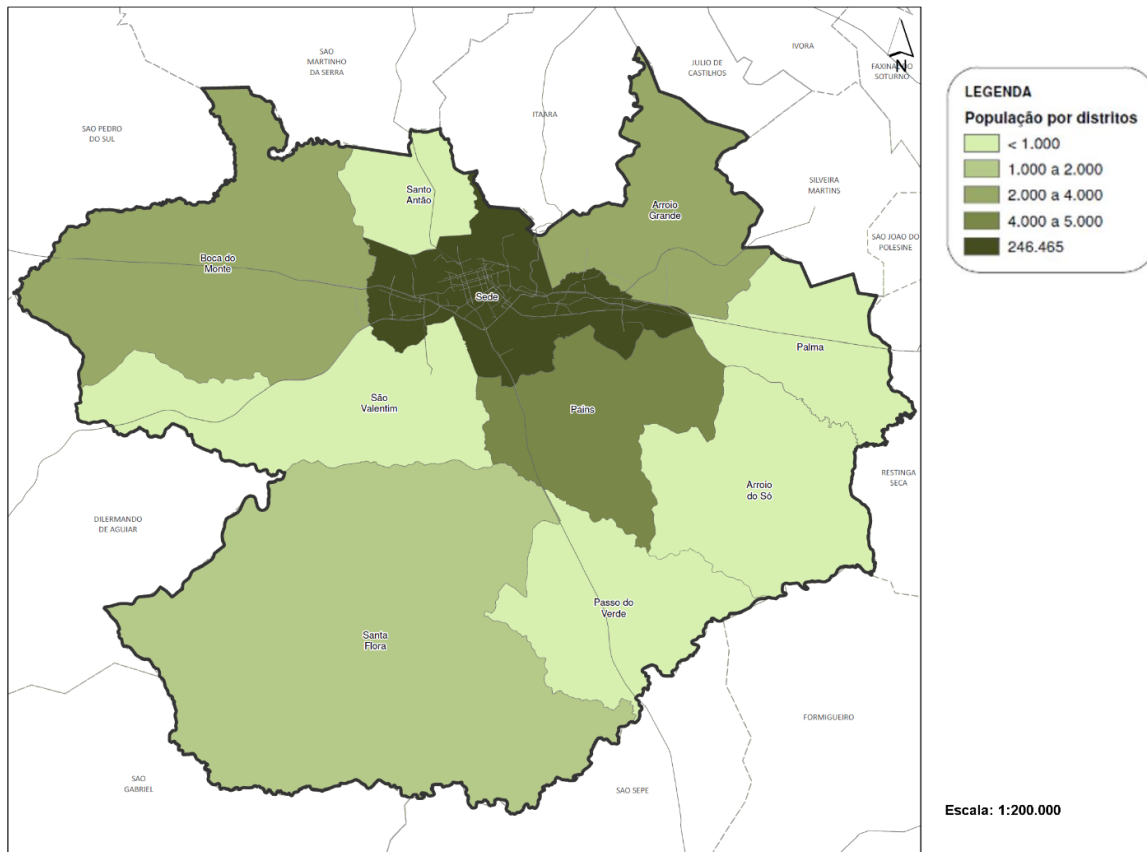


Figura 3.1: Localização do distrito Sede e distribuição populacional no município de Santa Maria.

Fonte: IPLAN, 2013.

A estrutura da cidade é influenciada por características físicas locais, como a linha ferroviária que divide o centro da cidade. Essa barreira urbana dificulta a acessibilidade em alguns bairros, principalmente aos pedestres, e isola bairros periféricos, pois existem grandes distâncias entre alguns dos 11 pontos de cruzamento existentes. A Figura 3.2 ilustra a disposição dos 41 bairros, bem como a localização do município de Santa Maria no estado do Rio Grande do Sul.

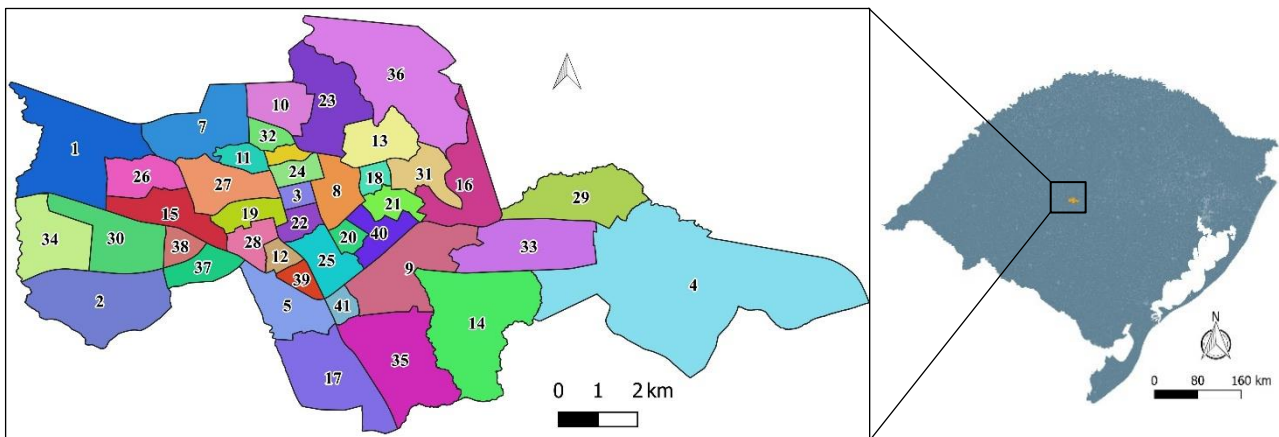


Figura 3.2. Disposição dos bairros de Santa Maria (RS).



A Tabela 3.1 a seguir identifica os bairros representados na Figura 3.2.

Tabela 3.1: Identificação dos bairros.

Bairro	Nome	Bairro	Nome
1	Agro-Industrial	22	Nossa Senhora de Fátima
2	Boi Morto	23	Nossa Senhora do Perpétuo Socorro
3	Bonfim	24	Nossa Senhora do Rosário
4	Camobi	25	Nossa Senhora Medianeira
5	Urlândia	26	Nova Santa Marta
6	Carolina	27	Passo D'Areia
7	Caturrita	28	Patronato
8	Centro	29	Pé-de-Platano
9	Cerrito	30	Pinheiro Machado
10	Chácara das Flores	31	Presidente João Goulart
11	Divina Providência	32	Salgado Filho
12	Duque de Caxias	33	São José
13	Itararé	34	Tancredo Neves
14	João Luiz Pozzobon	35	Tomazetti
15	Juscelino Kubitschek	36	Campestre do Menino Deus
16	Km 3	37	Renascença
17	Lorenzi	38	São João
18	Menino Jesus	39	Uglione
19	Noal	40	Nossa Senhora de Lourdes
20	Nonoai	41	Dom Antônio Reis
21	Nossa Senhora das Dores		

De acordo com a pesquisa do IBGE em 2010, a cidade em estudo tem como característica uma população considerada jovem, na qual predominam residentes com idades entre 15 e 29 anos. Os homens são maioria entre 01 e 14 anos de idade, mas, na população a partir de 15 anos, o número de mulheres ultrapassa o número de homens. O crescimento da população urbana de Santa Maria em relação à rural ocorreu a partir de 1970 e a evolução da população de 2002 a 2012 se manteve constante, com incremento anual de 6%. O crescimento mais relevante ocorreu nos bairros mais distantes do centro, como o Nova Santa Marta, Pinheiro Machado e Camobi (identificados por 26, 30 e 4 na Figura 3.2, respectivamente). A maior concentração de moradores foi identificada nos bairros Camobi, Centro, Juscelino Kubitschek e Nova Santa Marta (identificados por 4, 8, 15 e 26 na Figura 3.2, respectivamente) que juntos representam cerca de 30% da população.

Em relação ao perfil da renda dos habitantes, apenas 49% dos indivíduos possuem renda acima do salário mínimo mensal. Entretanto Santa Maria é considerada a 28ª cidade com maior número de moradores pertencentes à classe A (com renda familiar igual ou superior a vinte salários mínimos) e é uma das três cidades do estado com maior potencial de consumo. A maior parte da riqueza (representado pela soma de salários mensais no domicílio) está concentrada nos bairros do Centro, Nossa Senhora de Fátima e Camobi (identificados por 8, 22 e 4 na Figura 3.2, respectivamente).

A cidade é considerada um polo atrativo regional, pois é referência em serviços para 36 municípios da região central do estado e influencia mais de um milhão de pessoas (IPLAN, 2013). Com relação à economia da cidade, destaca-se o setor de serviços, o qual representa cerca de 81% dos ganhos gerados. Além do comércio forte e variado, a cidade oferece ao público quatro *shoppings* e sedia os principais órgãos federais e estaduais.

Santa Maria também se destaca por ter o *status* de cidade da educação. A cidade possui mais de sete instituições de ensino superior e mais de 35 mil estudantes, sendo considerada a terceira cidade do Brasil com maior número de mestres e doutores per capita. Aproximadamente 27% da população possui ensino superior incompleto e 13% ensino superior completo. Esses dados apontam para um alto nível formativo superior da população, uma vez que apenas 8,3% da população nacional possui ensino superior completo. As instituições de ensino representam grandes polos geradores de tráfego, pois somadas (fundamental, médio e superior) totalizam 95.784 alunos e 5.611 professores (27% dos alunos e 30,7% dos professores pertencem à Universidade Federal de Santa Maria - UFSM).

Os principais polos atrativos de tráfego na cidade são os equipamentos de saúde, educativos, industriais e de lazer. Os equipamentos de saúde contam com 41 unidades básicas, 5 unidades de emergência e 11 de pronto atendimento, que juntos somam uma capacidade de 1.200 leitos hospitalares e 1.100 profissionais médicos atuando na cidade, sendo os bairros Centro e Camobi (identificados por 8 e 4 na Figura 3.2, respectivamente) os que mais concentram essas unidades. Os equipamentos de educação incluem 39 escolas estaduais, 80 escolas municipais, 13 colégios privados e 2 colégios militares. Com relação às universidades, os maiores polos são a Universidade Federal de Santa Maria (UFSM), localizada no bairro Camobi e na qual acessam diariamente mais de 30.000 pessoas, e a Universidade Franciscana, localizada no bairro Centro. Os equipamentos industriais são representados por um total de 572 indústrias, que, no ano de 2010, contavam com 6.234 empregos formais. Os principais locais de lazer estão localizados no bairro Nossa Senhora das Dores (identificado por 21 na Figura 3.2), com os maiores *shopping* e clube da cidade. Há, ainda, o Parque Itaimbé, localizado no

Centro, e a UFSM, no bairro Camobi, que abriga exposições, centros de convivência e apresentações.

O levantamento de dados locais foi desenvolvido como parte do desenvolvimento do Plano Diretor de Mobilidade Urbana e seus dados foram utilizados como base para o desenvolvimento e compreensão deste estudo.

### 3.1.1 Pesquisa Domiciliar

As pesquisas domiciliares são uma ferramenta auxiliar e essencial para o planejamento urbano e de transportes. Os dados obtidos por meio delas são referentes aos domicílios, aos moradores e aos deslocamentos realizados.

No caso da pesquisa domiciliar utilizada neste estudo, realizada para a elaboração do Plano Diretor de Mobilidade Urbana (IPLAN,2013), para obtenção da amostra de viagens representativas para o município de Santa Maria, considerou-se nível de confiança de 95%, erro amostral máximo de 10% e valor  $p$  (proporção de viagens com um destino determinado) de 0,5 (cenário mais conservador). Para determinar o tamanho amostral, a pesquisa utilizou a Equação (3.1) a seguir (Ortúzar e Willumsen, 2011). Ao final das entrevistas, chegou-se ao total de 3.758 indivíduos, valor superior ao estimado para o tamanho mínimo da amostra e, portanto, válido diante das considerações realizadas.

$$n \geq \frac{p(1-p)}{\left(\frac{e}{Z}\right)^2 + \frac{p(1-p)}{N}} \quad (3.1)$$

Onde:

$n$ : tamanho amostral;

$p$ : proporção de viagens com um destino determinado;

$e$ : nível de erro aceitável;

$N$ : tamanho da população;

$Z$ : valor padronizado para o nível de confiança estabelecido.

O formulário utilizado para a pesquisa OD de Santa Maria encontra-se no Anexo A.

## 3.1.2 Tratamento dos dados

Antes de utilizar os dados oriundos da pesquisa domiciliar, com 3.758 viagens desagregadas por indivíduos, foi necessário realizar um tratamento prévio dos dados, com o intuito de reparar possíveis inconsistências, como valores discrepantes ou não condizentes com a realidade, e dados faltantes, ou seja, registros incompletos.

Primeiramente, retirou-se os registros cujos dados mostraram-se não condizentes com a realidade ou que continham dados faltantes para muitas variáveis. Em seguida, realizou-se uma análise de dados faltantes para as variáveis “auto” e “moto”. Do tipo quantitativa, essas variáveis registram a contagem de automóveis e motocicletas nos domicílios, respectivamente, e não apresentaram respostas para uma quantidade significativa de registros (cerca de 11,8% e 25%, respectivamente). Assim, como a retirada dos registros com dados de “auto” ou “moto” faltantes significaria uma alta redução do banco de dados, aplicou-se uma técnica para reposição dos dados faltantes, disponível no *software* IBM SPSS 24. O método de reposição aplicado utiliza regressão linear múltipla, e os termos do erro são escolhidos aleatoriamente a partir dos resíduos de registros cujos dados estão completos (IBM, 2016). Ao final do tratamento, o banco de dados continha 3.136 observações, sendo que os valores para as variáveis foram normalizados pelo método min-max, o qual redimensiona linearmente cada variável para o intervalo [0, 1].

A Figura 3.3 exibe a taxa de escolha de cada bairro. Nota-se que os dados não possuem viagens que se destinem aos bairros 36 (Campestre do Menino Deus), 37 (Renascença), 38 (São João), 39 (Uglione), 40 (Nossa Senhora de Lourdes) e 41 (Dom Antônio Reis).

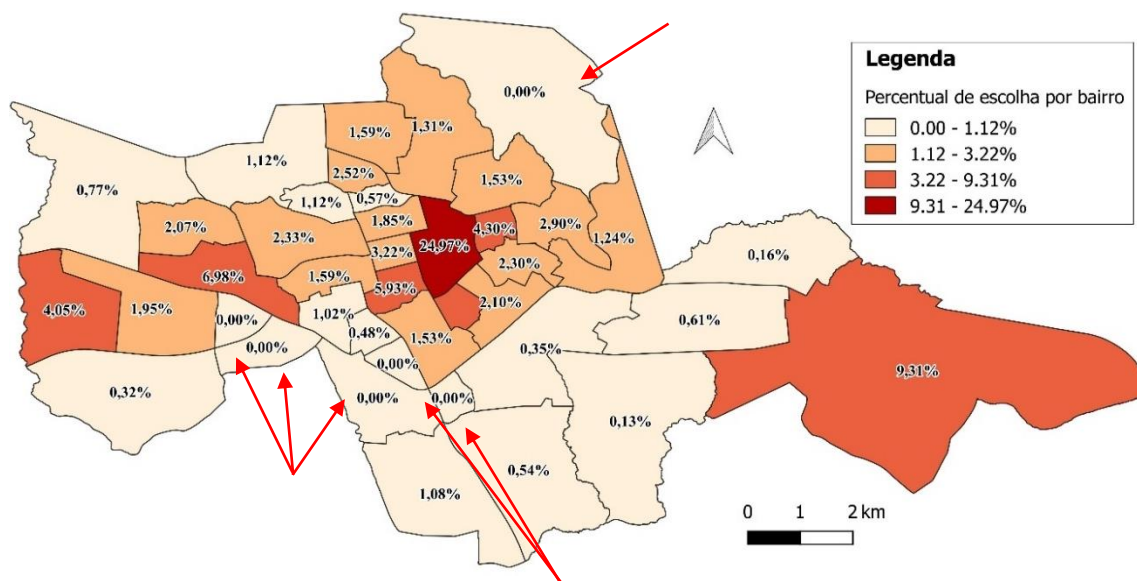
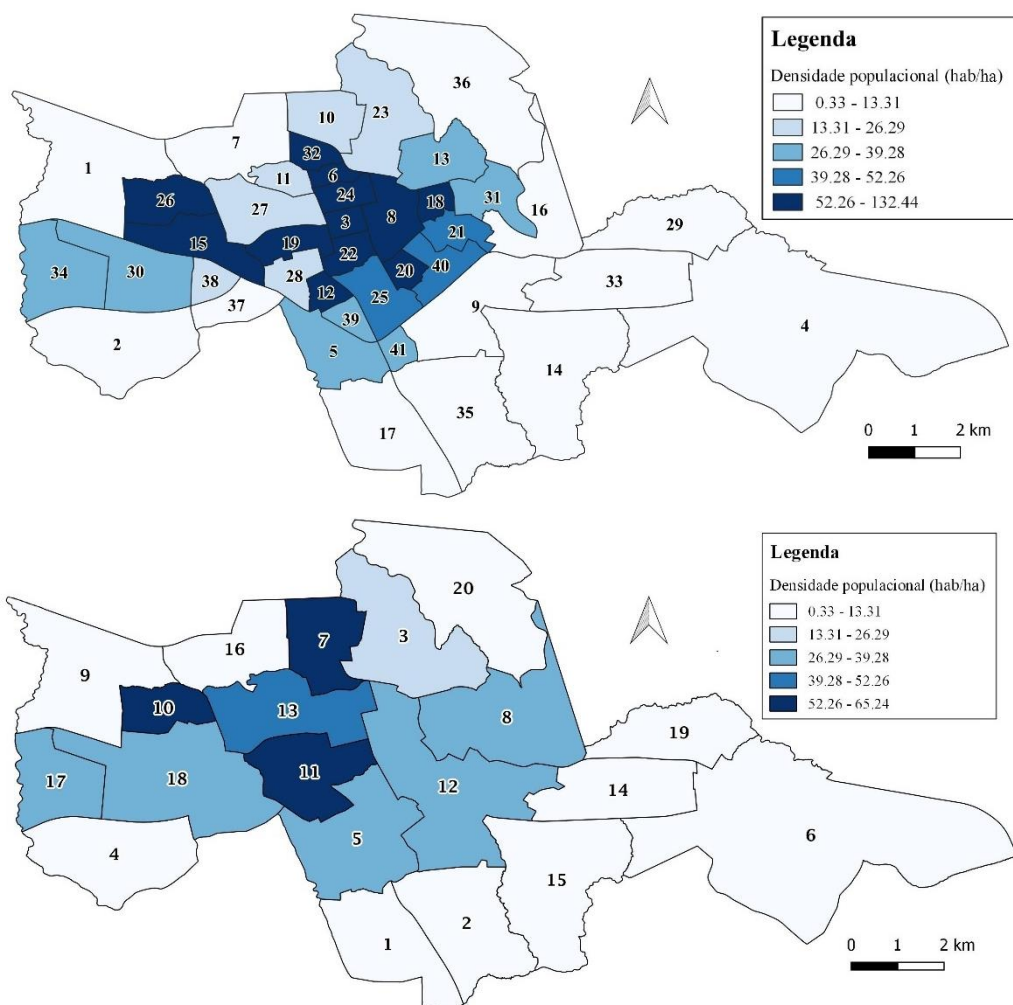
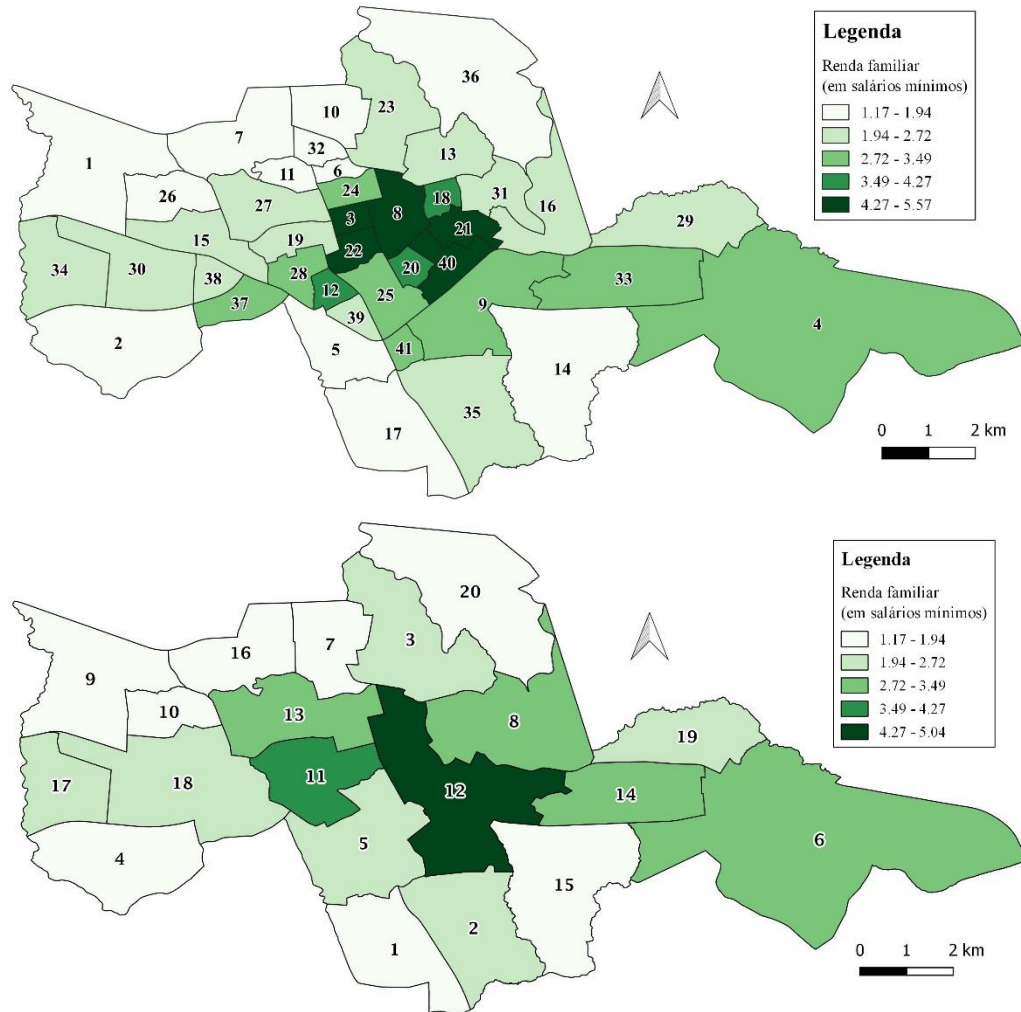


Figura 3.3: Caracterização da escolha dos destinos por bairro.

Para a formulação do procedimento para redução na quantidade de parâmetros a serem estimados em modelos com grande conjunto de alternativas, agregou-se os 41 bairros em 20 “macrobairros”, de modo que cada um deles representasse um destino possível. O objetivo desta agregação foi reduzir o número de alternativas de forma a viabilizar operacionalmente a calibração. Essa agregação foi realizada a partir de uma análise de *Cluster k*-médias, na qual se considerou as coordenadas geográficas dos centroides dos bairros como critério para os agrupamentos. A Figura 3.4 (a, b) ilustra o *layout* dos 41 bairros e dos 20 macrobairros considerados, trazendo informações relativas à densidade demográfica e à renda domiciliar média, respectivamente. Nela, é possível perceber que a classificação das unidades de área segundo classes de densidade populacional e de renda apresentam configuração similar entre os mapas de bairros e de “macrobairros”, com valores, no geral, menores ao longo da periferia (cores claras), e que aumentam conforme se aproximam da região central (cores escuras).



(a) Caracterização das unidades de área (bairros e “macrobairros”) segundo densidade populacional.



(a) Caracterização das unidades de área (bairros e “macrobairros”) segundo renda domiciliar média.  
 Figura 3.4: Bairros e “macrobairros” da cidade de Santa Maria (RS, Brasil), caracterizadas segundo (a) densidade populacional e (b) renda familiar média.

Além disso, incluiu-se ao banco de dados da pesquisa domiciliar informações acerca dos bairros e dos “macrobairros” de origem e das distâncias de viagens, que correspondem às distâncias euclidianas entre os centroides dos bairros e dos “macrobairros”. No caso de viagens dentro de uma mesma região, para impedir erros sistemáticos por distâncias nulas, considerou-se o quarto da distância entre o centroide da unidade de área em questão e o centroide da área vizinha mais próxima (Thomas e Hugget, 1980). Os leitores interessados em utilizar métodos mais recentes para calcular distâncias intrazonais de viagens podem consultar o estudo realizado por Plaza e Rodrigues da Silva (2015).

As Tabelas 3.2 e 3.3 caracterizam, respectivamente, as variáveis explicativas utilizadas nos modelos e a distribuição dos destinos de viagem segundo a variável dependente “macrobairro”, a partir da amostra final de 3.136 registros, obtida após tratamento dos dados.

Tabela 3.2: Variáveis utilizadas nos modelos.

Variável	Descrição
(A) Frequência Escolar	(0) Não; (1) Sim
(B) Gênero	(1) Masculino; (2) Feminino
(C) Grau de Instrução	(1) Analfabeto; (2) Alfabetização; (3) Fundamental I incompleto; (4) Fundamental I completo; (5) Fundamental II incompleto; (6) Fundamental II completo; (7) Ensino médio incompleto; (8) Ensino médio completo; (9) Superior incompleto; (10) Superior Completo
(D) Idade	(1) Até 17 anos; (2) 18 a 28 anos; (3) 29 a 39 anos; (4) 40 a 51 anos; (5) 52 a 65 anos; (6) Acima de 66 anos
(E) Moradores	Quantidade de moradores na residência
(F) Motivo Estudo	(0) Não; (1) Sim
(G) Motivo Integração	(0) Não; (1) Sim
(H) Motivo Lazer	(0) Não; (1) Sim
(I) Motivo Residência	(0) Não; (1) Sim
(J) Motivo Saúde	(0) Não; (1) Sim
(K) Motivo Trabalho	(0) Não; (1) Sim
(L) Renda	(1) Até 1 salário mínimo*; (2) 1 a 2 salários mínimos*; (3) 2 a 5 salários mínimos*; (4) 5 a 10 salários mínimos*; (5) Mais de 10 salários mínimos*
(M) Trabalha	(0) Não; (1) Sim
(N) Veículos Motorizados	Quantidade de veículos motorizados na residência
(O) População Origem	População residente na origem
(P) Distância	Distância euclidiana da origem aos possíveis destinos
(Q) Paradas de Ônibus	Quantidade de paradas de ônibus nos possíveis destinos

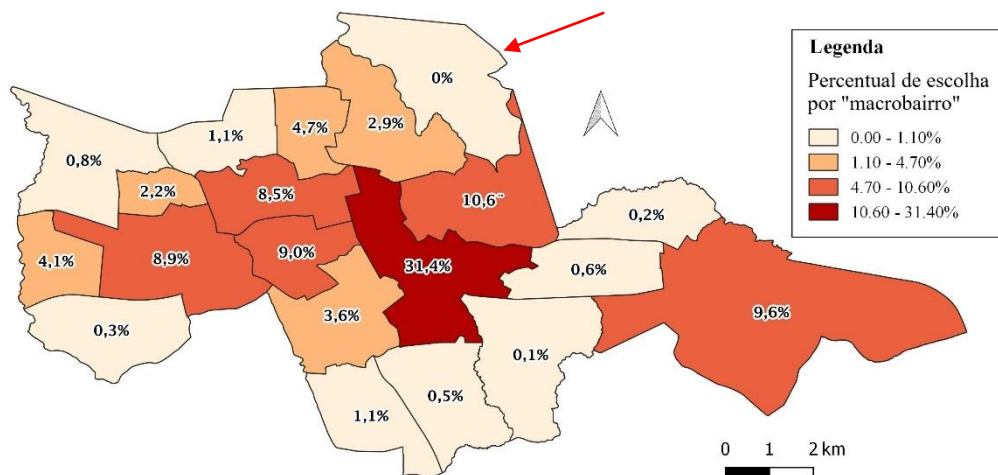
\* Valor de referência do salário mínimo em 2013: R\$678,00. Valor em 2020: R\$1045,00.

Tabela 3.3: Caracterização da variável dependente.

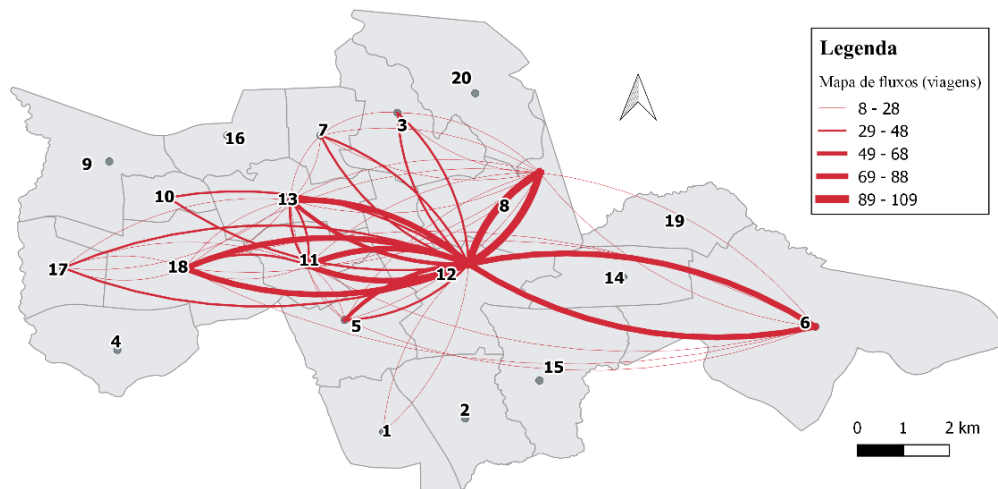
Macrobairro	Bairros contidos	% de viagens
1	Lorenzi	1,1
2	Tomazetti	0,5
3	Itararé, Nossa Senhora do Perpétuo Socorro	2,9
4	Boi-Morto	0,3
5	Nossa Senhora Medianeira, Urlândia, Uglione; Dom Antônio Reis	3,6
6	Camobi	9,3
7	Carolina, Chácara das Flores, Salgado Filho	4,7
8	Km 3, Menino Jesus, Nossa Senhora das Dores, Presidente João Goulart	10,6
9	Agro-Industrial	0,8
10	Nova Santa Marta	2,2
11	Duque de Caxias, Noal, Nossa Senhora de Fátima, Patronato	9,0
12	Centro, Cerrito, Nonoai, Nossa Senhora de Lourdes	31,4
13	Bonfim, Divina Providência, Nossa Senhora do Rosário, Passo D'Areia	8,5
14	São José	0,6
15	João Luiz Pozzobon	0,1
16	Caturrita	1,1

17	Tancredo Neves	4,1
18	Juscelino Kubitschek, Pinheiro Machado, Renascença, São João	8,9
19	Pé-de-Platano	0,2
20	Campestre do Menino Deus	0,0

A Figura 3.5 (a, b), por sua vez, caracteriza (a) a variável dependente quanto ao percentual de escolha de cada “macrobairros” e (b) as linhas de desejo que representam os pares origem-destino que contém quantidade de viagens acima da média. Observa-se que o banco de dados não possui viagens que se destinem ao macrobairro 20, constituído pelo bairro “Campestre do Menino Deus”, conforme sinalizado na figura. Por esse motivo, as análises aqui realizadas para a formulação da estratégia para redução da quantidade de parâmetros não incluem essa alternativa em suas previsões.



(a) Caracterização da variável dependente quanto ao percentual de escolha por “macrobairro”.



(b) Linhas de desejo representativas da maioria das viagens contidas no banco de dados.

Figure 3.5: (a) Distribuição das escolhas de destinos por “macrobairro” e (b) linhas de desejo que representam os pares origem-destino mais frequentes.



## 3.2 MÉTODO

O procedimento metodológico deste trabalho está a seguir apresentado na Figura 3.6, na qual estão representadas as etapas desenvolvidas neste trabalho.

Para a análise desagregada da distribuição de viagens urbanas, realizou-se, primeiramente, o tratamento dos dados oriundos da Pesquisa OD. Em seguida, foram efetuadas as calibrações do modelo Logit Multinomial: (1) modelagem clássica e (2) modelagem a partir de Algoritmos Genéticos.

A modelagem clássica (1) utiliza a técnica usual de calibração do modelo Logit Multinomial tradicional. Durante a modelagem Clássica, testou-se um procedimento metodológico, elaborado a partir de algoritmos de Árvore de Decisão (algoritmo CART), para redução do número de parâmetros a serem estimados. Esta etapa foi realizada para possibilitar a modelagem de escolha dos bairros, devido à dificuldade operacional de calibrar modelos para um grande conjunto de alternativas e de funções utilidade. O método aplicado para a elaboração e validação do procedimento proposto está detalhado na subseção 3.2.1.

A segunda modelagem deste trabalho foi realizada a partir de Algoritmos Genéticos (2), que consistem numa técnica supervisionada que não assume as limitações da calibração tradicional de modelos Logit Multinomiais com grandes conjuntos de alternativas. A fim de determinar a configuração mais eficiente, elaborou-se três tipos diferentes de modelagem com Algoritmos Genéticos: livre, semi-restrito e restrito, os quais estão detalhadas no Capítulo 5.

Finalmente, foram realizadas comparações dos resultados das calibrações e análises de desempenho das técnicas utilizadas nas modelagens e da sua aplicabilidade para a etapa de distribuição de viagens urbanas. Para isso, foram realizadas diversas apreciações, tais como análise qualitativa e distribuições das distâncias de viagens, obtidas a partir da amostra de validação e teste (30%). O uso da distribuição de distâncias de viagens baseia-se na calibração do modelo de distribuição gravitacional, a partir do ajuste das curvas de distribuição de comprimentos de viagens modelada e observada (Ortuzar e Willumsen, 2011).

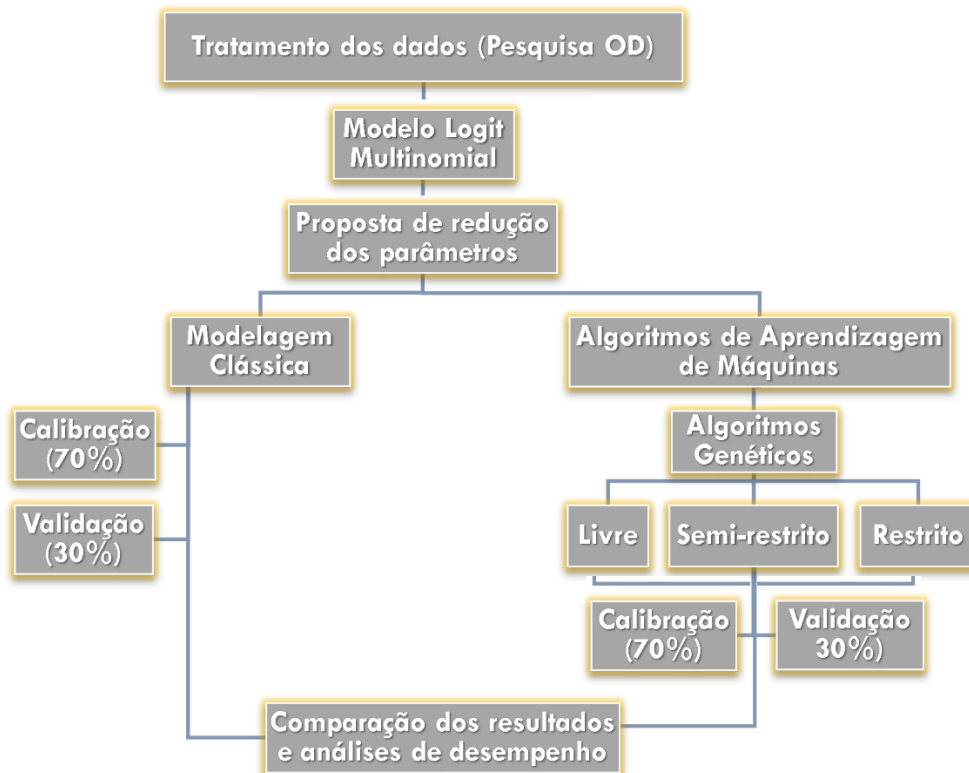


Figura 3.6: Método proposto para calibração de modelo Logit para distribuição de viagens urbanas.

### 3.2.1 Método para estratégia de redução da quantidade de parâmetros

A elaboração de um artifício para redução na quantidade de parâmetros para casos com muitas alternativas seguiu o seguinte procedimento metodológico: primeiramente, realizou-se o tratamento dos dados oriundos da pesquisa domiciliar por “macrobairro”. Não se utilizou o banco de dados organizado por bairros devido à impossibilidade de realizar, a título de validação metodológica, a calibração sem o procedimento para esse nível de agregação, haja vista a grande quantidade de parâmetros a serem estimados nesse caso. Em seguida, separou-se duas amostras de dados, sendo uma para calibração, com 70% dos dados, e outra para validação, contendo os 30% restantes. Definiu-se, então, com o auxílio do teste da razão de verossimilhança, a formulação das funções utilidade a serem usadas no modelo Logit Multinomial. A partir disso, a amostra de calibração foi utilizada, a título de comparação, para a calibração do modelo Logit Multinomial tradicional, sem redução da quantidade de parâmetros. Em seguida, aplicou-se o algoritmo CART para redução da quantidade de parâmetros. Assim, novos modelos Logit Multinomiais foram calibrados, considerando crescentes níveis de redução de parâmetros. Por fim, a partir da amostra de validação, foram feitas comparações e análises de desempenho entre as modelagens com e sem a aplicação do

procedimento metodológico proposto. Para isso, avaliou-se taxas de acertos, distribuição das escolhas dos destinos de viagens e medidas estatísticas, como o rho-quadrado. A Figura 3.7 esquematiza o procedimento utilizado que propicia a redução de parâmetros associados às variáveis do indivíduo.

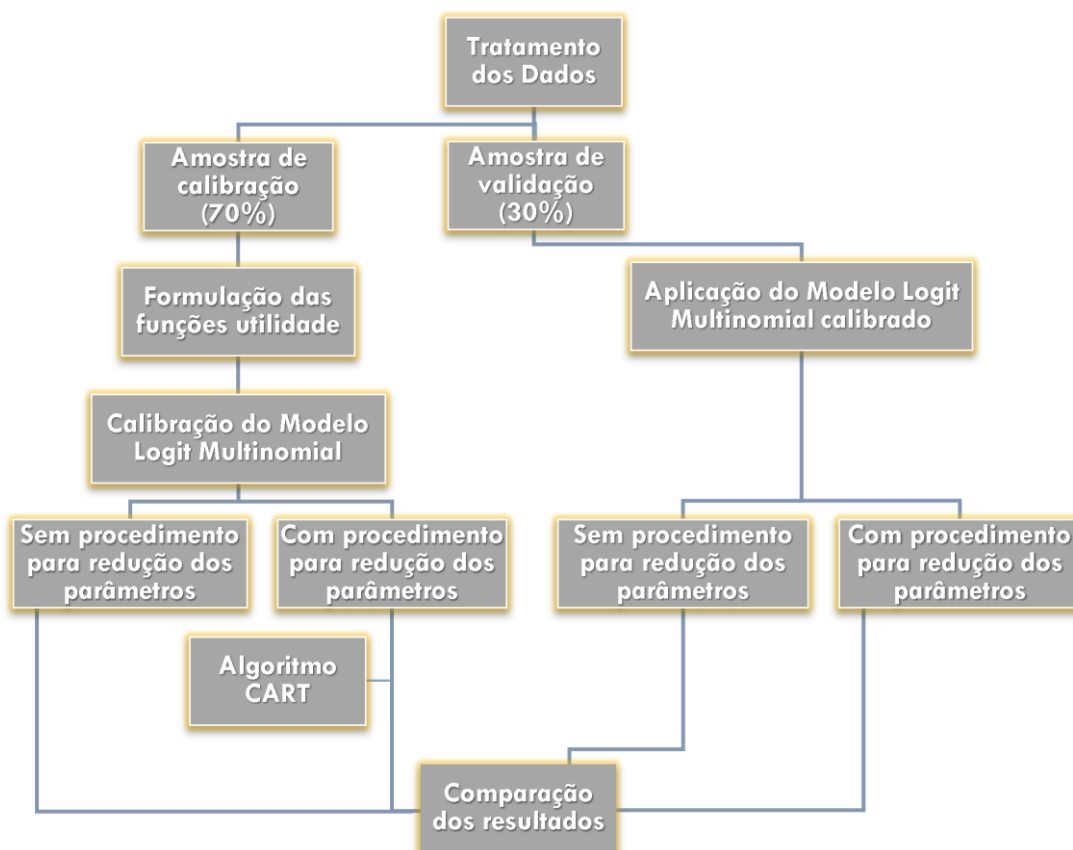


Figura 3.7: Método proposto para estratégia proposta de redução da quantidade de parâmetros do modelo Logit Multinomial para grandes conjuntos de alternativas.

### 3.2.1.1 Detalhamento da estratégia proposta

Conforme exposto no Capítulo 2 deste trabalho, a aplicação de modelos Logit para a etapa de distribuição de viagens urbanas, com grandes conjuntos de alternativas e variáveis, pode significar uma calibração complexa e exigir recursos computacionais muitas vezes não disponíveis. É comum, no entanto, a existência de variáveis que, embora incluídas ao modelo, pouca ou nenhuma influência geram sobre a qualidade do modelo como um todo, causando pouco efeito a sua retirada no que diz respeito à capacidade de representar determinado comportamento, mas que, computacionalmente, geraria um expressivo ganho operacional.

Devido ao caráter paramétrico das calibrações clássicas do modelo Logit Multinomial, essas variáveis são facilmente identificáveis, uma vez que, ao final da calibração, dados sobre a significância dos parâmetros estimados são fornecidos, e, a partir deles e de um nível de significância adotado, é possível a detecção e retirada das variáveis cujos parâmetros resultaram não significativos. Apesar disso, essa vantagem não evita problemas computacionais, uma vez que ainda é necessário que haja a calibração para que se identifique as variáveis que poderiam ser retiradas.

Isso posto, o procedimento proposto neste trabalho, que utiliza algoritmo CART, possibilita a retirada de variáveis potencialmente não relevantes em modelos Logit Multinomiais, na formulação das funções utilidade, etapa prévia à calibração. A retirada de tais variáveis permite viabilizar, operacionalmente, muitas modelagens sem que haja perda relevante de poder preditivo. Vale ressaltar que o procedimento proposto é válido somente para características dos indivíduos. Assim, no caso aplicado neste estudo, de previsão dos destinos de viagens urbanas, foi possível reduzir a quantidade de parâmetros referentes aos indivíduos (socioeconômicos e motivos de viagem).

O procedimento metodológico proposto para redução da quantidade de parâmetros a serem estimados constitui-se de seis etapas, brevemente ilustradas no fluxograma da Figura 3.8 e detalhadas a seguir.

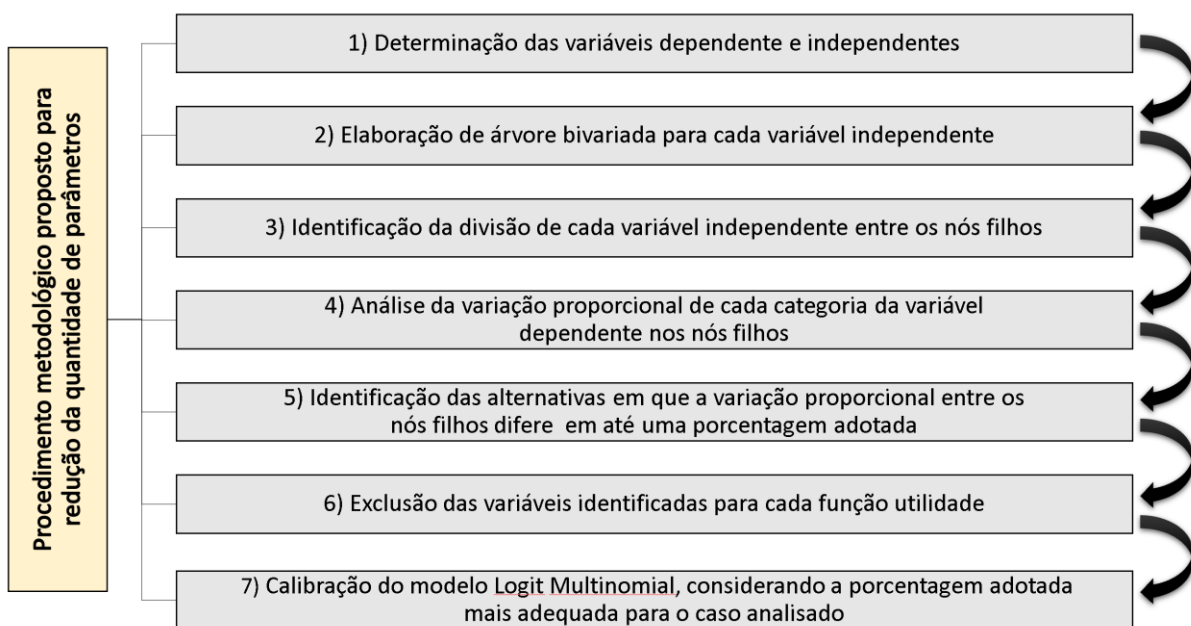


Figura 3.8: Procedimento metodológico proposto para redução da quantidade de parâmetros.

1) Inicialmente, determina-se a variável dependente e as variáveis independentes. Neste trabalho, a variável dependente é categórica, composta de 19 categorias (19 macrobairros possíveis como destino). As 17 variáveis independentes são aquelas descritas na Tabela 3.2, sendo que 15 dessas variáveis não dependem do destino (representadas por A a O), referentes ao indivíduo, e 2 referentes aos possíveis destinos (representadas por P e Q). Assim, neste trabalho, serão consideradas no procedimento 15 variáveis independentes, referentes ao indivíduo.

2) Em seguida, são geradas árvores bivariadas (uma para cada variável que independe da alternativa), com apenas uma divisão do conjunto total dos dados (variável dependente x variável independente 1, variável dependente x variável independente 2, ..., variável dependente x variável independente 16). A Figura 3.9 ilustra um exemplo de uma árvore relacionada a uma variável independente 1 categórica (com quatro categorias) e a uma variável dependente com x categorias (x alternativas).

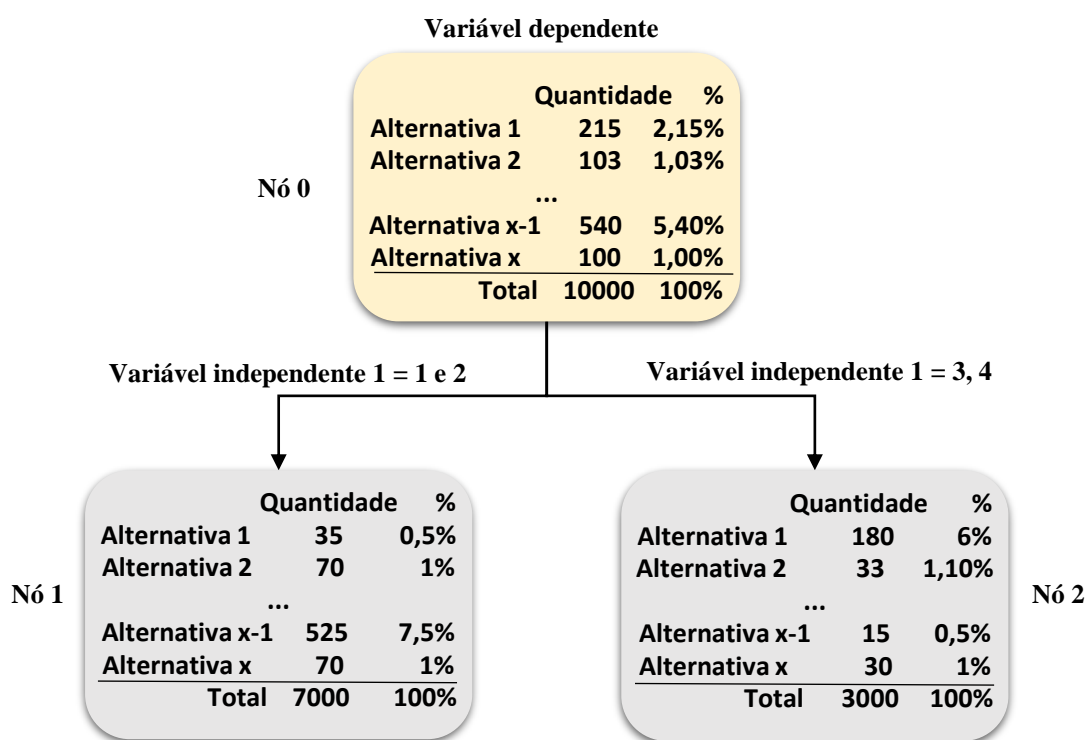


Figura 3.9: Exemplo do procedimento adotado com Árvore de Decisão para redução da quantidade de parâmetros a serem estimados.

3) Observa-se, para cada árvore gerada, a divisão das categorias ou valores da variável independente entre os nós filhos (ou nós terminais), mediante divisão do nó pai. No caso do

exemplo da Figura 3.9, o Nó 1 contém as ocorrências com variável independente 1 de categorias 1 ou 2, enquanto o Nó 2 contém os casos em que a variável independente 1 vale 3 ou 4.

4) Analisa-se, nos nós filhos (Nó 1 e Nó 2), a variação proporcional de cada categoria da variável dependente, mediante divisão do nó pai, tomando como base o maior valor entre essas proporções. Vale ressaltar que cada categoria da variável dependente corresponde a uma alternativa no modelo Logit.

No exemplo da Figura 3.9, a alternativa 1 responde por 0,5% das escolhas representadas pelo Nó 1, ou seja, 0,5% das escolhas cuja variável independente 1 vale 1 ou 2 apresentam a alternativa 1 como alternativa escolhida. Já para escolhas representadas pelo Nó 2, que são aquelas cujo valor da variável independente é 3 ou 4, a alternativa 1 surge como escolha em 6% dos casos. Assim, a variação proporcional da escolha da alternativa 1 entre os nós filhos, tomando como base o maior valor (6%), é de aproximadamente 92%.

$$Variação_1 = \frac{(6\% - 0,5\%)}{6\%} = 0,9167 = 91,67\% \quad (3.2)$$

5) Identifica-se, para cada uma das variáveis independentes, as alternativas/categorias em que a sua variação proporcional entre os nós filhos difere em até uma porcentagem determinada. O objetivo deste método é identificar, para cada variável independente, variações relativamente altas nas categorias da variável dependente. Esses casos sinalizam uma influência e/ou relevância de determinada variável explicativa na utilidade de uma alternativa específica.

Por exemplo, ao analisar a Figura 3.9, percebe-se que a variável independente 1, quando assume os valores 1 e 2, influencia positivamente a escolha da alternativa x-1 (7,5% das escolhas do Nó 1 são atribuídas a esta). Isso pode ser verificado ao observar que, quando a mesma variável assume os valores 3 e 4, a proporção de escolha da alternativa x-1 entre as escolhas classificadas no Nó 2 é bem pequena (0,5%). Assim, há uma alta variação proporcional (93,33%) na alternativa x-1, mediante divisão do nó pai a partir da variável independente 1. Para o caso da alternativa 2, no entanto, percebe-se que a proporção de indivíduos que a escolhem é semelhante entre os nós filhos 1 e 2 (1% e 1,10%, respectivamente), indicando que a variável independente 1 não exerce influência significativa sobre a escolha dessa alternativa (variação de 9,09%).

6) As variáveis independentes relacionadas às alternativas identificadas na etapa (5) são excluídas das respectivas funções utilidade. No exemplo indicado, seria excluída a variável independente 1 das funções utilidade 2 e x.

### 3.2.1.2 Restrições Metodológicas

O algoritmo CART, utilizado para realização do procedimento proposto, realiza divisões binárias a partir dos valores das variáveis independentes segundo as categorias da variável dependente (Breiman et al., 1984). A partir disso, propõe-se a identificação, para cada variável independentes, das alternativas/categorias em que a sua variação proporcional entre os nós filhos difere em até uma porcentagem estabelecida. No caso avaliado neste trabalho, no qual as alternativas referem-se às escolhas dos destinos, as variáveis que não são referentes aos indivíduos, “pontos de ônibus” e “distância”, caracterizam as alternativas, e, por isso, não podem ser utilizadas na aplicação da estratégia descrita.

Para melhor ilustrar o motivo do método proposto não poder ser aplicado a variáveis que caracterizem as alternativas, exhibe-se um exemplo de banco de dados (Tabela 3.4). Nele, cada observação relaciona-se ao indivíduo. Assim, as variáveis relativas aos indivíduos variam entre os indivíduos e não variam entre as alternativas. Por outro lado, as variáveis relativas às alternativas variam entre as alternativas, mas não variam entre os indivíduos. Para o caso do modelo Logit Multinomial, no qual a característica de determinada alternativa compõe apenas a utilidade daquela alternativa, isso não seria um problema. No entanto, para a aplicação do algoritmo CART, o que interessa é a variação das variáveis explicativas entre as observações (ou seja, entre os indivíduos). Neste caso, como pode ser observado no exemplo da Tabela 3.4, as variáveis relativas às alternativas (pontos de parada de ônibus nos destinos) tornam-se uma constante quando observadas sob o ponto de vista do indivíduo.

Tabela 3.4: Exemplo de banco de dados com variáveis relativas ao indivíduo e às alternativas.

Id	Renda (em US\$)	Idade (em anos)	Pontos de ônibus no destino 1	Pontos de ônibus no destino 2	...	Pontos de ônibus no destino 19
1	1,000.00	32	8	15	...	51
2	1,500.00	51	8	15	...	51
3	800.00	20	8	15	...	51
4	3,000.00	60	8	15	...	51
...	...	...	...	...	...	...

Outra restrição metodológica que vale a pena ser ressaltada é o fato de o procedimento proposto necessitar da adoção de uma porcentagem para análise da variação proporcional das variáveis independentes entre os nós filhos, conforme citado no item (5) da subseção 3.2.1.1. A adoção dessa porcentagem configura-se como restrição metodológica pelo fato de não ocorrer de maneira automática, sendo necessário conhecimento prévio acerca das alternativas e do

contexto no qual o estudo de caso está inserido. Dessa forma, trata-se de uma estratégia exploratória, pois se assume a necessidade de realização de testes para determinação da variação proporcional adequada.

Além disso, para os leitores que queiram replicar o método para casos em que são consideradas variáveis espaciais e que necessitem agrupamento de áreas geográficas, como no caso aqui apresentado, orienta-se que tais variáveis sejam consideradas no agrupamento, juntamente com as coordenadas geográficas, de forma a evitar problemas de Unidade de Área Modificável (MAUP - *Modifiable Areal Unit Problem*).

### 3.3 APLICATIVOS UTILIZADOS

Os aplicativos Microsoft Excel e IBM SPSS 24.0 foram utilizados para o tratamento do banco de dados da Pesquisa Domiciliar. A análise de *Cluster*, para agrupamento dos bairros, e a Árvore de Decisão, para redução de parâmetros para calibração do Logit, foram realizadas por meio do *software* IBM SPSS 24.0. Os centroides dos “macro Bairros” e dos bairros e as distâncias entre eles foram obtidos com auxílio do *software* QGIS (versão 3.6.3). A formulação das funções utilidade e as calibrações clássicas do modelo Logit Multinomial, por sua vez, foram efetuadas com o auxílio da versão 3.1.1 do aplicativo aberto BIOGEME - PandasBiogeme (Bierlaire, 2018). Por fim, o aplicativo Visual Studio 2017 foi utilizado para a elaboração do código para a calibração do modelo Logit Multinomial por meio de Algoritmos Genéticos.

Vale destacar que, de forma a evitar interrupções e minimizar os problemas operacionais advindos das operações computacionalmente onerosas, as calibrações, tanto por meio do aplicativo PandasBiogeme (clássica) como por Algoritmos Genéticos, foram realizadas em uma máquina virtual robusta, individual e personalizada, disponibilizada pela Seção Técnica de Informática da Escola de Engenharia de São Carlos da Universidade de São Paulo. A máquina virtual era acessada remotamente e apresentava as seguintes configurações: sistema operacional Windows Server 2012, 16GB de memória RAM, CPU com 8 processadores virtuais e HD com 200GB.

A Tabela 3.5 sintetiza as informações supracitadas, expondo as unidades de análise e os aplicativos utilizados em cada etapa metodológica.



Tabela 3.5. Síntese dos dados e aplicativos utilizados em cada etapa metodológica.

Etapa metodológica	Unidade de Análise	Software	
Tratamento dos dados	Pesquisa domiciliar	Indivíduo	Microsoft Excel; IBM SPSS 24.0
	Agrupamento dos bairros	Bairros	IBM SPSS 24.0
	Obtenção das distâncias	Bairros e “macrobairros”	QGIS (3.6.3)
Formulação das funções utilidade	Bairros e “macrobairros”	PandasBiogeme	
Calibração clássica do modelo Logit Multinomial	Bairros e “macrobairros”	PandasBiogeme	
Procedimento para redução de parâmetros	Macro Bairros	IBM SPSS 24.0	
Calibração do modelo Logit Multinomial por Algoritmos Genéticos	Bairros	Visual Studio 2017	



## 4 RESULTADOS ASSOCIADOS À ESTRATÉGIA PARA REDUÇÃO DA QUANTIDADE DE PARÂMETROS A SEREM ESTIMADOS EM MODELOS DE ESCOLHA DISCRETA PARA CASOS COM GRANDES CONJUNTOS DE ALTERNATIVAS

*Este capítulo discorre sobre os resultados relativos à estratégia elaborada para redução no número de parâmetros em situações com grandes conjuntos de alternativas, como o caso da escolha de destinos urbanos. A eficácia do procedimento proposto é discutida ao final desta seção, no qual são apresentadas comparações entre os modelos com e sem a aplicação da estratégia proposta.*

### 4.1 FORMULAÇÃO DAS FUNÇÕES UTILIDADE

Conforme citado no Capítulo 3 deste trabalho, a estratégia proposta para redução da quantidade de parâmetros foi testada a partir dos 20 “macro Bairros” obtidos com a agregação dos 41 bairros do município de Santa Maria (RS). Essa agregação foi necessária para que a calibração sem a estratégia testada pudesse ser realizada para posterior comparação e análise da eficácia do procedimento.

A partir da amostra de calibração, composta por 70% dos dados organizados por “macro Bairros”, as calibrações foram realizadas. As funções utilidades utilizadas para a calibração inicial do modelo Logit Multinomial contém as seguintes variáveis explicativas, já expostas no Capítulo 3: (A) Frequência Escolar; (B) Gênero; (C) Grau de instrução; (D) Idade; (E) Moradores; (F) Motivo estudo; (G) Motivo integração; (H) Motivo lazer; (I) Motivo residência; (J) Motivo saúde; (K) Motivo trabalho; (L) Renda; (M) Trabalha; (N) Veículos motorizados; (O) População origem; (P) Distância e (Q) Parada de ônibus.

Com relação aos atributos relativos aos indivíduos, como não variam entre as alternativas, devem ser incluídos ao modelo de forma específica, de modo a criar diferenças entre as utilidades das alternativas (Train, 2009). Contudo atributos relativos às alternativas variam entre as mesmas. Nesse caso, há duas formulações possíveis para essas variáveis: genérica e específica.

Para a formulação genérica, assume-se que os coeficientes dessas variáveis possuem o mesmo valor para todas as alternativas, ou seja, a variação dessas variáveis é sentida da mesma forma, qualquer que seja a alternativa analisada (utilidades marginais iguais) (Ben-Akiva e Lerman, 1985). Assim, no caso da situação em questão, na qual os atributos sobre as alternativas

são “Distância” e “Pontos de Ônibus”, os parâmetros dessas variáveis não dependem dos destinos, ou seja, possuem o mesmo valor para todas as funções utilidade.

No caso da formulação específica, assume-se que variações nas variáveis sobre as alternativas são sentidas de forma diferente para cada alternativa. Assim, a utilidade de cada alternativa apresenta parâmetros distintos para essas variáveis (Ben-Akiva e Lerman, 1985).

Para determinar qual formulação é a mais adequada para as variáveis “Distância” e “Pontos de Ônibus”, realizou-se o teste da razão de verossimilhança, conforme orientado por Bierlaire et al. (2016). A hipótese nula assume, então, que os coeficientes dessas variáveis são iguais para todas as alternativas.

A hipótese nula é rejeitada para o caso em que:

$$-2 \cdot (\mathcal{L}_R - \mathcal{L}_U) > \chi^2_{((1-\alpha), df)} \quad (4.1)$$

Onde:

$\mathcal{L}_R$ : logaritmo da verossimilhança final para formulação genérica;

$\mathcal{L}_U$ : logaritmo da verossimilhança final para formulação específica;

$\alpha$ : nível de significância adotado;

$df$ : graus de liberdade.

Para a calibração, assume-se nível de significância de 5% e 18 graus de liberdade (obtido pela diferença entre a quantidade de parâmetros estimados entre os modelos específico e genérico). Assim, calibrados os modelos genérico e específico para a variável “Distância”, tem-se:

$$-2 \cdot (-4077,066 - (-4037,637)) = 78,858 > 28,865 \quad (4.2)$$

Logo, a hipótese nula foi rejeitada, o que significa que a formulação específica é a mais adequada para a variável “Distância”.

Para a variável “Pontos de Ônibus”, por sua vez, tem-se:

$$-2 \cdot (-4077,066 - (-4077,033)) = 0,066 < 28,865 \quad (4.3)$$

Logo, a hipótese nula foi aceita, o que significa que a formulação genérica é a mais adequada para a variável “Pontos de Ônibus”.

Assim, a configuração geral das funções utilidade para a calibração por “macro Bairros” seguiu a seguinte configuração:

$$\begin{aligned}
 V_{ij} = & a_i + b_i \cdot \text{FrequênciaEscolar}_j + c_i \cdot \text{Gênero}_j + d_i \cdot \text{GraudeInstrução}_j + e_i \\
 & \cdot \text{Idade}_j + f_i \cdot \text{Moradores}_j + g_i \cdot \text{MotivoEstudo}_j + h_i \\
 & \cdot \text{MotivoIntegração}_j + i_i \cdot \text{MotivoLazer}_j + j_i \cdot \text{MotivoResidência}_j \\
 & + k_i \cdot \text{MotivoSaúde}_j + l_i \cdot \text{MotivoTrabalho}_j + m_i \cdot \text{Renda}_j + n_i \\
 & \cdot \text{Trabalha}_j + o_i \cdot \text{VeículosMotorizados}_j + p_i \cdot \text{PopulaçãoOrigem}_j \\
 & + q_i \cdot \text{Distância}_{ij} + r \cdot \text{PontosdeÔnibus}_{ij}
 \end{aligned} \tag{4.4}$$

Onde:

$V_{ij}$ : utilidade da alternativa  $i$  para o indivíduo  $j$ ;

$a_i, b_i, c_i, \dots, q_i$ : coeficientes a serem estimados para a alternativa  $i$ ;

$r$ : coeficiente estimado para todas as alternativas, referente à variável “Pontos de Ônibus”.

## 4.2 APLICAÇÃO DO PROCEDIMENTO PROPOSTO

O procedimento detalhado na subseção 3.2.1 do Capítulo 3 foi aplicado para o estudo da previsão de destinos urbanos da cidade de Santa Maria (RS), com o banco de dados detalhado na subseção 3.1 Assim, 15 árvores binárias foram geradas, uma para cada variável independente relacionada ao indivíduo.

A seguir, apresenta-se a árvore resultante para a variável “Motivo Estudo”, em que 0 significa que a viagem realizada pelo indivíduo não teve o estudo como motivo, enquanto o valor 1 é atribuído às viagens com motivo estudo. A caixa superior (Nó 0) refere-se ao nó pai, e as caixas subsequentes são os nós filhos (ou nós terminais). Os símbolos coloridos representam os destinos possíveis (“macro Bairros”).

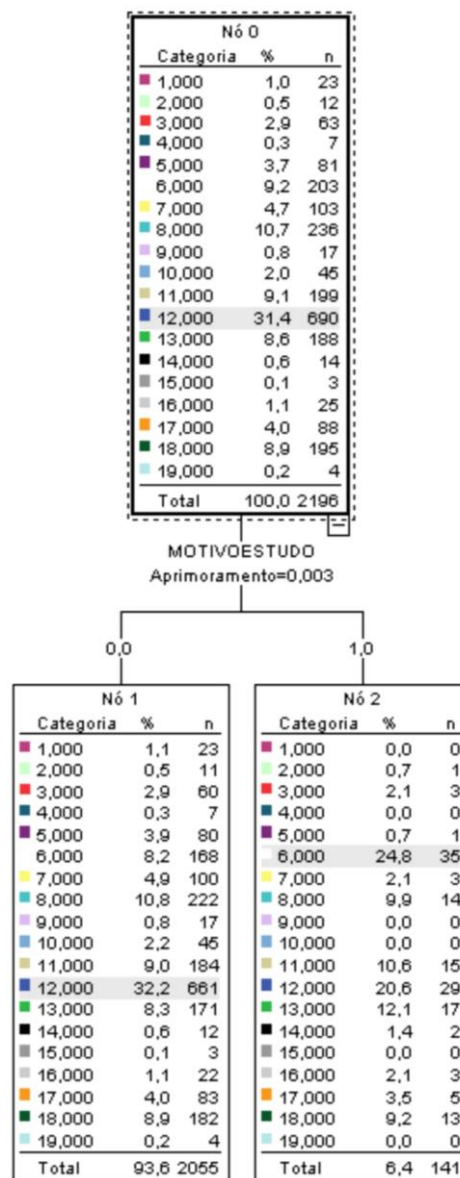


Figura 4.1: CART para a variável independente “Motivo Estudo” e variável dependente “macro Bairros”.

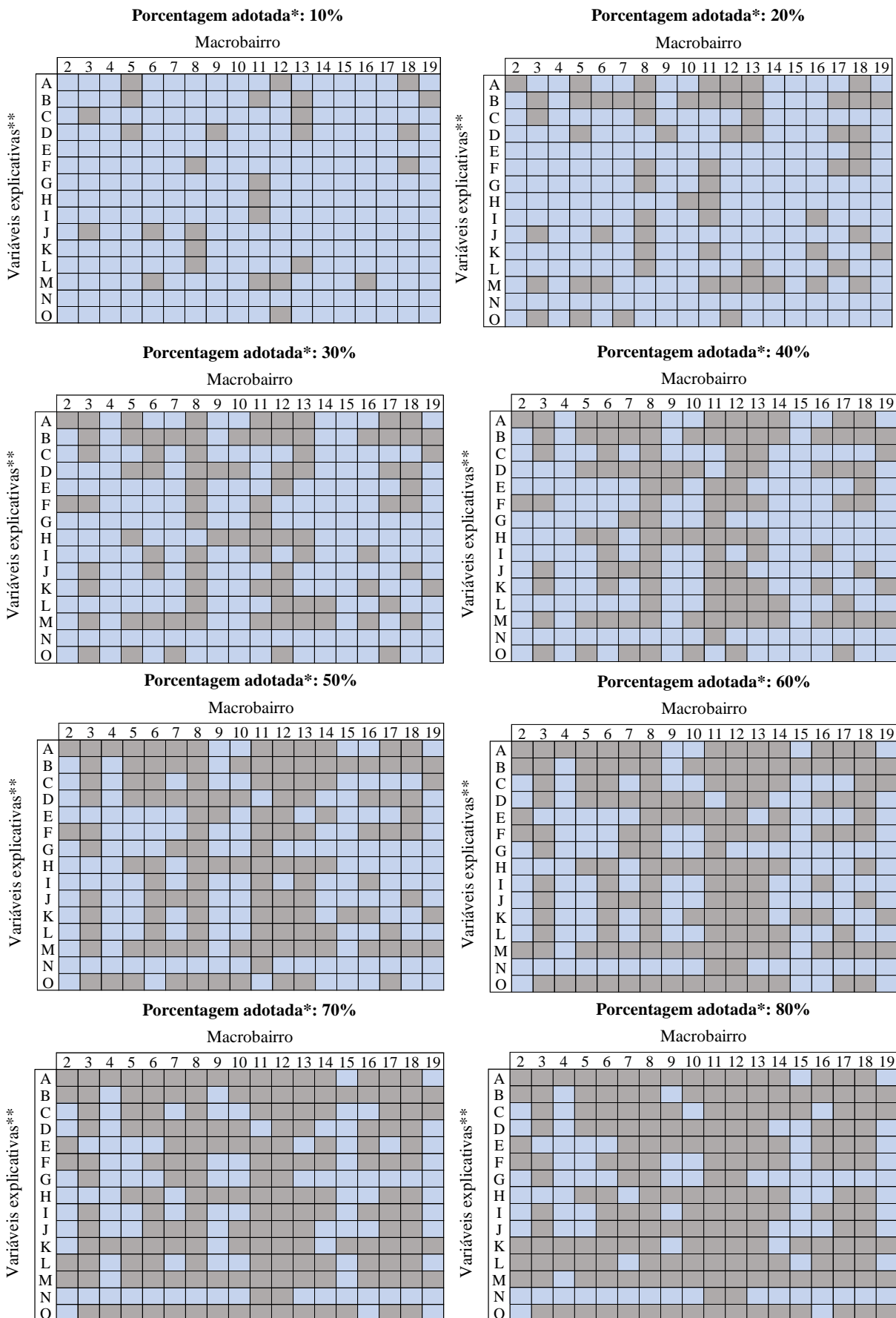
Seguindo o raciocínio aplicado para o exemplo anterior, a divisão proposta pelo algoritmo CART indica que o fato de a viagem ter ou não estudo como motivação influencia na escolha de alguns “macro Bairros” como destino.

O “macro bairro” 6, por exemplo, é destino de 8,2% de viagens que não são motivadas por estudo, mas é destino de 24,8% das viagens com essa motivação. Essas proporções possuem uma variação de aproximadamente 67%, o que indica possível influência positiva dessa variável para a utilidade do “macro bairro” 6. Esse resultado é facilmente compreendido ao constatar nesse “macro bairro”, constituído pelo bairro Camobi, a presença da Universidade Federal de Santa Maria (UFSM), importante polo gerador de viagens para estudantes no município.

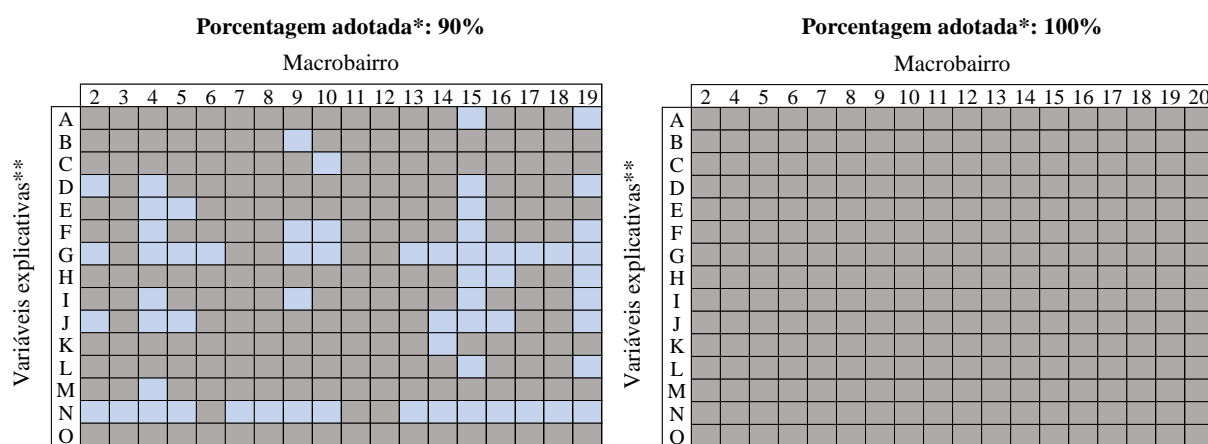
Ressalta-se, no entanto, que nem sempre o fato de a maioria das viagens que se destinem a determinado “macrobairro” apresentarem a motivação estudo indica que essa motivação é uma variável relevante na utilidade do referente destino. A discrepância nas quantidades pode ser justificada por discrepâncias já existentes no banco de dados como um todo. Observando o exemplo do “macrobairro” 8 na Figura 4.1, verifica-se que, embora sejam encontradas viagens com motivo estudo (14, no total - 9,9% das viagens representadas no nó 2), a grande maioria não é influenciada por essa motivação. Esse fato pode ser explicado pela diferença já existente no banco de dados com relação à variável “Motivo Estudo” (141 viagens ocorrem por motivo estudo e 2.055 não apresentam essa motivação). Desse modo, justifica-se a necessidade de analisar, para cada alternativa, as proporções das quantidades escolhidas em cada nó terminal, e não os valores absolutos.

Após a retirada dos parâmetros como indicado na subseção 3.2.1 do Capítulo 3, realizou-se calibrações do modelo Logit Multinomial para porcentagens adotadas de 10%, 20%, 30%, ..., 100%. Para cada modelagem, diversas calibrações foram executadas, de modo que, para cada uma delas, excluiu-se os parâmetros que resultaram não significativos na calibração anterior, ao considerar níveis de significância de 5% e 10% (p-valor associado ao teste t maior do que 0,05 e 0,1, respectivamente). Assim, repetiu-se esse procedimento até que cada calibração final contivesse todos os parâmetros significativos.

A Figura 4.2 apresenta um resumo das variáveis (representadas por A a O, conforme Tabela 3.2) que foram incluídas (representadas em azul) e excluídas (representadas em cinza) das funções utilidade a partir do procedimento testado e de diferentes valores de porcentagens (variações proporcional das alternativas/categorias entre os nós terminais) adotadas, com exceção do “macrobairro” 1, não representada por sua utilidade ter sido adotada como referência para a calibração do modelo, e do “macrobairro 3”, uma vez que não há registros de viagens que se destinem à essa localidade. Assim, o modelo contém 18 funções que seguem a formulação geral representada pela Equação (4.4), o que representa um total de 270 parâmetros possíveis de serem eliminados pelo procedimento proposto.







\* Porcentagem adotada como limite da variação proporcional entre os nós filhos obtidos na árvore de decisão para cada alternativa de destino.

\*\* A: Frequência escolar; B: Gênero; C: Grau de instrução; D: Idade; E: Moradores; F: Motivo estudo; G: Motivo integração; H: Motivo lazer; I: Motivo residência; J: Motivo saúde; K: Motivo trabalho; L: Renda; M: Trabalha; N: Veículos motorizados; O: População origem.

Figura 4.2: Variáveis associadas a parâmetros significativos (em azul) e não significativos (em cinza) das funções utilidade de cada “macrobairro”, a partir de diferentes porcentagens adotadas.

Percebe-se que, quanto maior a porcentagem adotada para analisar a variação proporcional das categorias da variável dependente entre os nós filhos, maior é a quantidade de parâmetros excluídos do modelo. Maior será também, no entanto, a possibilidade de excluir parâmetros significativos ao modelo. No caso da variável “Motivo Estudo”, representada pela letra “F” e avaliada na Figura 4.1, por exemplo, o parâmetro se mantém na utilidade do “macrobairro” 6 para porcentagens de até 60%.

Diferentemente da aplicação de modelos Logit para a etapa de escolha modal, na qual a compreensão da influência das variáveis sobre as utilidades dos modos de transporte ocorre facilmente de maneira intuitiva, a análise qualitativa dos parâmetros nas funções utilidade no caso da escolha dos destinos urbanos não é facilmente verificada, haja vista as diferentes características e equipamentos presentes em cada alternativa.

Nesse sentido, características marcantes e a presença de estabelecimentos tidos como referência auxiliam na compreensão dos fatores que influenciam a utilidade de determinado destino de viagem, como no caso da análise do “macrobairro” 6, na qual a presença de uma grande universidade (Universidade Federal de Santa Maria) explicaria a grande porcentagem de viagens por motivo estudo que se destinam a ela.

Outra característica marcante que pode auxiliar na interpretação da Figura 4.2 é o fato de o “macrobairro” 9, constituído unicamente pelo bairro “Agro-industrial”, ser o único com



**Sem procedimento** (nível de significância: 10%)

		Macrobairro																	
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Variáveis explicativas	Frequência Escolar																		
	Gênero																		
	Grau de Instrução																		
	Idade																		
	Moradores																		
	Motivo Estudo																		
	Motivo Integração																		
	Motivo Lazer																		
	Motivo Residência																		
	Motivo Saúde																		
	Motivo Trabalho																		
	Renda																		
	Trabalha																		
	Veículos Motorizados																		
População Origem																			

Figura 4.3: Variáveis associadas a parâmetros significativos (em azul) e não significativos (em cinza) das funções utilidade de cada “macrobairro” após as calibrações sem o procedimento proposto, segundo níveis de significância de 5% e 10%.

Na Figura 4.4, os valores dos parâmetros estimados significativamente diferentes de zero estão ilustrados para os cenários sem a aplicação do procedimento proposto e com a aplicação dele, considerando porcentagens adotadas de 10%, 50% e 80%. Analisando, de maneira geral, percebe-se que os parâmetros que permanecem na calibração com o procedimento têm sinais e valores similares àqueles estimados sem o procedimento.

Observando-se o exemplo da função utilidade do “macrobairro” 6, constituído pelo bairro Camobi, associado à Localização da Universidade Federal de Santa Maria, inicialmente são calibrados (sem o procedimento) 6 parâmetros, associados a características do indivíduo e do domicílio (E: Moradores) aos motivos de viagem (F: Motivo estudo; G: Motivo integração; H: Motivo lazer e I: Motivo residência) e ao “macrobairro” de origem (O: População origem). Considerando uma porcentagem adotada equivalente a 10%, a calibração permanece bastante similar, em termos de valores e sinais de parâmetros. Adotando-se a porcentagem de 50%, apenas três parâmetros permanecem na equação, relacionados ao motivo de viagem (F: Motivo estudo e G: Motivo integração) e ao “macrobairro” de origem (O: População Origem). Verifica-se que, em módulo, há um aumento nos valores estimados. Resultado este coerente, uma vez que a influência das variáveis retiradas possivelmente é compensada pelas variáveis remanescentes, que continuam com capacidade explicativa da escolha do “macrobairro” 6. Finalmente, para porcentagens ainda mais altas (80%), é estimado um único parâmetro,

associado à variável “G: Motivo integração”. Como não há nenhum terminal de integração do “macrobairro” 6, justifica-se o sinal negativo associado a este parâmetro.

De maneira análoga, para a função utilidade do “macrobairro” 10, constituído pelo bairro “Nova Santa Marta”, inicialmente (sem o procedimento) são calibrados 9 parâmetros, sendo 4 associados a características do indivíduo e do domicílio (A: Frequência escolar; C: Grau de instrução; D: Idade e E: Moradores), 4 aos motivos de viagem (H: Motivo lazer; I: Motivo residência; J: Motivo saúde e K: Motivo trabalho) e 1 ao “macrobairro” de origem (O: População origem). Considerando uma percentagem adotada equivalente a 10%, a calibração permanece bastante similar, com exclusão apenas da variável relacionada ao “macrobairro” de origem (O: População origem). Adotando-se a percentagem de 50%, por sua vez, as variáveis “D: Idade” e “E: Moradores” também são excluídas. A exclusão dessas variáveis, no entanto, é compensada pela presença das variáveis “F: Motivo estudo” e “G: Motivo integração”, que possivelmente ganharam capacidade explicativa com a exclusão de variáveis devido à aplicação do procedimento. Já para a porcentagem de 80%, somente as variáveis “C: Grau de instrução”, “F: Motivo estudo” e “G: Motivo integração” se mantém, porém com considerável alteração nos valores estimados, o que pode ser explicado pela readequação do modelo às variáveis remanescentes.

**Sem procedimento**

		Macrobairro																		
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Variáveis explicativas**	A									-2,64					6,43				-6,61	
	B								-1,49											
	C			4,77						-3,65	1,19	1,39			8,59					
	D									-4,29										
	E	-5,89			-2,06	-1,86		-1,63		-2,03	-1,91	-2,03						-2,55	-2,69	
	F			-4,08		1,66				-6,64					2,33					0,70
	G		5,25		-2,01	-4,00	5,24	4,92			5,04	5,70	-4,05				-1,38	-2,49	-3,48	
	H	2,60				0,71				5,47					6,07	-5,32				
	I			-9,06		-2,46		-2,08	-3,73	4,78	-1,38	-2,61	1,60		-6,18					-8,70
	J				-5,68					6,13					-3,12		-5,16			-5,10
	K	2,09								4,31					-5,20	6,11				
	L							-2,64								-4,39			-1,77	
	M			5,42												-2,53				
	N															8,29				
	O		1,45			3,16		0,79	3,02	2,28	1,16	1,07	2,58					1,61	2,28	

**Porcentagem adotada\*: 10%**

Macrobairro

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A		0,88							-2,61					7,35				-6,47
B								-1,54										
C			3,25	-1,23		-1,20			-4,52							-1,97	-1,30	
D		1,90							-4,29									
E				-2,55	-1,60		-1,56		-2,54	-1,38	-2,28	-1,35				-3,01	-2,71	
F			-4,53		1,32			-4,70			-0,71			-3,22				
G	-3,52		-4,64	-6,51	-7,96			-3,92				-7,86	-5,64		-6,64	-7,31	-7,53	-5,06
H	3,33		1,81		-0,64				8,34					6,17	-6,38			
I			-7,67	1,29	-1,35	1,32	-0,96		8,75	-1,57				-4,42			1,21	-7,44
J	-2,88		-3,14	-6,10				3,19	9,08				-5,22		-6,29			-4,89
K	2,51							2,87	7,09				-6,93	5,28		-1,22		
L				-0,93		-2,71									-1,44		-1,61	
M			6,62															
N																		
O	-2,10				2,10										-2,32		1,46	

**Porcentagem adotada\*: 50%**

Macrobairro

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A									-1,23					1,64				-6,22
B								-1,53										
C			3,46			-1,18			-2,72					2,53		-1,74	-1,25	
D													-13,00					
E		3,08				1,95												4,64
F			-3,72		1,76			-3,96	-5,58					-6,52				
G	-2,50		-3,19	-6,18	-7,32			-3,04	-5,34	0,73	-7,68	-5,07		-5,12	-6,41	-6,28	-7,10	-3,60
H	3,34		1,88												-6,28			
I		1,17	-6,61	1,59		1,90			2,22	-0,80				-7,15		1,73	1,79	-7,21
J	-2,62			-6,07				3,03	1,34				-4,96	-3,08	-5,91			-4,71
K	2,76					-2,98		2,61					-6,94					
L				-2,25					-1,85						-1,87		-1,93	
M			5,74															
N	-6,44																	
O	-1,64				1,74					0,56					-2,50		1,39	

**Porcentagem adotada\*: 80%**

Macrobairro

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
A														8,98				-7,62	
B								-1,41											
C			3,00						-3,11										
D																			
E		3,22																4,67	
F			-6,93					-7,15	-8,52					-8,27				-4,54	
G	-7,19		-6,41	-8,84	-9,44			-6,89	-8,10				-9,36	-7,76	-7,53	-7,97	-8,65	-9,37	-6,75
H	1,91		1,67			-2,10										-7,74			
I			-8,55	0,86										-8,91				-9,19	
J	-6,89		-5,33	-8,52									-7,47	-5,76	-7,59			-7,22	
K								2,29					-8,82						
L						-2,59													
M			7,85																
N	-6,19																		
O															-2,65				

\* Porcentagem adotada como limite da variação proporcional entre os nós filhos obtidos na árvore de decisão para cada alternativa de destino.

\*\* A: Frequência escolar; B: Gênero; C: Grau de instrução; D: Idade; E: Moradores; F: Motivo estudo; G: Motivo integração; H: Motivo lazer; I: Motivo residência; J: Motivo saúde; K: Motivo trabalho; L: Renda; M: Trabalha; N: Veículos motorizados; O: População origem.

Figura 4.4: Detalhamento das calibrações para o modelo sem o procedimento e com o procedimento (porcentagens adotadas de 10%, 50% e 80%).

A Figura 4.5 ilustra a quantidade de parâmetros excluídos do modelo à medida em que se aumenta a porcentagem adotada para o procedimento metodológico proposto. Conforme dito anteriormente, quanto maior a porcentagem (variação proporcional da categoria da variável dependente/alternativa) adotada, maior é a quantidade de parâmetros excluídos do modelo.

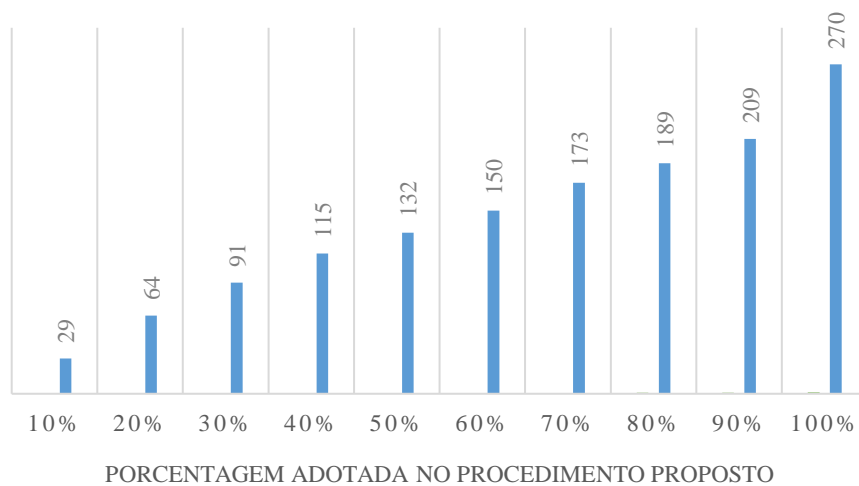


Figura 4.5: Quantidade de parâmetros excluídos conforme aumento da porcentagem adotada para o procedimento proposto.

### 4.3 COMPARAÇÕES E ANÁLISES DE DESEMPENHO

A fim de analisar a eficácia da estratégia proposta em diminuir a quantidade de parâmetros em modelos de escolha discreta desagregados, comparou-se a calibração realizada sem a aplicação do método e as calibrações executadas para diferentes níveis de porcentagem adotados para retirada de parâmetros.

As comparações foram realizadas a partir dos parâmetros retirados em cada modelo, das medidas de qualidade de ajuste alcançadas com as calibrações (rho-quadrado, rho-quadrado ajustado, log-verossimilhança final e Critério de Informação de Akaike), a análise da taxa de acertos e da distribuição das escolhas dos destinos, obtidos a partir da amostra de validação.

- Medidas de qualidade de ajuste

A Figura 4.6 apresenta os valores de  $\rho^2$  e  $\rho^2$  ajustado para a modelagem sem a aplicação do método proposto e com a aplicação, considerando diferentes níveis de porcentagens adotados na etapa de análise de variação proporcional de cada alternativa entre os nós filhos

gerados pelo algoritmo CART e considerando a remoção de parâmetros que resultaram não significativos para os níveis de significância de 10% e 5% durante a calibração do modelo Logit Multinomial. Ambos os valores foram fornecidos no arquivo de saída do *software* PandasBiogeme (Bierlaire, 2018). Observa-se que os valores iniciais estão no intervalo entre 0,2 e 0,4, considerado bom para esse tipo de modelo (McFadden, 1978b). Nota-se, também, uma tendência decrescente dos valores de  $\rho^2$  a medida em que se aumenta a porcentagem adotada no método proposto. Essa tendência ocorre como o esperado, uma vez que a probabilidade de exclusão de variáveis significativas aumenta à medida que se aumenta a porcentagem adotada no procedimento.

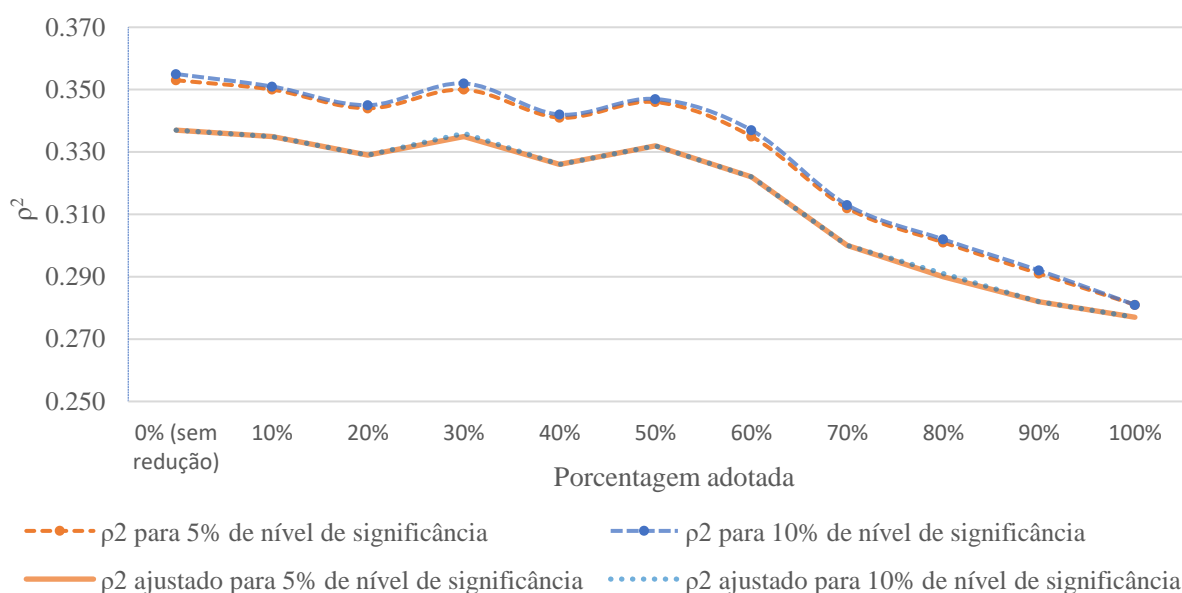


Figura 4.6. Valores de  $\rho^2$  e  $\rho^2$  ajustado para cada calibração executada.

Outra medida utilizada para comparação foi o valor final do log da verossimilhança, obtido na etapa de calibração. A Figura 4.7 ilustra os valores finais do log da verossimilhança usando a estratégia de redução (para 10 e 5% de nível de significância) e o mesmo valor sem o procedimento (linha verde). Pode-se facilmente verificar similaridade nestes valores para porcentagens adotadas de até 60% para a variação das categorias de variáveis dependentes (alternativas) no procedimento CART.

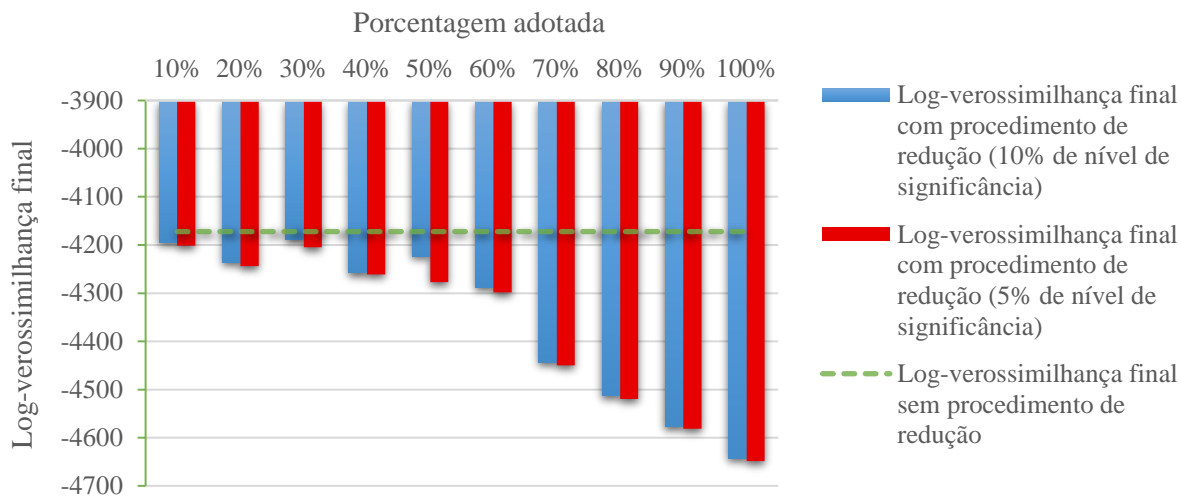


Figura 4.7: Valores de log-verossimilhança finais para cada calibração.

Além disso, os valores do Critério de Informação de Akaike (AIC) também foram usados para comparar os modelos estimados. Como essa medida penaliza a presença de diversas variáveis no modelo, pode-se analisar a eficácia do método proposto em eliminar variáveis potencialmente irrelevantes para o modelo. Também se verificam melhores resultados para porcentagem adotadas de até 60% no procedimento testado.

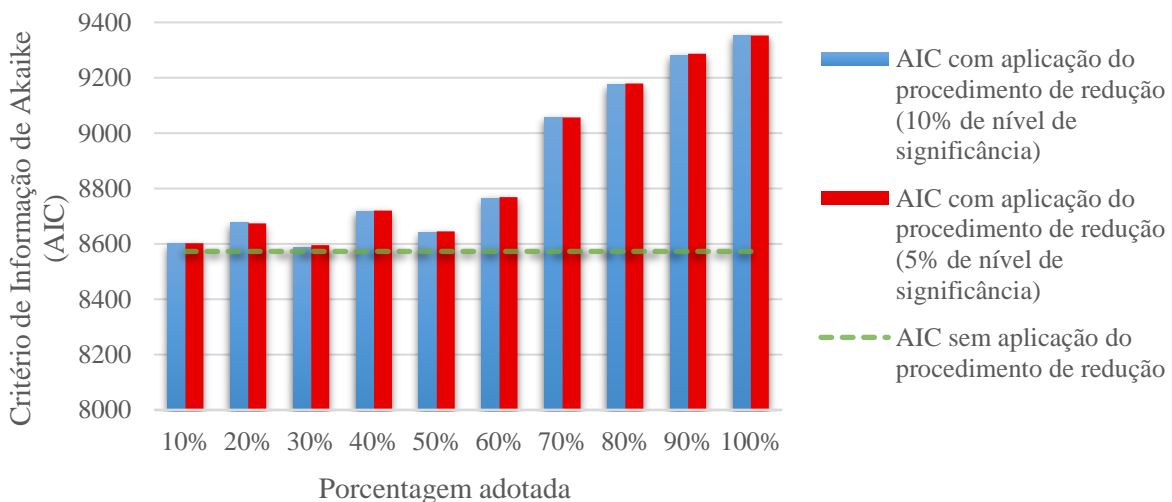


Figura 4.8: Valores do Critério de Informação de Akaike para cada calibração.

A Figura 4.8 ilustra a variação dos valores de AIC para os diferentes níveis percentuais adotados na etapa de análise da variação proporcional de cada alternativa entre os nós filhos gerados pelo algoritmo CART. Observa-se que, ao aplicar o procedimento testado, os valores de AIC flutuam próximos ao valor do procedimento sem redução (indicado pela linha verde)



para percentuais adotados de até 50%. A partir de 60%, no entanto, os valores da AIC mostram fortes tendências de crescimento.

- Taxas de acertos

Aplicando os modelos calibrados para a amostra de validação, comparou-se as escolhas de destinos estimadas com as observadas, obtendo-se, então, as taxas de acertos expostas na Tabela 4.1.

Tabela 4.1: Taxas de acertos para cada modelo, obtidas a partir da amostra de validação.

Calibração	Taxa de acertos (5% de nível de significância)	Taxa de acertos (10% de nível de significância)
Sem procedimento para redução	38,09%	37,70%
Com procedimento para redução (10%)	38,83%	39,04%
Com procedimento para redução (20%)	37,77%	38,09%
Com procedimento para redução (30%)	37,66%	38,72%
Com procedimento para redução (40%)	38,40%	38,30%
Com procedimento para redução (50%)	38,62%	38,51%
Com procedimento para redução (60%)	36,38%	36,60%
Com procedimento para redução (70%)	35,00%	35,00%
Com procedimento para redução (80%)	34,26%	34,26%
Com procedimento para redução (90%)	34,47%	34,47%
Com procedimento para redução (100%)	33,83%	30,11%

Observa-se que, para ambos os níveis de significância adotados, as taxas de acertos com o método oscilam por volta do mesmo intervalo, e próximas das taxas obtidas sem a aplicação do procedimento proposto, para porcentagens adotadas entre 10% e 50%. A partir de 60%, no entanto, as taxas de acertos apresentam uma tendência descendente.

- Distribuição das escolhas de destinos

Analisou-se a distribuição das escolhas de destinos estimadas para a amostra de validação dos modelos executados para diferentes níveis de porcentagem adotados para retirada de parâmetros, de forma a identificar aqueles que mais se assemelham à distribuição das escolhas de destinos do modelo obtido sem a aplicação do procedimento proposto.

Para isso, realizaram-se os testes estatísticos não paramétricos *Mann-Whitney* e *Kolmogorov-Smirnov*, nos quais se considerou-se a hipótese nula como a igualdade entre as frequências de escolha das alternativas obtidas nos modelos em que o procedimento para redução da quantidade de parâmetros foi aplicado e no modelo sem aplicação do artifício proposto. Os resultados para cada teste estão abaixo registrados na Tabela 4.2.

Tabela 4.2. Resultados dos testes estatísticos para cada comparação.

Comparação	<i>Kolmogorov-Smirnov</i> (5% de nível de significância)	<i>Mann-Whitney</i> (5% de nível de significância)	<i>Kolmogorov-Smirnov</i> (10% de nível de significância)	<i>Mann-Whitney</i> (10% de nível de significância)
Sem procedimento x Com procedimento (10%)	Não rejeita $H_0$	Não rejeita $H_0$	Não rejeita $H_0$	Não rejeita $H_0$
Sem procedimento x Com procedimento (20%)	Não rejeita $H_0$	Não rejeita $H_0$	Não rejeita $H_0$	Não rejeita $H_0$
Sem procedimento x Com procedimento (30%)	Não rejeita $H_0$	Não rejeita $H_0$	Rejeita $H_0$	Não rejeita $H_0$
Sem procedimento x Com procedimento (40%)	Não rejeita $H_0$	Não rejeita $H_0$	Rejeita $H_0$	Não rejeita $H_0$
Sem procedimento x Com procedimento (50%)	Não rejeita $H_0$	Não rejeita $H_0$	Rejeita $H_0$	Não rejeita $H_0$
Sem procedimento x Com procedimento (60%)	Rejeita $H_0$	Não rejeita $H_0$	Rejeita $H_0$	Não rejeita $H_0$
Sem procedimento x Com procedimento (70%)	Rejeita $H_0$	Rejeita $H_0$	Rejeita $H_0$	Rejeita $H_0$
Sem procedimento x Com procedimento (80%)	Rejeita $H_0$	Rejeita $H_0$	Rejeita $H_0$	Rejeita $H_0$
Sem procedimento x Com procedimento (90%)	Rejeita $H_0$	Rejeita $H_0$	Rejeita $H_0$	Rejeita $H_0$
Sem procedimento x Com procedimento (100%)	Rejeita $H_0$	Rejeita $H_0$	Rejeita $H_0$	Rejeita $H_0$

Nota-se que, para o caso em que o nível de significância adotado é de 5%, a hipótese nula não é rejeitada para ambos os testes estatísticos aplicados nas análises cujas porcentagens adotadas no procedimento para análise das variações proporcionais de escolha na árvore de decisão eram de até 50%. Para o caso de 60%, a hipótese nula não é rejeitada somente pelo teste de *Mann-Whitney*, e, a partir de 70%, é rejeitada por todos. Esses resultados corroboram os resultados anteriores relacionados às métricas do modelo (rho-quadado, Akaike, final log-likelihood), nos quais, a partir da adoção de 60%, se observa uma diminuição da acurácia dos modelos. Além disso, os resultados confirmam a análise feita sobre a variável “motivo estudo”, ilustrada na Figura 4.2. Por meio dos conhecimentos prévios acerca do município de Santa Maria, constatou-se a importância dessa variável na utilidade do “macrobairro” 6, que é mantida no modelo para porcentagens até 60%.

Os resultados obtidos mostraram que a estratégia proposta pode ser aplicada para a situação analisada de escolha de destinos de viagens urbanas no município de Santa Maria, sem prejuízo à qualidade geral do modelo (medidas de qualidade do ajuste e taxa de acertos) e à distribuição das escolhas dos destinos. Porém, os resultados evidenciam, também, que essa conclusão somente é válida para os casos em que a utilização do procedimento proposto adota porcentagens de até 50% na etapa de análise da variação proporcional de escolha das alternativas entre os nós da árvore de decisão gerada. Vale destacar, no entanto, que essa ressalva diz respeito especificamente à situação testada neste trabalho de distribuição de viagens urbanas para a cidade de Santa Maria. Outras situações, portanto, podem indicar outras porcentagens máximas aceitáveis.

Como o intuito do procedimento é simplificar os modelos para melhorar o custo operacional da calibração dos modelos desagregados com grande quantidade de alternativas e/ou variáveis explicativas, não há como definir um valor máximo aceitável para essa porcentagem com base em comparações como as que aqui foram realizadas, uma vez que elas foram aplicadas para validar o procedimento, além de serem operacionalmente onerosas, o que contrariaria o objetivo estabelecido. Desse modo, sugere-se que, baseado no conhecimento prévio acerca do problema com o qual se deseja trabalhar, seja adotada a porcentagem que o usuário julgue necessária para que o modelo seja suficientemente reduzido para que viabilize sua calibração, desde que essa não implique a retirada de variáveis que, diante do contexto do problema com o qual se trabalha, se mostram intuitivamente relevantes, como a variável “motivo estudo” para o “macrobairro” 6 do modelo trabalhado neste artigo. Sugere-se também uma possível análise exploratória, de sensibilidade, testando-se diferentes porcentagens adotadas e analisando o efeito de cada uma na formulação do modelo.

Assim, embora só possa ser aplicado para variáveis no âmbito do indivíduo, a estratégia aqui exposta para redução da quantidade de parâmetros com uso da técnica de árvore de decisão mostra-se uma alternativa viável para qualquer modelo de escolha discreta que contenha um grande conjunto de alternativas e um grande conjunto de variáveis explicativas associadas ao indivíduo. Essa abordagem deve ser complementada pelo conhecimento do pesquisador ou por informações da literatura sobre a influência das variáveis independentes na previsão do fenômeno em estudo.



## 5 MODELO LOGIT MULTINOMIAL PARA DISTRIBUIÇÃO DE VIAGENS URBANAS: CALIBRAÇÕES CLÁSSICA E POR ALGORITMOS GENÉTICOS

*Este capítulo apresenta os resultados referentes às calibrações clássica e por meio de Algoritmos Genéticos do modelo Logit Multinomial para uma abordagem desagregada de distribuição de viagens urbanas. Inicialmente, aplicou-se a estratégia proposta anteriormente, para redução de parâmetros e, em seguida, determinou-se a formulação a ser adotada nos modelos. Por fim, os resultados das calibrações finais e os obtidos por meio da amostra de validação são apresentados e comparados.*

### 5.1 CALIBRAÇÃO CLÁSSICA

Embora o enfoque deste capítulo seja a aplicação de Algoritmos Genéticos para calibração do modelo Logit Multinomial, como forma de superar as limitações apresentadas no Capítulo 2, a calibração clássica ainda é necessária, uma vez que serve como ferramenta para comparação com a calibração proposta por Algoritmos Genéticos e validação metodológica.

Assim, calibrou-se um modelo referente à unidade de área “bairros”, portanto mais desagregado do que os “macrobairros” referentes à calibração realizada para a proposição de uma estratégia para reduzir a quantidade de parâmetros estimados, cujos resultados foram expostos no Capítulo 4 deste trabalho. Logo, os 41 bairros constituintes do município de Santa Maria e ilustrados na Figura 5.1 compõem os destinos possíveis.

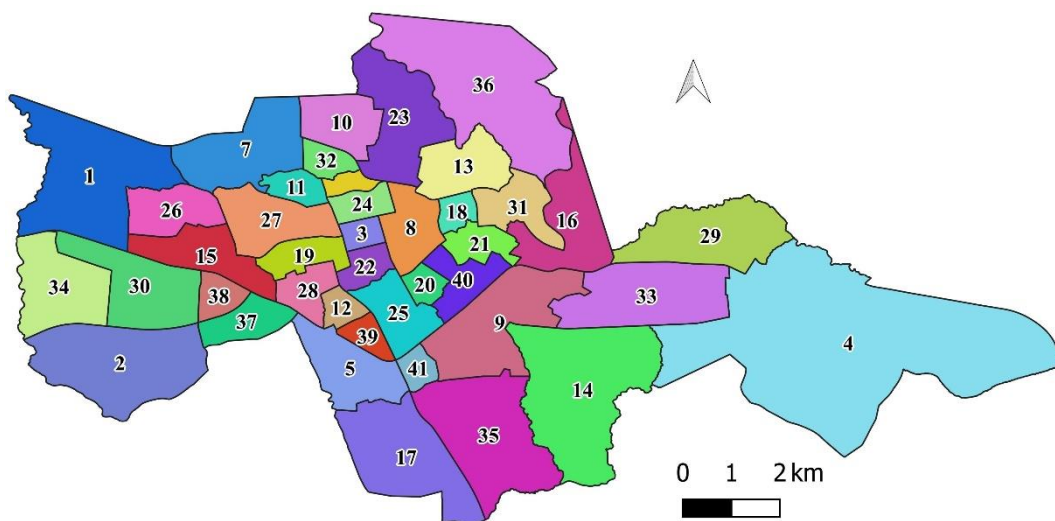


Figura 5.1: Disposição dos bairros de Santa Maria (RS).

Devido à grande quantidade de alternativas e variáveis envolvidas na modelagem por “bairros”, aplicou-se a estratégia proposta, de forma a viabilizar a calibração. Vale ressaltar que os dados utilizados correspondem a 70% do total de observações (amostra de calibração). Os 30% restantes foram utilizados para validação do modelo calibrado.

Nas subseções a seguir, detalha-se a aplicação da estratégia proposta e o tipo de formulação adotada (genérica ou específica) e expõe-se o resultado da calibração final e o obtido por meio da amostra de validação.

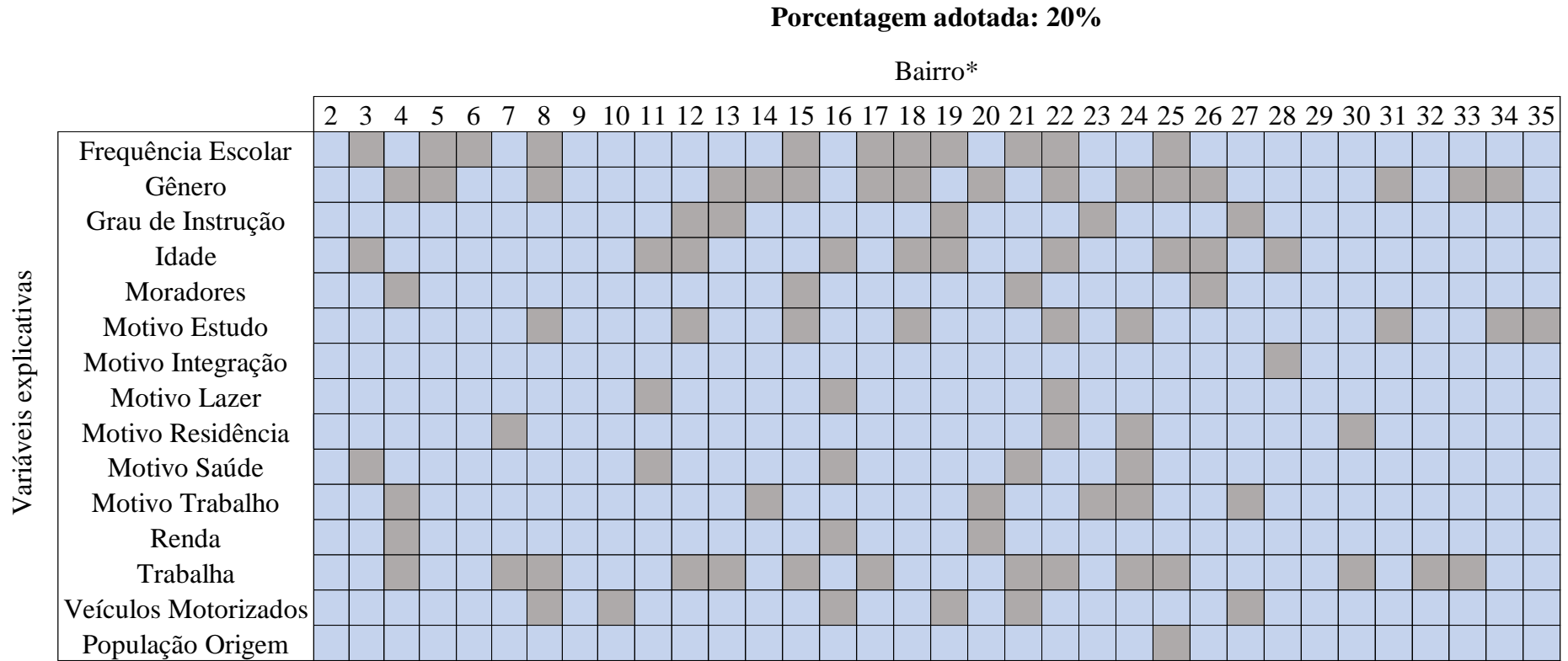
### 5.1.1 Aplicação da estratégia proposta

Tendo em vista as evidências associadas à eficácia da estratégia proposta para redução da quantidade de parâmetros a estimar em modelos de escolha discreta, sobretudo no caso de viagens urbanas para a cidade de Santa Maria (RS), e as dificuldades operacionais envolvendo calibrações com muitas alternativas, como é o caso da calibração por “bairros”, seguiu-se o passo a passo da estratégia, descrita anteriormente.

Para isso, adotou-se a porcentagem de 20%, ou seja 20% como limite de variação proporcional de categorias da variável dependente (alternativas) entre os nós filhos de cada árvore de decisão, associada a cada uma das variáveis independentes relativas ao indivíduo. Essa porcentagem foi escolhida devido ao fato de proporcionar uma redução significativa na quantidade de parâmetros, viabilizando sua calibração no aplicativo utilizado (PandasBiogeme), e pelos resultados analisados na proposição da estratégia indicarem que esta poderia ser aplicada para porcentagens adotadas de até 50% para o mesmo banco de dados.

Vale ressaltar que a estratégia proposta é aplicável somente para as variáveis relativas ao indivíduo. Assim, somente quinze variáveis (descritas na Figura 5.2) foram utilizadas. Além disso, destaca-se que a função utilidade relativa ao bairro 1 (Agro-Industrial) foi utilizada como referência para a calibração do modelo, logo não há necessidade de utilizar o procedimento para analisar as variáveis para essa alternativa.

Ressalta-se, ainda, que 6 dos 41 bairros analisados não apresentaram registros no banco de dados (bairros 36, 37, 38, 39, 40 e 41), sendo, portanto, desconsiderados nas calibrações dos modelos. Logo, a quantidade de parâmetros possíveis de serem retiradas resultou em 510 (15 variáveis x 34 alternativas). Aplicado o procedimento proposto, identificou-se a possível retirada de 98 parâmetros, conforme indicado na Figura 5.2, o que representa cerca de 19% do total de parâmetros analisados.



\* Bairro 1 não foi incluído no procedimento, pois foi utilizado como referência para a calibração das funções utilidade. Bairros 36, 37, 38, 39, 40 e 41 não foram incluídos, pois não há viagens que se destinem a essas localidades no banco de dados.

Figura 5.2: Variáveis associadas a parâmetros significativos (em azul) e não significativos (em cinza) das funções utilidade de cada bairro, a partir da porcentagem de variação adotada de 20%.

### 5.1.2 Formulação das funções utilidade

A calibração do modelo Logit Multinomial foi realizada com o banco de dados contendo 70% das observações, as variáveis do indivíduo, apontadas na Figura 5.2, e as duas variáveis associadas às alternativas: “Distância” e “Pontos de Ônibus”.

Assim, fez-se necessário determinar a formulação (genérica ou específica) para o caso das duas variáveis relativas às alternativas. Para determinar qual formulação é a mais adequada para as variáveis “Distância” e “Pontos de Ônibus”, realizou-se novamente o teste da razão de verossimilhança (Bierlaire et al., 2016). A hipótese nula assume que os coeficientes dessas variáveis são iguais para todas as alternativas.

Para a calibração, assume-se nível de significância de 5% e 34 graus de liberdade (obtido pela diferença entre a quantidade de parâmetros estimados entre os modelos específico e genérico). Assim, calibrados os modelos genérico e específico para a variável “Distância”, tem-se, para a Expressão (5.1):

$$-2 \cdot (-5325,883 - (-5238,509)) = 174,748 > 51,965 \quad (5.1)$$

Logo, a hipótese nula foi rejeitada, o que significa que a formulação específica é a mais adequada para a variável “Distância”.

Para a variável “Pontos de Ônibus”, por sua vez, tem-se:

$$-2 \cdot (-5325,883 - (-5325,911)) = 0,056 < 51,965 \quad (5.2)$$

Logo, a hipótese nula foi aceita, o que significa que a formulação genérica é a mais adequada para a variável “Pontos de Ônibus”.

Assim, a configuração geral das funções utilidade para a calibração por “bairros” ocorreu da forma:

$$\begin{aligned} V_{ij} = & a_i + b_i \cdot \text{FrequênciaEscolar}_j + c_i \cdot \text{Gênero}_j + d_i \cdot \text{GraudeInstrução}_j + e_i \\ & \cdot \text{Idade}_j + f_i \cdot \text{Moradores}_j + g_i \cdot \text{MotivoEstudo}_j + h_i \\ & \cdot \text{MotivoIntegração}_j + i_i \cdot \text{MotivoLazer}_j + j_i \cdot \text{MotivoResidência}_j \\ & + k_i \cdot \text{MotivoSaúde}_j + l_i \cdot \text{MotivoTrabalho}_j + m_i \cdot \text{Renda}_j + n_i \\ & \cdot \text{Trabalha}_j + o_i \cdot \text{VeículosMotorizados}_j + p_i \cdot \text{PopulaçãoOrigem}_j \\ & + q_i \cdot \text{Distância}_{ij} + r \cdot \text{PontosdeÔnibus}_{ij} \end{aligned} \quad (5.3)$$

Onde:



$V_{ij}$ : utilidade da alternativa  $i$  para o indivíduo  $j$ ;

$a_i, b_i, c_i, \dots, q_i$ : coeficientes a serem estimados para a alternativa  $i$ ;

$r$ : coeficiente estimado para todas as alternativas, referente à variável “Pontos de Ônibus”.

### 5.1.3 Resultados da calibração

Definida a formulação das funções utilidade e excluídos os parâmetros com o auxílio do algoritmo CART, indicados na Figura 5.2, a calibração clássica do modelo Logit Multinomial para análise da distribuição de viagens urbanas na cidade de Santa Maria (RS), considerando os bairros como opções de destinos, pode ser realizada.

Diversas calibrações foram realizadas, de modo que, para cada uma delas, excluiu-se os parâmetros que resultaram não significativos na calibração anterior, ao considerar um nível de significância de 95% (p-valor associado ao teste t maior do que 0,05). Assim, repetiu-se esse procedimento até que uma calibração contivesse todos os parâmetros significativos.

A Figura 5.3 contém a indicação dos parâmetros relacionados aos indivíduos retirados (em cinza) e mantidos (em azul) de cada função utilidade na calibração final. Além disso, as funções utilidades obtidas estão apresentadas na Tabela 5.1. Em resumo, a calibração final totalizou 200 parâmetros estimados, obtendo-se  $\rho^2$  com valor 0,308 e  $\rho^2$  ajustado com valor 0,282, ou seja, dentro do intervalo entre 0,2 e 0,4 considerado bom para esse tipo de modelo (McFadden, 1978b), além de Critério de Informação Akaike (AIC) de 11.208,95 e verossimilhança final igual a -5.404,475. Vale ressaltar que a informação contida na Figura 5.3 é referente aos valores de coeficientes estimados apenas para as variáveis dos indivíduos. Essa informação é complementada pela Tabela 5.1, a qual traz também os valores dos parâmetros associados às variáveis relativas às alternativas.

		Calibração Final																																				
		Bairro*																																				
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35			
Variáveis explicativas	Frequência Escolar								-1,56			1,43		7,50									1,08							-6,32		-0,64						
	Gênero																																					
	Grau de Instrução	5,23			-1,59			1,70	-2,06						-0,93						1,55	2,02				-2,68											-1,53	
	Idade				-2,18	2,96		1,20	2,05				2,56												2,00													
	Moradores				-3,05						2,88	2,70	-3,87	3,47				5,17	3,34	1,66					1,82					5,19		1,46	4,01					
	Motivo Estudo	-3,41	1,21	1,67	-2,23	-1,42			6,82					-5,40	-5,91	-5,96	-3,88																					-3,10
	Motivo Integração	-3,80	-6,27	-7,53	-1,75		-5,87	0,70				-5,59	-5,18		-4,14	-6,72		-2,78		-6,09					-5,84	-5,99	-5,84	-1,85	-5,46		-2,88	-6,32	-6,06		-5,24	-6,21	-5,54	
	Motivo Lazer	2,05	1,33		6,60	-2,07	-5,98		-2,29	-5,69								2,23					1,02	-6,06	-6,50	6,09				2,48							1,33	
	Motivo Residência	-6,08			7,48	6,91		-1,06	4,01	1,80		-7,36	1,26	-6,33	1,98	3,24				0,85	-0,90		1,00		0,86	7,35	1,67		-5,37		1,12	2,07		1,64				
	Motivo Saúde	-2,98			-1,25	-1,53	-5,58		-1,66			-4,73					-3,63							-5,17	-5,56	6,60		-4,87	-3,75		-6,55	-5,01		-5,34				
	Motivo Trabalho				4,33	4,37		-0,59	5,59							-1,24				-1,69		-0,98				4,72					-0,93	-1,54	-7,34	-1,22				
	Renda					-1,36	-2,03			-1,98		-3,52	-0,86		-1,86				1,41							-1,69										-3,21		
	Trabalha				0,94																																	
	Veículos Motorizados												2,83		3,64	1,27										-2,94									-3,23			-11,0
População Origem			1,04				-5,12							-2,39	1,87			-0,93								1,88										2,82		

\* Bairro 1 não foi incluído no procedimento, pois foi utilizado como referência para a calibração das funções utilidade. Bairros 36, 37, 38, 39, 40 e 41 não foram incluídos, pois não há viagens que se destinem a essas localidades no banco de dados.

Figura 5.3: Variáveis associadas a parâmetros significativos (em azul) e não significativos (em cinza) das funções utilidade de cada bairro na calibração final.

Tabela 5.1: Funções utilidade para cada bairro de destino.

$V_{1j} = -3,47 \cdot \text{Distância}_{1j} + 1,67 \cdot \text{PontosdeOnibus}_{1j}$
$V_{2j} = -11,2 + 5,23 \cdot \text{GraudeInstrução}_j + 6,21 \cdot \text{Trabalha}_j - 3,41 \cdot \text{MotivoEstudo}_j - 3,8$ $\cdot \text{MotivoIntegração}_j + 2,05 \cdot \text{MotivoLazer}_j - 6,08 \cdot \text{MotivoResidência}_j - 2,98$ $\cdot \text{MotivoSaúde}_j - 2,30 \cdot \text{Distância}_{2j} + 1,67 \cdot \text{PontosdeOnibus}_{2j}$
$V_{3j} = 1,21 \cdot \text{MotivoEstudo}_j - 6,27 \cdot \text{MotivoIntegração}_j + 1,33 \cdot \text{MotivoLazer}_j - 4,84 \cdot \text{Distância}_{3j}$ $+ 1,67 \cdot \text{PontosdeOnibus}_{3j}$
$V_{4j} = 1,67 \cdot \text{MotivoEstudo}_j - 7,53 \cdot \text{MotivoIntegração}_j + 1,04 \cdot \text{PopulaçãoOrigem}_j - 1,81$ $\cdot \text{Distância}_{4j} + 1,67 \cdot \text{PontosdeOnibus}_{4j}$
$V_{5j} = -3,27 - 1,59 \cdot \text{GraudeInstrução}_j - 2,18 \cdot \text{Idade}_j + 0,94 \cdot \text{Trabalha}_j - 3,05 \cdot \text{Moradores}_j$ $- 2,23 \cdot \text{MotivoEstudo}_j - 1,75 \cdot \text{MotivoIntegração}_j + 6,60 \cdot \text{MotivoLazer}_j + 7,48$ $\cdot \text{MotivoResidência}_j - 1,25 \cdot \text{MotivoSaúde}_j + 4,33 \cdot \text{MotivoTrabalho}_j - 7,79$ $\cdot \text{Distância}_{5j} + 1,67 \cdot \text{PontosdeOnibus}_{5j}$
$V_{6j} = -7,86 + 2,96 \cdot \text{Idade}_j + 4,37 \cdot \text{MotivoTrabalho}_j - 1,42 \cdot \text{MotivoEstudo}_j - 2,07 \cdot \text{MotivoLazer}_j$ $+ 6,91 \cdot \text{MotivoResidência}_j - 1,53 \cdot \text{MotivoSaúde}_j - 1,36 \cdot \text{Renda}_j + 1,67$ $\cdot \text{PontosdeOnibus}_{6j}$
$V_{7j} = 3,1 - 5,87 \cdot \text{MotivoIntegração}_j - 5,98 \cdot \text{MotivoLazer}_j - 5,58 \cdot \text{MotivoSaúde}_j - 5,12$ $\cdot \text{PopulaçãoOrigem}_j - 2,03 \cdot \text{Renda}_j - 7,05 \cdot \text{Distância}_{7j} + 1,67 \cdot \text{PontosdeOnibus}_{7j}$
$V_{8j} = 1,70 \cdot \text{GraudeInstrução}_j + 1,20 \cdot \text{Idade}_j + 0,70 \cdot \text{MotivoIntegração}_j - 1,06$ $\cdot \text{MotivoResidência}_j - 0,59 \cdot \text{MotivoTrabalho}_j - 1,77 \cdot \text{Distância}_{8j} + 1,67$ $\cdot \text{PontosdeOnibus}_{8j}$
$V_{9j} = -6,70 - 1,56 \cdot \text{FrequênciaEscolar}_j + 6,82 \cdot \text{MotivoEstudo}_j + 5,59 \cdot \text{MotivoTrabalho}_j + 4,01$ $\cdot \text{MotivoResidência}_j - 1,66 \cdot \text{MotivoSaúde}_j - 2,29 \cdot \text{MotivoLazer}_j + 1,67$ $\cdot \text{PontosdeOnibus}_{9j}$
$V_{10j} = -2,06 \cdot \text{GraudeInstrução}_j + 2,05 \cdot \text{Idade}_j + 2,88 \cdot \text{Moradores}_j - 5,69 \cdot \text{MotivoLazer}_j + 1,80$ $\cdot \text{MotivoResidência}_j - 1,98 \cdot \text{Renda}_j - 2,91 \cdot \text{Distância}_{10j} + 1,67$ $\cdot \text{PontosdeOnibus}_{10j}$
$V_{11j} = 2,70 \cdot \text{Moradores}_j - 5,59 \cdot \text{MotivoIntegração}_j - 5,34 \cdot \text{Distância}_{11j} + 1,67$ $\cdot \text{PontosdeOnibus}_{11j}$
$V_{12j} = 1,43 \cdot \text{FrequênciaEscolar}_j - 3,87 \cdot \text{Moradores}_j + 2,83 \cdot \text{VeículosMotorizados}_j - 5,18$ $\cdot \text{MotivoIntegração}_j - 7,36 \cdot \text{MotivoResidência}_j - 4,73 \cdot \text{MotivoSaúde}_j - 3,52$ $\cdot \text{Renda}_j + 1,67 \cdot \text{PontosdeOnibus}_{12j}$
$V_{13j} = -2,41 + 2,56 \cdot \text{Idade}_j + 3,47 \cdot \text{Moradores}_j - 5,40 \cdot \text{MotivoEstudo}_j + 1,26$ $\cdot \text{MotivoResidência}_j - 0,86 \cdot \text{Renda}_j - 2,97 \cdot \text{Distância}_{13j} + 1,67$ $\cdot \text{PontosdeOnibus}_{13j}$
$V_{14j} = -8,00 + 7,50 \cdot \text{FrequênciaEscolar}_j + 3,64 \cdot \text{VeículosMotorizados}_j - 5,91 \cdot \text{MotivoEstudo}_j$ $- 4,14 \cdot \text{MotivoIntegração}_j - 6,33 \cdot \text{MotivoResidência}_j - 2,39 \cdot \text{PopulaçãoOrigem}_j$ $+ 1,67 \cdot \text{PontosdeOnibus}_{14j}$
$V_{15j} = 1,35 - 0,934 \cdot \text{GraudeInstrução}_j + 1,27 \cdot \text{VeículosMotorizados}_j - 6,72 \cdot \text{MotivoIntegração}_j$ $+ 1,98 \cdot \text{MotivoResidência}_j - 1,24 \cdot \text{MotivoTrabalho}_j + 1,87 \cdot \text{PopulaçãoOrigem}_j$ $- 1,86 \cdot \text{Renda}_j - 5,78 \cdot \text{Distância}_{15j} + 1,67 \cdot \text{PontosdeOnibus}_{15j}$
$V_{16j} = 0,962 - 5,96 \cdot \text{MotivoEstudo}_j - 7,17 \cdot \text{Distância}_{16j} + 1,67 \cdot \text{PontosdeOnibus}_{16j}$
$V_{17j} = -3,48 + 5,17 \cdot \text{Moradores}_j - 3,88 \cdot \text{MotivoEstudo}_j - 2,78 \cdot \text{MotivoIntegração}_j + 2,23$ $\cdot \text{MotivoLazer}_j + 3,24 \cdot \text{MotivoResidência}_j - 3,63 \cdot \text{MotivoSaúde}_j - 4,44$ $\cdot \text{Distância}_{17j} + 1,67 \cdot \text{PontosdeOnibus}_{17j}$
$V_{18j} = 1,54 - 0,93 \cdot \text{PopulaçãoOrigem}_j + 1,41 \cdot \text{Renda}_j - 0,60 \cdot \text{Trabalha}_j - 4,72 \cdot \text{Distância}_{18j}$ $+ 1,67 \cdot \text{PontosdeOnibus}_{18j}$
$V_{19j} = 3,34 \cdot \text{Moradores}_j - 6,09 \cdot \text{MotivoIntegração}_j - 1,69 \cdot \text{MotivoTrabalho}_j - 8,77 \cdot \text{Distância}_{19j}$ $+ 1,67 \cdot \text{PontosdeOnibus}_{19j}$
$V_{20j} = 1,55 \cdot \text{GraudeInstrução}_j + 1,66 \cdot \text{Moradores}_j + 0,85 \cdot \text{MotivoResidência}_j - 5,33$ $\cdot \text{Distância}_{20j} + 1,67 \cdot \text{PontosdeOnibus}_{20j}$

---


$$V_{21j} = 1,02 \cdot \text{MotivoLazer}_j - 0,90 \cdot \text{MotivoResidência}_j - 3,51 \cdot \text{Distância}_{21j} + 1,67$$

$$\cdot \text{PontosdeOnibus}_{21j}$$


---


$$V_{22j} = 2,02 \cdot \text{GraudeInstrução}_j - 0,98 \cdot \text{MotivoTrabalho}_j - 4,24 \cdot \text{Distância}_{22j} + 1,67$$

$$\cdot \text{PontosdeOnibus}_{22j}$$


---


$$V_{23j} = 1,08 \cdot \text{FrequênciaEscolar}_j - 5,84 \cdot \text{MotivoIntegração}_j - 6,06 \cdot \text{MotivoLazer}_j + 1,00$$

$$\cdot \text{MotivoResidência}_j - 5,17 \cdot \text{MotivoSaúde}_j - 6,89 \cdot \text{Distância}_{23j} + 1,67$$

$$\cdot \text{PontosdeOnibus}_{23j}$$


---


$$V_{24j} = 2,00 \cdot \text{Idade}_j + 1,82 \cdot \text{Moradores}_j - 2,94 \cdot \text{VeículosMotorizados}_j - 5,99 \cdot \text{MotivoIntegração}_j$$

$$- 8,62 \cdot \text{Distância}_{24j} + 1,67 \cdot \text{PontosdeOnibus}_{24j}$$


---


$$V_{25j} = -5,84 \cdot \text{MotivoIntegração}_j - 6,50 \cdot \text{MotivoLazer}_j + 0,86 \cdot \text{MotivoResidência}_j - 5,56$$

$$\cdot \text{MotivoSaúde}_j - 5,80 \cdot \text{Distância}_{25j} + 1,67 \cdot \text{PontosdeOnibus}_{25j}$$


---


$$V_{26j} = -3,60 - 2,68 \cdot \text{GraudeInstrução}_j - 3,10 \cdot \text{MotivoEstudo}_j - 1,85 \cdot \text{MotivoIntegração}_j + 6,09$$

$$\cdot \text{MotivoLazer}_j + 7,35 \cdot \text{MotivoResidência}_j + 6,60 \cdot \text{MotivoSaúde}_j + 4,72$$

$$\cdot \text{MotivoTrabalho}_j + 1,88 \cdot \text{PopulaçãoOrigem}_j - 1,69 \cdot \text{Renda}_j - 4,35 \cdot \text{Distância}_{26j}$$

$$+ 1,67 \cdot \text{PontosdeOnibus}_{26j}$$


---


$$V_{27j} = -5,46 \cdot \text{MotivoIntegração}_j + 1,67 \cdot \text{MotivoResidência}_j - 4,95 \cdot \text{Distância}_{27j} + 1,67$$

$$\cdot \text{PontosdeOnibus}_{27j}$$


---


$$V_{28j} = -4,87 \cdot \text{MotivoSaúde}_j - 8,14 \cdot \text{Distância}_{28j} + 1,67 \cdot \text{PontosdeOnibus}_{28j}$$


---


$$V_{29j} = -4,12 - 6,32 \cdot \text{FrequênciaEscolar}_j + 5,19 \cdot \text{Moradores}_j - 2,88 \cdot \text{MotivoIntegração}_j + 2,48$$

$$\cdot \text{MotivoLazer}_j - 5,37 \cdot \text{MotivoResidência}_j - 3,75 \cdot \text{MotivoSaúde}_j + 1,67$$

$$\cdot \text{PontosdeOnibus}_{29j}$$


---


$$V_{30j} = 0,69 - 6,32 \cdot \text{MotivoIntegração}_j - 0,93 \cdot \text{MotivoTrabalho}_j - 3,34 \cdot \text{Distância}_{30j} + 1,67$$

$$\cdot \text{PontosdeOnibus}_{30j}$$


---


$$V_{31j} = 2,22 - 0,64 \cdot \text{FrequênciaEscolar}_j + 1,46 \cdot \text{Moradores}_j - 3,23 \cdot \text{VeículosMotorizados}_j - 6,06$$

$$\cdot \text{MotivoIntegração}_j + 1,12 \cdot \text{MotivoResidência}_j - 6,55 \cdot \text{MotivoSaúde}_j - 1,54$$

$$\cdot \text{MotivoTrabalho}_j - 12,10 \cdot \text{Distância}_{31j} + 1,67 \cdot \text{PontosdeOnibus}_{31j}$$


---


$$V_{32j} = 4,01 \cdot \text{Moradores}_j + 2,07 \cdot \text{MotivoResidência}_j - 3,21 \cdot \text{Renda}_j - 5,65 \cdot \text{Distância}_{32j} + 1,67$$

$$\cdot \text{PontosdeOnibus}_{32j}$$


---


$$V_{33j} = -5,24 \cdot \text{MotivoIntegração}_j - 5,01 \cdot \text{MotivoSaúde}_j - 7,34 \cdot \text{MotivoTrabalho}_j - 2,92$$

$$\cdot \text{Distância}_{33j} + 1,67 \cdot \text{PontosdeOnibus}_{33j}$$


---


$$V_{34j} = 1,35 - 1,53 \cdot \text{GraudeInstrução}_j - 6,21 \cdot \text{MotivoIntegração}_j + 1,64 \cdot \text{MotivoResidência}_j$$

$$- 1,22 \cdot \text{MotivoTrabalho}_j + 2,82 \cdot \text{PopulaçãoOrigem}_j - 5,75 \cdot \text{Distância}_{34j} + 1,67$$

$$\cdot \text{PontosdeOnibus}_{34j}$$


---


$$V_{35j} = -11,00 \cdot \text{VeículosMotorizados}_j - 5,54 \cdot \text{MotivoIntegração}_j + 1,33 \cdot \text{MotivoLazer}_j - 5,34$$

$$\cdot \text{MotivoSaúde}_j + 1,67 \cdot \text{PontosdeOnibus}_{35j}$$


---

Diferentemente da aplicação de modelos Logit para a etapa de escolha modal, na qual a compreensão da influência das variáveis sobre as utilidades dos modos de transporte ocorre facilmente de maneira intuitiva e baseada na literatura, a análise qualitativa dos parâmetros nas funções utilidade no caso da escolha dos destinos não é facilmente verificada, haja vista as diferentes características e equipamentos presentes em cada alternativa.

Nesse sentido, características marcantes e a presença de estabelecimentos tidos como referência auxiliam na compreensão dos fatores que influenciam a utilidade de determinado destino de viagem. Isso posto, ao analisar os resultados da Tabela 5.1, pode-se compreender determinados fatores que influenciam a escolha de alguns destinos pelos indivíduos. Na equação representativa do bairro 4 (Camobi), por exemplo, observa-se que o fato de a viagem

ser por motivo estudo aumenta a utilidade desse bairro, e, conseqüentemente, a probabilidade de o indivíduo escolhê-la como destino. O contrário ocorre se a viagem for motivada por integração. Esse resultado é facilmente compreendido ao constatar no bairro 4 a presença da Universidade Federal de Santa Maria (UFSM), importante polo atrativo para estudantes no município. Por outro lado, não há nessa região polos atrativos significativos para integração, o que justificaria o efeito negativo dessa motivação ao destino em questão.

A equação representativa do bairro 21 (Nossa Senhora de Lourdes), por sua vez, aponta que viagens motivadas pelo lazer aumentam a utilidade desse bairro, o que pode ser explicado pela presença do *Royal Plaza Shopping*, importante e conhecido polo atrativo para deslocamentos com viés comercial ou recreativo.

Com relação à “idade”, percebe-se a influência positiva dessa variável para a escolha dos bairros “Carolina”, “Centro”, “Chácara das Flores”, “Itararé” e “Nossa Senhora do Rosário”, representados pelas utilidades  $V_6$ ,  $V_8$ ,  $V_{10}$ ,  $V_{13}$  e  $V_{24}$ , respectivamente. Isso pode ser explicado pelo fato de esses bairros constituírem a zona central da cidade, e que, segundo o Plano Diretor de Mobilidade Urbana (IPLAN, 2013), o grupo de idade 0 a 14 anos é o menor, sendo a faixa entre 25 e 64 anos a mais representativa dessa região. O oposto, no entanto, ocorre com o Bairro “Urlândia”, representado pela utilidade  $V_5$ , na qual a variável “idade” apresenta influência negativa. Para o censo demográfico de 2010, esse bairro era o 37º em percentual de população na terceira idade e o 11º em percentual de população na menoridade.

Além disso, a variável “motivo residência” para os bairros “Juscelino Kubitschek” e “Nova Santa Marta”, representados pelas funções  $V_{15}$  e  $V_{26}$ , respectivamente, relaciona-se positivamente com a utilidade desses locais, o que pode ser explicado por serem alguns dos bairros com maior população residente do município (IPLAN, 2013).

#### 5.1.4 Validação

O modelo calibrado foi aplicado para a amostra de validação (com 30% das observações), obtendo-se as escolhas dos destinos estimadas. Comparando-as com as escolhas observadas, alcançou-se taxa de acertos global de 31,91% para a calibração por bairros.

## 5.2 CALIBRAÇÃO POR ALGORITMOS GENÉTICOS

A calibração por Algoritmos Genéticos ocorreu a partir de um programa executável originado de código elaborado por meio do aplicativo Visual Studio 2017. A linguagem

utilizada foi a C# e o código pode ser visualizado através da biblioteca virtual disponibilizada no link [https://github.com/marinaurano/Algoritmo\\_Genetico-escolha\\_de\\_destinos\\_urbanos](https://github.com/marinaurano/Algoritmo_Genetico-escolha_de_destinos_urbanos). O banco de dados e as informações sobre os limites dos parâmetros para cada variável compuseram os arquivos de entrada (com extensão *.txt*) para execução do programa e estão na biblioteca mencionada.

O processo de otimização realizado por Algoritmo Genéticos utiliza uma série de parâmetros para realizar as etapas referentes ao cálculo das aptidões, seleção, cruzamento e mutação. Na literatura, no entanto, não há resultados conclusivos sobre os melhores valores desses parâmetros para todas as aplicações. Alguns autores utilizam valores reportados em outros trabalhos, outros definem os seus próprios valores com base em experimentação (Rodrigues et al., 2004). Assim, os valores considerados neste trabalho, expostos na Tabela 5.2, basearam-se nos trabalhos que também utilizaram Algoritmos Genéticos para modelos de distribuição de viagens (Souza et al., 2018; Sung e Rakha, 2009). Vale ressaltar que o cruzamento aplicado para o algoritmo elaborado foi do tipo BLX- $\alpha$  (cruzamento *blend*), por ser mais adequado para formatos reais.

Tabela 5.2: Parâmetros considerados para a calibração por Algoritmos Genéticos.

População	100
Taxa de Mutação	0,2
Fração de crossover	0,5 (em 100% dos indivíduos)
Elitismo	1

Por se tratar de um modelo Logit Multinomial, portanto estimado por máxima verossimilhança, a função a ser otimizada pelo algoritmo (função objetivo) compreendeu o valor do somatório dos logaritmos das probabilidades de escolha de cada destino efetivamente escolhido. No algoritmo, conforme pode ser observado no retrato do código ilustrado na Figura 5.4, optou-se por determinar a função objetivo como o valor em módulo dessa expressão. Assim, o menor valor da função objetivo representa a solução com a máxima verossimilhança.

```

232     somaprob = 0; //somaprob = somatório dos logaritmos das probabilidades das alternativas escolhidas
233     for (var i = 1; i <= 2196; i++)
234     {
235         for (var j = 1; j <= 35; j++) { v3[j] = mu[i, j]; }
236         soma = 0;
237         for (var j = 1; j <= 35; j++) { soma = soma + Math.Exp(v3[j]); }
238         //Exponencial da utilidade de cada alternativa sobre o somatório das exponenciais de todas as alternativas
239         vprob[i] = Math.Exp(mu[i, vcod[i]]) / soma;
240         if (vprob[i] < 0.0000000001) { vprob[i] = 0.0000000001; }
241         somaprob = somaprob + Math.Log(vprob[i]);
242     }
243
244     return Math.Abs(somaprob); //Retorna o valor absoluto do somaprob. Observe que somaprob é um valor negativo,
245     //pois os logaritmos das probabilidades são valores negativos.
246     //O maior valor de somaprob (máxima verossimilhança) ocorre com o menor valor absoluto de somaprob,
247     //que corresponde à função objetivo do AG.

```

Figura 5.4: Código com a função objetivo do Algoritmo Genético.

Sobre os critérios de parada, para acompanhar o processo de otimização da função objetivo, optou-se por não estabelecer uma quantidade específica de gerações. O fim do processo de calibração dependeu da finalização do programa pelo usuário. Apesar disso, uma estratégia para diversificar as soluções foi adotada: caso a diferença entre os valores mínimo e máximo da função objetivo de uma população se mantenha menor do que 1 durante 50 gerações, a população será resetada e gerada de forma aleatória, de forma que somente a melhor solução se mantenha. No geral, a cada execução do algoritmo, o programa foi encerrado após cerca de cem mil gerações.

A Figura 5.5 a seguir apresenta a evolução do valor da função objetivo para uma das execuções realizadas. O comportamento da curva mostrou-se similar para todas as execuções e ilustram a melhoria expressiva no valor da função objetivo, sobretudo nas primeiras gerações, o qual obteve certa estabilidade após aproximadamente 50.000 gerações.

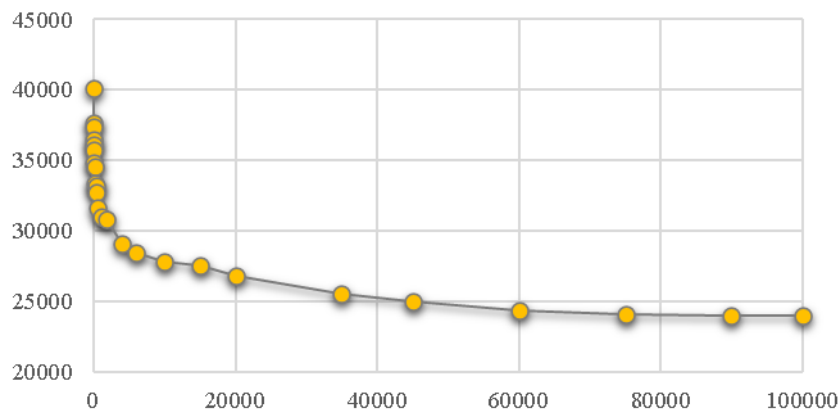


Figura 5.5: Valores da função objetivo ao longo das gerações.

De forma a analisar as soluções obtidas considerando diferentes configurações para a quantidade de parâmetros e os limites de valores a eles estabelecidos, três cenários foram analisados: livre, semi-restrito e restrito.

### 5.2.1 Livre

No cenário livre, desconsidera-se qualquer informação obtida com a calibração clássica realizada previamente. Além disso, não se aplicou procedimento para redução na quantidade de parâmetros, diferentemente do caso da calibração clássica.

Assim, um total de 580 parâmetros foram calibrados pelo Algoritmo Genético, sem qualquer grande limitação nos intervalos de valores do arquivo de entrada. Adotou-se o intervalo entre -50 e 50 para todos os parâmetros.

Como os Algoritmos Genéticos não fornecem a significância dos parâmetros estimados, não foi possível identificar parâmetros não significativos em cada função utilidade. Assim, os modelos finais continham os 580 parâmetros.

Ao todo, o algoritmo no cenário livre foi executado 63 vezes. Cada execução correspondia a 100 valores diferentes de função objetivo. Assim, ao final das simulações, havia 6300 valores diferentes de funções objetivos. Com relação à aplicação dos modelos gerados à amostra de validação, nota-se grande disparidade nos dados, com taxa de acertos variando entre aproximadamente 4% e 19% e média de 10,8%. As funções utilidade resultantes da calibração que resultou menor valor da função objetivo (melhor resultado global) encontram-se no Apêndice E, ao final deste trabalho.

As distribuições dos valores das funções objetivo e das taxas de acertos para a melhor solução de cada rodada estão a seguir ilustradas nas Figuras 5.6 e 5.7, respectivamente.



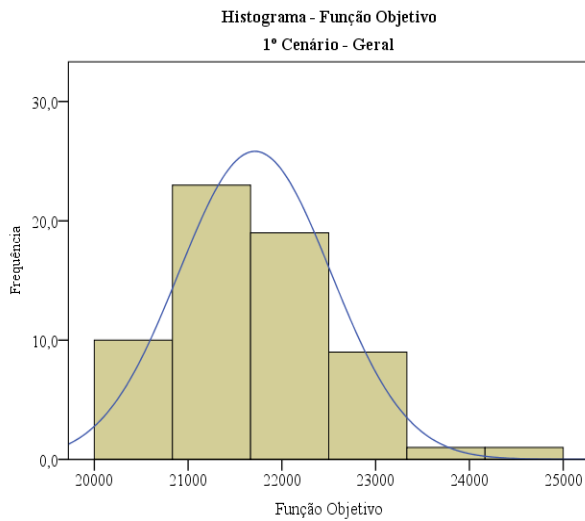


Figura 5.6: Histograma dos valores de função objetivo para as melhores soluções no cenário livre.

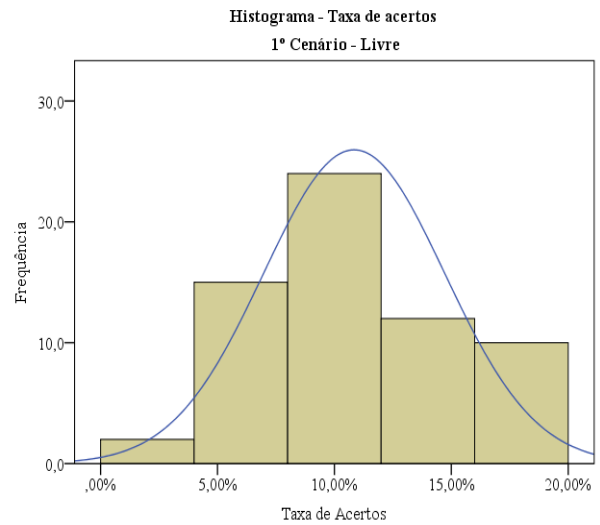


Figura 5.7: Histograma das taxas de acertos para as melhores soluções no cenário livre.

Além disso, elaborou-se um gráfico de dispersão relacionando as taxas de acertos obtidas com a validação e os valores das funções objetivo de cada rodada, evidenciando que não há relação entre essas medidas. O gráfico de dispersão está a seguir ilustrado na Figura 5.8.

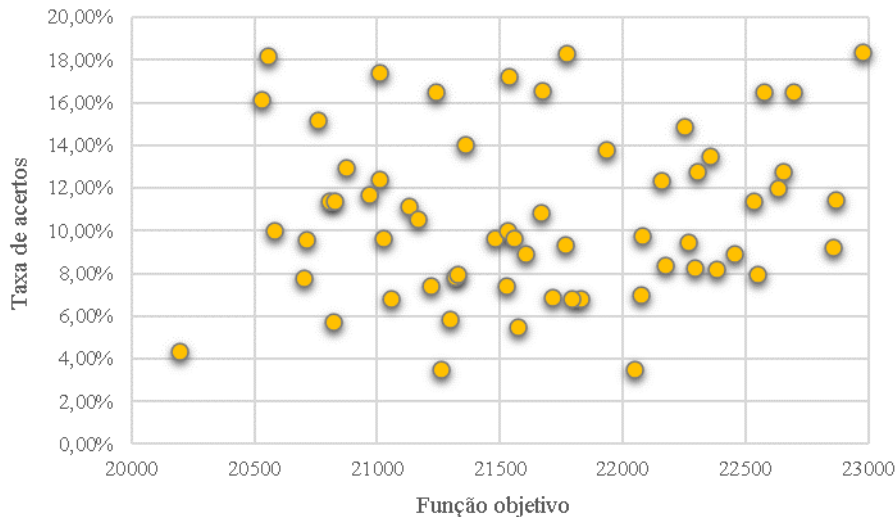


Figura 5.8: Gráfico de dispersão das taxas de acertos e valores de função objetivo no cenário livre.

Ressalta-se que as taxas de acertos, sobretudo globais, não seriam as melhores métricas para avaliação da qualidade da calibração de modelos Logit Multinomiais. Na Seção 5.3, são apresentadas outras análises para os três cenários, tais como: (1) distribuição da impedância de viagem (distâncias de viagens) e (2) análise qualitativa dos parâmetros estimados.

### 5.2.2 Semi Restrito

No cenário semi-restrito, por sua vez, desconsiderou-se, assim como no cenário livre, qualquer informação obtida com a calibração clássica realizada previamente. Aplicou-se, porém, a estratégia para redução na quantidade de parâmetros.

Dessa forma, um total de 482 parâmetros foram calibrados, mantendo-se o intervalo entre -50 e 50 para todos os parâmetros no arquivo de entrada do algoritmo.

Ao todo, o algoritmo no cenário semi-restrito foi executado 60 vezes, com 6000 valores de funções objetivos. Com relação à aplicação dos modelos gerados à amostra de validação, nota-se, assim como no caso livre, grande disparidade nos dados, com taxa de acertos variando entre aproximadamente 4% e 16% e média de 9,2%. As funções utilidade resultantes da calibração que resultou o menor valor da função objetivo encontram-se no Apêndices F, ao final deste trabalho.

As distribuições dos valores das funções objetivo e das taxas de acertos para a melhor solução de cada rodada estão a seguir ilustradas nas Figuras 5.9 e 5.10, respectivamente.

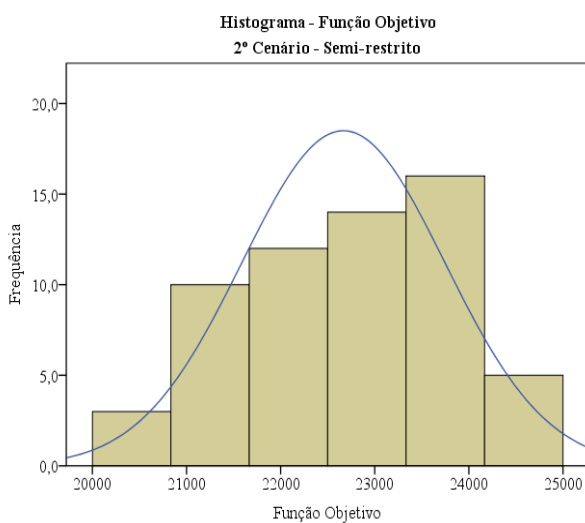


Figura 5.9: Histograma dos valores de função objetivo para as melhores soluções no cenário semi-restrito.

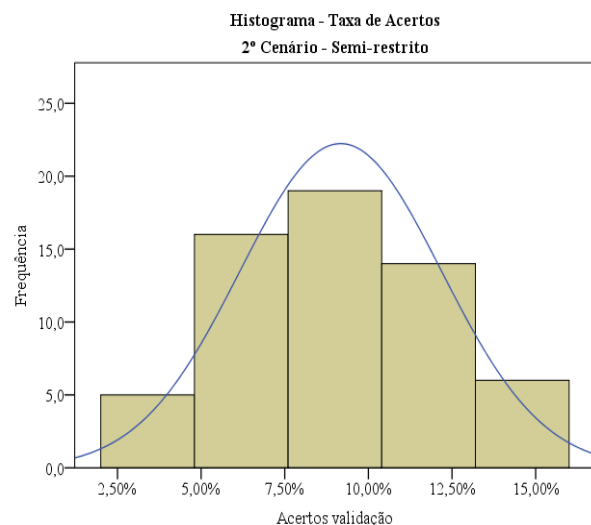


Figura 5.10: Histograma das taxas de acertos para as soluções no cenário semi-restrito.

O gráfico de dispersão relacionando as taxas de acertos obtidas com a validação e os valores das funções objetivo de cada rodada no cenário semi-restrito também evidenciou que não há relação entre essas medidas. O gráfico de dispersão está a seguir ilustrado na Figura 5.11.

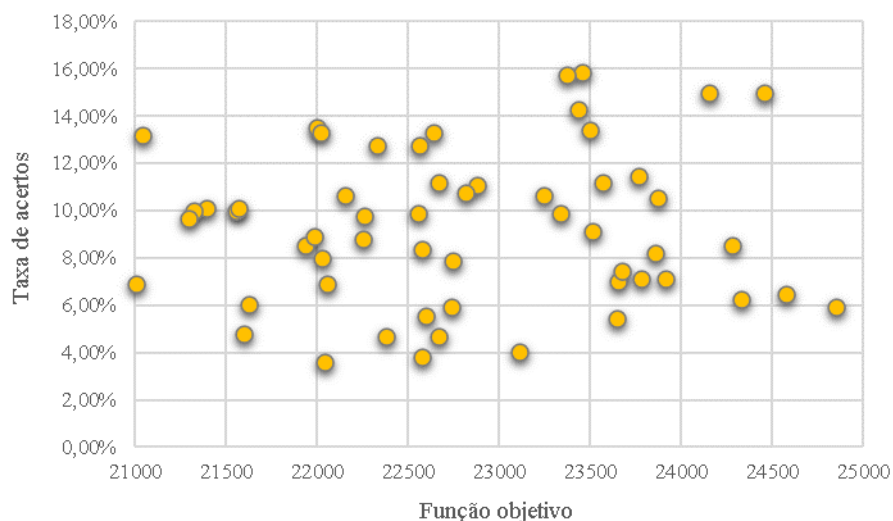


Figura 5.11: Gráfico de dispersão das taxas de acertos e valores de função objetivo no cenário semi-restrito.

### 5.2.3 Restrito

Por fim, no cenário restrito, considerou-se o modelo final obtido com a modelagem clássica. Assim, o modelo calibrado no cenário restrito constituiu-se dos mesmos 200 parâmetros estimados com a calibração clássica. Além disso, o arquivo de entrada contendo os intervalos de valores para os parâmetros foram definidos a partir dos intervalos de confiança disponibilizados no arquivo de saída da calibração clássica.

Ao todo, o algoritmo no cenário restrito foi executado 100 vezes (10.000 valores de funções objetivos). Com relação à aplicação dos modelos gerados à amostra de validação, nota-se grande disparidade nos dados, com taxa de acertos variando entre aproximadamente 12% e 28% e média de 21,7%. As funções utilidade resultantes da calibração que resultou o menor valor da função objetivo encontram-se no Apêndice H, ao final deste trabalho.

As distribuições dos valores das funções objetivo e das taxas de acertos para a melhor solução de cada rodada estão a seguir ilustradas nas Figuras 5.12 e 5.13, respectivamente.

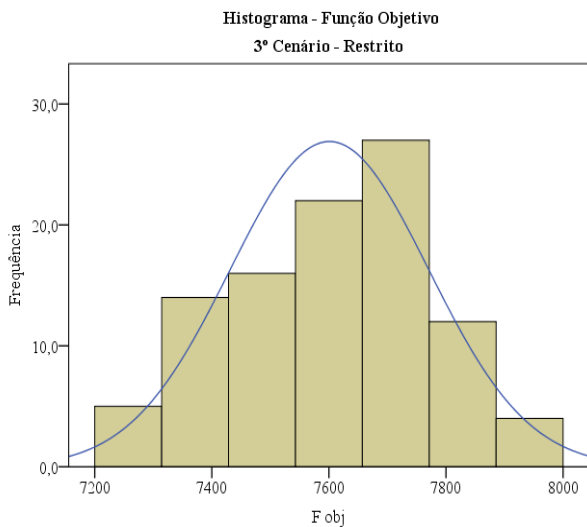


Figura 5.12: Histograma dos valores de função objetivo para as melhores soluções no cenário restrito.

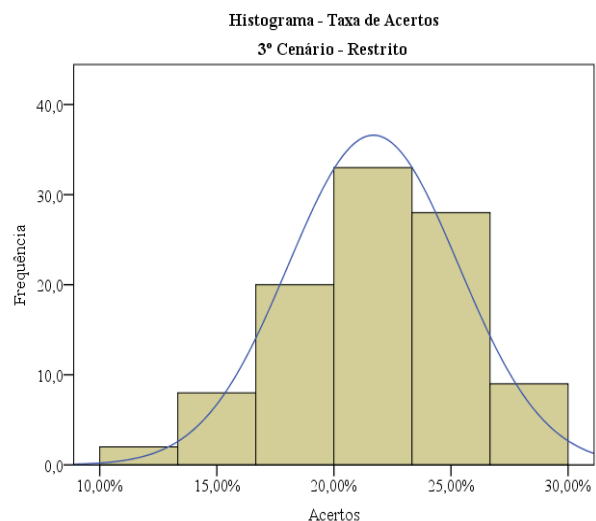


Figura 5.13: Histograma das taxas de acertos para as melhores soluções no cenário restrito.

O gráfico de dispersão relacionando as taxas de acertos obtidos com a validação e os valores das funções objetivo de cada rodada no cenário restrito, ilustrado na Figura 5.14, também evidenciou que não há relação entre essas medidas.

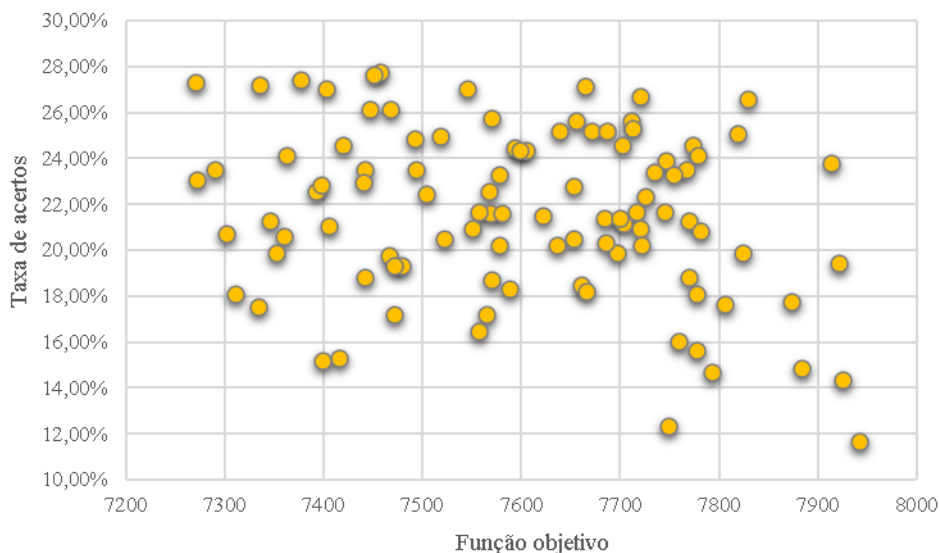


Figura 5.14: Gráfico de dispersão das taxas de acertos e valores de função objetivo no cenário restrito.

### 5.3 COMPARAÇÕES ENTRE AS ABORDAGENS

Apesar de as calibrações por Algoritmos Genéticos não fornecerem informações estatísticas sobre a qualidade do modelo ou dos parâmetros estimados, comparações desses

modelos com o obtido por meio da calibração clássica foram possíveis a partir dos histogramas de distâncias obtidos com a amostra de validação contendo 30% dos registros totais e da análise qualitativa dos parâmetros estimados.

- Histograma de distâncias

O uso da distribuição das distâncias de viagens tem como fundamento a calibração do modelo de distribuição gravitacional e baseia-se no ajuste das curvas de distribuição de comprimentos de viagens (ou qualquer outra impedância de viagem) modelada e observada (Ortuzar e Willumsen, 2011). Assim, a exemplo do realizado por Roma *et al.* (2018), analisou-se a distribuição das frequências das distâncias de viagens observadas e estimadas para a amostra de validação, de forma a identificar o método que minimiza a diferença entre elas.

Nesse contexto, analisou-se a distribuição de frequências de viagens de acordo com faixas de distâncias, considerando os destinos observados para a amostra de validação e os destinos calibrados para a modelagem clássica e para as modelos que resultaram maior verossimilhança para cada cenário de calibração por Algoritmos Genéticos (livre, semi-restrito e restrito). Vale ressaltar que as distâncias foram normalizadas pelo método min-max – ou seja, o intervalo de 0,12 km a 16,49 km foi rearranjado para o intervalo de 0 a 1 – e distribuídas em 8 intervalos, conforme ilustrado na Figura 5.15.

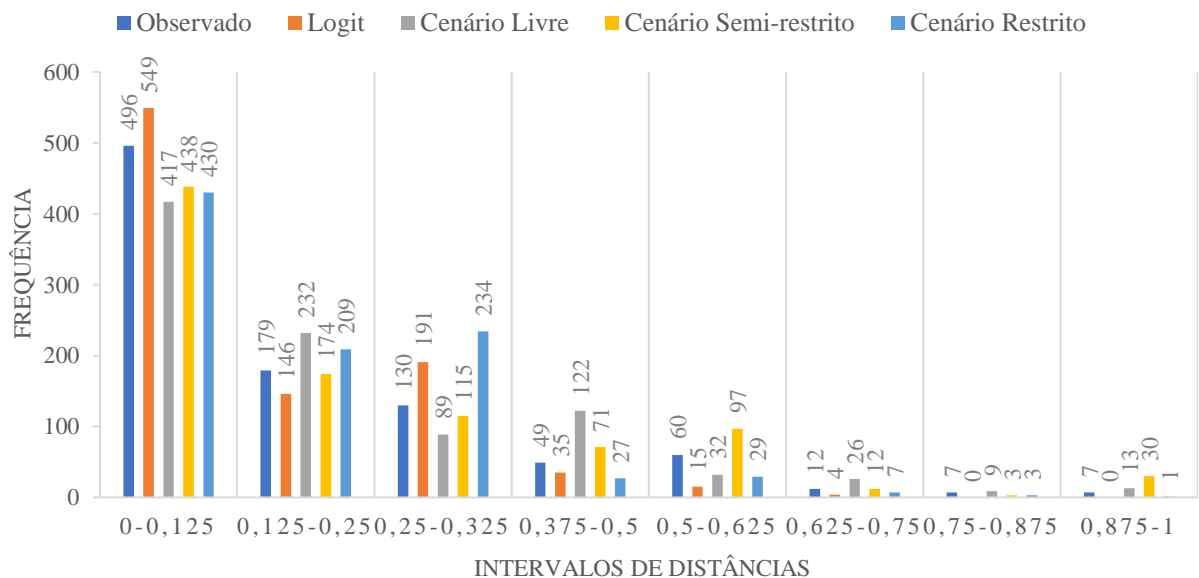


Figura 5.15: Histograma das distâncias de viagem observadas e estimadas.

Para menores faixas de distâncias (0 - 0,125, equivalente a 0,12 km – 2,17 km), observa-se melhores desempenhos para os modelos Logit e AG semi-restrito. Já para faixas entre 0,125 – 0,250 (equivalente a 2,17 km – 4,21 km) verifica-se um desempenho muito superior do cenário AG semi-restrito, seguido pelo cenário restrito. De maneira geral, até distâncias de 6,26 km (0,375), nota-se um desempenho melhor para o modelo calibrado por AG no cenário semi-restrito. Em termos absolutos, isso representa um desempenho superior desse modelo para cerca de 86% das distâncias observadas, haja vista a concentração de observações dentro do intervalo 0 – 0,375 (0,12 km – 6,26 km). Para distâncias superiores, no entanto, as quais concentram menos observações, observa-se maior variabilidade entre os desempenhos dos modelos.

A Tabela 5.3 sintetiza os desempenhos de cada cenário frente à distribuição de distâncias observada para cada intervalo, sendo os modelos considerados superiores classificados com “XXX”, os modelos intermediários com “XX” e os modelos considerados inferiores com “X”. Nela, percebe-se desempenho geral superior do modelo calibrado por Algoritmos Genéticos no cenário semi-restrito, seguido pelo cenário restrito.

Tabela 5.3: Desempenho dos diferentes modelos quanto à distribuição de distâncias.

Distâncias		Modelo			
		Clássico	AG Livre	AG Semi-restrito	AG Restrito
0 – 0,125	0,12 km – 2,17 km	XXX	X	XX	XXX
0,125 – 0,25	2,17 km – 4,21 km	X	X	XXX	XX
0,25 – 0,375	4,21 km – 6,26 km	X	XX	XXX	X
0,375 – 0,5	6,26 km – 8,31 km	XXX	X	XX	XX
0,5 – 0,625	8,31 km – 10,35 km	X	XXX	X	XX
0,625 – 0,75	10,35 km – 12,40 km	X	X	XXX	XX
0,75 – 0,875	12,40 km – 14,44 km	X	XXX	XX	XX
0,875 - 1	14,44 km – 16,49 km	X	X	XXX	XX

Além das análises supracitadas, foram realizados os testes não paramétricos Mann-Whitney e Kolmogorov-Smirnov para comparações das frequências de distâncias. Realizou-se comparações entre os seguintes pares de frequências: (1) valores observados x valores estimados pela modelagem clássica; (2) valores observados x valores estimados pelo AG livre; (3) valores observados x valores estimados pelo AG semi-restrito e (4) valores observados x valores estimados pelo AG restrito, sendo que, em ambos os testes, a hipótese nula, a qual considera a igualdade entre as frequências observadas e estimadas, não foi rejeitada para todos os pares analisados.

- Análise qualitativa

Conforme analisado com os resultados do modelo obtido pela calibração clássica, algumas variáveis apresentam um comportamento já esperado na formulação final das utilidades de cada alternativa, haja vista o conhecimento prévio de características acerca dos destinos possíveis (bairros). As variáveis então analisadas foram: “motivo estudo” e “motivo integração” para o bairro 4 (Camobi), “motivo lazer” para o bairro 21 (Nossa Senhora de Lourdes), “idade” para os bairros 5 (Urlândia), 6 (Carolina), 8 (Centro), 10 (Chácara das Flores), 12 (Itararé) e 24 (Nossa Senhora do Rosário) e “motivo residência” para os bairros 15 (Juscelino Kubitschek) e 26 (Nova Santa Marta).

A Tabela 5.4 expõe os sinais dos parâmetros associados às referidas variáveis para as soluções com o melhor valor de função objetivo (maior verossimilhança), assim como a proporção de ocorrência dos sinais para todas as calibrações dos cenários livre, semi-restrito e restrito. Os sinais na cor verde apresentaram o sinal esperado conforme calibração clássica (apresentados em azul) e prévio conhecimento acerca do bairro em questão. Os sinais na cor vermelha, no entanto, apresentaram o sinal contrário ao esperado, e, portanto, não condizentes com as informações prévias acerca do referido destino.

Tabela 5.4: Comparativo entre os sinais dos parâmetros das variáveis analisadas para cada calibração.

Parâmetro/ Calibração		MotivoEstudo <sub>4</sub>	MotivoIntegração <sub>4</sub>	MotivoLazer <sub>21</sub>	Idade <sub>5</sub>	Idade <sub>8</sub>	Idade <sub>10</sub>	Idade <sub>13</sub>	Idade <sub>24</sub>	MotivoResidência <sub>15</sub>	MotivoResidência <sub>26</sub>
<b>Clássica</b>		(+)	(-)	(+)	(-)	(+)	(+)	(+)	(+)	(+)	(+)
<b>LIVRE</b> (63 calibrações)	Menor função objetivo	(-)	(+)	(+)	(+)	(+)	(+)	(+)	(-)	(+)	(+)
	Geral	(-) (33%)	(-) (52%)	(-) (38%)	(-) (38%)	(-) (3%)	(-) (41%)	(-) (35%)	(-) (37%)	(-) (3%)	(-) (43%)
		(+) (67%)	(+) (48%)	(+) (62%)	(+) (62%)	(+) (97%)	(+) (59%)	(+) (65%)	(+) (63%)	(+) (97%)	(+) (57%)
<b>SEMI-RESTRITO</b> (60 calibrações)	Menor função objetivo	(-)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	Geral	(-) (40%)	(-) (43%)	(-) (43%)	(-) (37%)	(-) (3%)	(-) (35%)	(-) (38%)	(-) (30%)	(-) (5%)	(-) (35%)
		(+) (60%)	(+) (57%)	(+) (57%)	(+) (63%)	(+) (97%)	(+) (65%)	(+) (62%)	(+) (70%)	(+) (95%)	(+) (65%)
<b>RESTRITO</b> (100 calibrações)	Menor função objetivo	(+)	(-)	(+)	(+)	(+)	(+)	(+)	(+)	(+)	(+)
	Geral	(-) (21%)	(-) (80%)	(-) (41%)	(-) (60%)	(-) (35%)	(-) (33%)	(-) (27%)	(-) (19%)	(-) (17%)	(-) (0%)
		(+) (79%)	(+) (20%)	(+) (59%)	(+) (40%)	(+) (65%)	(+) (67%)	(+) (73%)	(+) (81%)	(+) (83%)	(+) (100%)



De maneira geral, observa-se que, para o cenário restrito, uma maioria considerável das calibrações resultaram nos sinais esperados para as variáveis analisadas, o que já era previsto diante das imposições feitas às calibrações nesse cenário (mesma quantidade de parâmetros da calibração clássica e intervalo de valores possíveis para os parâmetros baseado no intervalo de confiança fornecido no arquivo de saída da calibração clássica).

O cenário sem-restrito, por sua vez, apesar de representar menos imposições, apresentou maioria semelhante de sinais esperados. Na calibração com menor valor de função objetivo, no entanto, exibiu mais falhas quando comparada ao cenário restrito.

Por fim, o cenário livre também apresentou os sinais esperados para a maioria das calibrações. Apesar disso, analisando as calibrações com menor valor de função objetivo, foi o cenário que mais apresentou sinais contrários ao esperado.

Em resumo, a despeito do número reduzido de simulações, resultante do longo período de restrições impostas pela pandemia de covid-19 durante o ano de 2020, percebe-se a predominância de acertos dos sinais dos parâmetros associados às variáveis das quais se pode prever o comportamento. Ressalta-se, no entanto, que, tendo em vista a aleatoriedade característica de estimativas realizadas por Algoritmos Genéticos (Mullainathan e Spiess, 2017), sobretudo nos valores “semente” dos parâmetros, e a grande quantidade de parâmetros (580) a serem estimados no presente estudo, uma quantidade muito maior de simulações seria necessária para gerar críticas mais consistentes acerca da análise qualitativa geral dos modelos calibrados por Algoritmos Genéticos.

Em relação à distribuição das distâncias de viagens, os modelos baseados em AG tiveram um desempenho superior em relação ao modelo Logit tradicional. Para o caso de modelos de distribuição de viagens, a ideia mais tradicional de calibração por tentativa e erro visa ajustar as curvas de distribuição de comprimentos de viagens modelada e observada. Em cada caso, tem-se a matriz de viagens e a matriz de impedâncias. Com isso, constrói-se um histograma de quantidade de viagens por intervalos de impedâncias pré-definidas. A ideia clássica da calibração do gravitacional, por exemplo, seria variar o beta para tornar as curvas de impedâncias de viagens estimadas e observadas similares.

Assim, em termos de validação metodológica, a comparação das distribuições de viagens traz algumas vantagens. Usar as viagens estimadas para cada célula não seria o mais adequado considerando a quantidade de células com reduzidas observações no meio urbano. Ao construir a curva de comprimento de viagens, por exemplo, agrega-se todos os pares OD cuja distância está em um mesmo intervalo (definidos para construção do histograma). Portanto,

a amostra em cada intervalo de distâncias é muito superior quando comparada às subamostras contidas em cada célula da matriz OD.

## 6 CONSIDERAÇÕES FINAIS

*Este capítulo sintetiza a organização e o desenvolvimento deste estudo, além de apresentar as principais conclusões e considerações diante dos resultados obtidos com a estratégia proposta para redução da quantidade de parâmetros em modelos de escolha discreta e com a comparação entre os modelos calibrados tradicionalmente e por meio de Algoritmos Genéticos. Ademais, sugestões para trabalhos futuros são apresentadas ao final desta seção.*

### 6.1 SÍNTESE DO TRABALHO E HIPÓTESES DE PESQUISA

Diante da necessidade de considerar as características individuais dos viajantes nos estudos acerca da previsão de destinos, os modelos de escolha discreta configuram-se como uma ferramenta importante para esse fim. Assim, este trabalho abordou questões relativas à calibração do modelo Logit Multinomial para o estudo da previsão de destinos em viagens urbanas.

Problemas operacionais envolvendo a calibração de modelos desagregados com muitas alternativas, as restrições presentes nos métodos de calibração tradicionais e os benefícios dos algoritmos de Aprendizagem de Máquinas constituíram os fatores motivadores para realização deste trabalho, cujos objetivos principais abrangeram a proposição de um procedimento, utilizando algoritmo CART, para redução da quantidade de parâmetros em modelos de escolha discreta com muitas alternativas e o uso de Algoritmos Genéticos como ferramenta de calibração do modelo Logit Multinomial para previsão de escolha de destinos urbanos. Essas questões foram abordadas no primeiro capítulo desta dissertação.

Na sequência, o embasamento teórico necessário para o desenvolvimento desta pesquisa foi apresentado no capítulo 2. O terceiro capítulo, por sua vez, detalhou os dados empregados, desagregados por indivíduo e oriundos de pesquisa domiciliar realizada para o Plano de Diretor de Mobilidade de Santa Maria, o tratamento neles realizado, incluindo um procedimento a partir de clusterização dos destinos para viabilizar as análises realizadas na etapa de elaboração do procedimento para redução da quantidade de parâmetros. O método seguido para realização deste estudo e os aplicativos utilizados também foram apresentados no terceiro capítulo.

O quarto capítulo discorreu acerca dos resultados de um dos objetivos principais deste trabalho: a proposição de uma estratégia para redução de parâmetros em modelos com grandes conjuntos de alternativas, como no caso da escolha de destinos urbanos. Esse procedimento foi de suma importância para viabilizar a execução da calibração clássica do modelo Logit

Multinomial para escolha de destinos em um nível mais desagregado (por bairros), etapa realizada para comparação com a calibração por Algoritmos Genéticos.

O sexto capítulo apresentou e comparou os resultados referentes às calibrações clássica e por meio de Algoritmos Genéticos do modelo Logit Multinomial para uma abordagem desagregada de distribuição de viagens, além de fornecer detalhes sobre a configuração adotada para os parâmetros necessários ao funcionamento do Algoritmo Genético.

Finalmente, levando-se em conta as hipóteses de pesquisa, descritas no Capítulo 1, faz-se uma breve análise destas após os resultados obtidos:

- **Hipótese de pesquisa 1: É viável a aplicação de Algoritmos Genéticos para modelagens desagregadas de distribuição de viagens urbanas;**

As análises e resultados demonstram a viabilidade de Algoritmos Genéticos para estimativa de deslocamentos urbanos e composição de matrizes OD. No entanto, considerando a quantidade elevada de parâmetros a serem estimados e a natureza aleatória dos Algoritmos Genéticos, sugere-se a realização de um número alto de simulações.

- **Hipótese de pesquisa 2: Quando comparado ao modelo obtido pela calibração tradicional do modelo Logit Multinomial, o modelo obtido pela calibração por Algoritmos Genéticos é eficaz para análises desagregadas de distribuição de viagens urbanas;**

A análise baseada na distribuição de impedâncias de viagens (mais especificamente comprimentos de viagens) demonstra a superioridade da calibração por Algoritmos Genéticos. Tal análise seria mais adequada para o caso de matrizes OD urbanas, onde há um grande número de células nulas ou com reduzidas observações.

- **Hipótese de pesquisa 3: O algoritmo CART pode ser utilizado para simplificar a calibração de modelos Logit Multinomiais em casos com grande conjunto de alternativas e/ou muitas variáveis explicativas, associadas aos indivíduos.**

O sucesso da estratégia proposta e apresentada nos Capítulos 3 e 4 é evidenciado considerando a validação proposta e comparações de métricas de desempenho para o caso de adoção ou não da estratégia.

## 6.2 CONCLUSÕES

Os resultados apresentados neste trabalho para o estudo da previsão de destinos urbanos permitem concluir que, embora se apresentem como uma ferramenta poderosa e flexível, os algoritmos de Aprendizagem de Máquinas exibem características que demandam certa cautela com relação ao seu uso nas calibrações de modelos Logit Multinomiais. Ainda que aplicados a um mesmo problema central, os algoritmos de aprendizagem de máquinas utilizados neste trabalho (CART e Algoritmos Genéticos) foram utilizados com propósitos distintos, e, com isso, ofereceram também conclusões distintas.

Com relação ao procedimento elaborado com algoritmo CART para redução da quantidade de parâmetros a serem estimados em modelos de escolha discreta, a estratégia proposta mostrou-se eficiente, haja vista a redução significativa da demanda operacional frente à manutenção da qualidade geral do modelo. De maneira geral, conclui-se que o procedimento se mostra viável para casos com grandes conjuntos de alternativas. Ressalta-se, no entanto, que, apesar de bem-sucedido, o procedimento proposto neste trabalho requer conhecimento prévio do pesquisador ou informações da literatura acerca do que se pretende avaliar, de forma a impedir a retirada de variáveis potencialmente relevantes ao modelo.

Sobre a aplicação de Algoritmos Genéticos para a calibração de modelo Logit Multinomial para previsão de destinos urbanos, conclui-se que, assim como a calibração clássica, os modelos que utilizaram Algoritmos Genéticos apresentam vantagens e desvantagens. Embora não apresentem restrições operacionais para grandes conjuntos de alternativas e muitas variáveis relacionadas ao indivíduo, os modelos calibrados por Algoritmos Genéticos podem necessitar de um número alto de simulações para maior precisão e acurácia. A análise do histograma de distâncias indicou que os modelos calibrados por Algoritmos Genéticos, sobretudo nos cenários semi-restrito e restrito, apresentaram desempenho superior ao obtido por meio da calibração clássica. Ainda, as análises dos parâmetros referentes às variáveis cujo comportamento é sabido, devido ao conhecimento prévio e a informações da literatura acerca dos destinos estudados, demonstraram que os as calibrações por Algoritmos Genéticos produziram os sinais esperados para esses parâmetros na maioria das calibrações.

Em suma, conclui-se que a calibração do modelo Logit Multinomial por meio de Algoritmos Genéticos mostrou grande potencial, não somente para problemas de escolhas de destinos urbanos, mas para qualquer outro problema relacionado ao comportamento individual, como escolha modal, escolha de rota ou escolha de destinos intermunicipais. No entanto, para problemas de maior complexidade, como foi o caso do modelo com 35 destinos possíveis aqui

analisado, essa ferramenta apresenta uma restrição temporal a ser considerada, pois, devido à grande quantidade de parâmetros a serem estimados e à aleatoriedade característica de estimativas realizadas por Algoritmos Genéticos, é necessária uma grande quantidade de simulações para a obtenção de um modelo consistente.

### 6.3 SUGESTÕES PARA TRABALHOS FUTUROS

Diante do bom desempenho do procedimento proposto que utiliza algoritmo CART para redução da quantidade de parâmetros a serem estimados, sugere-se:

- Aplicação da estratégia proposta para diferentes estudos de caso, de modo a verificar sua aplicabilidade para outras análises que não a previsão de destinos urbanos;
- Verificar, por exemplo, a aplicabilidade do procedimento proposto para o caso de um problema de escolha modal, onde a retirada de uma variável independente pode ser melhor analisada pelo pesquisador, considerando relações prévias, conhecidas na literatura.

Ainda, tendo em vista que a propriedade da independência das alternativas irrelevantes (IIA), segundo a qual a proporção de probabilidade de escolha de uma alternativa sobre outra não é afetada pela inclusão ou retirada de novas alternativas, configura-se como uma das principais restrições do modelo Logit Multinomial sempre que houver alternativas correlacionadas, sugere-se:

- Aplicação de Algoritmos Genéticos para calibração das funções utilidade em modelos de distribuição de viagens urbanas nos moldes do modelo Logit Multinomial, porém com a função objetivo elaborada de forma a maximizar as taxas de acertos, ou mesmo buscando minimizar as diferenças entre as distâncias estimada e observada;
- Propor a calibração de outro modelo de escolha discreta que não tenha a suposição IIA como o Modelo Logit com Componente de Erro para replicação do método aqui proposto.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Antonini, G.; Bierlaire, M.; Weber, M. (2006) Discrete choice models of pedestrian walking behavior, *Transportation Research Part B: Methodological*, 40(8), p. 667-687. DOI: 10.1016/j.trb.2005.09.006.
- Ben-Akiva, M.; Bierlaire, M. (1999) Discrete Choice Methods and their Applications to Short Term Travel Decisions. In: Hall R.W. (Eds) *Handbook of Transportation Science*. International Series in Operations Research & Management Science. Springer, Boston, MA, pp. 5-33. DOI: 10.1007/978-1-4615-5203-1\_2.
- Ben-Akiva M.; Bierlaire M. (2003) Discrete Choice Models with Applications to Departure Time and Route Choice. In: Hall R.W. (Eds) *Handbook of Transportation Science*. International Series in Operations Research & Management Science. Springer, Boston, MA, pp. 7-37. DOI: 10.1007/0-306-48058-1\_2.
- Ben-Akiva, M.; Lerman, S. R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge, MA: MIT Press, 384 pp.
- Bhat, C. R. (1995) A heteroscedastic extreme value model of intercity travel mode choice, *Transportation Research Part B*, 29(6), p. 471-483. DOI: 10.1016/0191-2615(95)00015-6.
- Bhat, C. R. (1998) Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling. *Transportation Research Part A: Policy and Practice*, 32(7), p. 495-507. DOI: 10.1016/S0965-8564(98)00011-1.
- Bhat, C.; Govindarajan, A.; Pulugurta, V. (1998) Disaggregate Attraction-End Choice Modelling: Formulation and Empirical Analysis. *Transportation Research Board*, 1645, p. 60-68. DOI: 10.3141/1645-08.
- Bierlaire, M. (2003) “BIOGEME: a free package for the estimation of discrete choice models”, 3<sup>rd</sup> *Swiss Transport Research Conference, STRC 2003*.
- Bierlaire, M. (2018) *PandasBiogeme: a short introduction*. Technical report TRANSP-OR 181219. Transport and Mobility Laboratory, ENAC, EPFL.
- Bierlaire, M. e Feticarison, M. (2009) “Estimation of discrete choice models: extending BIOGEME”.
- Bierlaire, M.; Lapparent, M.; Sharif, S. (2016) Decision-aid Methodologies in Transportation, TRANSP-OR. Disponível em <[http://transp-or.epfl.ch/courses/decisionAid2016/labs/Lab\\_10/Exercise/MNLSwissmetro.pdf](http://transp-or.epfl.ch/courses/decisionAid2016/labs/Lab_10/Exercise/MNLSwissmetro.pdf)>. Acesso em: 18 de ago. de 2019.

- Bordley, R. (2013) Discrete choice with large choice sets, *Economics Letters*, 118(1), p. 13-15. DOI: 10.1016/j.econlet.2012.05.010.
- Bowman, J. L. e Ben-Akiva, M. E. (2001) Activity-based disaggregate travel demand model system with activity schedules, *Transportation Research Part A: Policy and Practice*, 35(1), p. 1-28. DOI: 10.1016/S0965-8564(99)00043-9.
- Breiman, L.; Friedman, J.H; Olshen, R.A.; Stone, C.J. (1984) *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA.
- Carvalho, A. C. P. L. F.; Galvão, C. O.; Lacerda, E. G. M.; Diniz, L. S.; Valença, M. J. S.; Ludermir, T. B.; Vieira, V. P. P. B. (1999). *Sistemas inteligentes: Aplicações a recursos hídricos e ambientais*. Porto Alegre: Editora Universidade/ UFRGS/ ABRH. ISBN 8570255276.
- Cascetta, E.; Pagliara, F.; Papola, A. (2007) Alternative approaches to trip distribution modelling: A retrospective review and suggestions for combining different approaches, *Papers in Regional Science*, 86(4), p. 597–620. DOI: 10.1111/j.1435-5957.2007.00135.x.
- Casey, H. J. (1955) Applications to traffic engineering of the law of retail gravitation. *Traffic Quarterly IX*, p. 23-35.
- Chen, L.; Duann, W. Hu. (2005) The estimation of discrete choice models with large choice-set. *Journal of the Eastern Asia Society for Transportation Studies*, v. 6, p. 1724-1739. DOI: 10.11175/easts.6.1724.
- Chow, L.; Zhao, F.; Li, M.; Li, S. (2005) Development and evaluation of aggregate destination choice models for trip distribution in Florida, *Transportation Research Record: 1931(1)*, pp. 18–27. DOI: 10.3141/1931-03.
- Cunha, A. L.; Bessa, J. E.; Setti, J. R. (2009) Genetic algorithm for the calibration of vehicle performance models of microscopic traffic simulators, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, v. 5816 LNAI, p. 3–14. DOI: 10.1007/978-3-642-04686-5\_1.
- Ding, L.; Zhang, N. (2016). A travel mode choice model using individual grouping based on cluster analysis. *Procedia Engineering*, v. 137, p. 786–795. DOI: 10.1016/j.proeng.2016.01.317.
- Domecich, T. A.; McFadden, D. (1975) *Urban travel demand: A behavioral analysis*, North Holland Publishing Co., NY, USA.
- Evans, A.W. (1970) Some properties of trip distribution methods. *Transportation Research*, v. 4(1), p. 19–36. DOI: 10.1016/0041-1647(70)90072-9.



- Evans, S.P.; Kirby, H.R. (1974) A three-dimensional furness procedure for calibrating gravity models. *Transportation Research*, v. 8(2), p. 105–122. DOI: 10.1016/0041-1647(74)90037-9
- Fotheringham, A.S. (1983) Some theoretical aspects of destination choice and their relevance to production-constrained gravity models. *Environment and Planning A*, v. 15 (8), p. 1121–1132. DOI: 10.1068/a151121
- Frejinger, E.; M. Bierlaire; M. Ben-Akiva (2009) Sampling of Alternatives for Route Choice Modeling. *Transportation Research Part B*, v. 43, 984-994. DOI: 10.1016/j.trb.2009.03.001.
- Friedman, J.; Gerlowski, D.; Silberman, J. (1992) What attracts foreign multinational corporations? Evidence from branch plant location in the united states, *Journal of Regional Science*, v. 32(4), p. 403-418. DOI: 10.1111/j.1467-9787.1992.tb00197.x.
- Genkin, A.; Lewis, D. D.; Madigan, D. (2007) Large-Scale Bayesian Logistic Regression for Text Categorization, *Technometrics*, v. 49(3), p.291-304. DOI: 10.1198/0040170070000000245.
- Goel, S., Singh, J. B.; Sinha, A. K. (2012) Trip Distribution Model for Delhi Urban Area Using Genetic Algorithm, *International journal of computer engineering science*, 2(3), p. 1–8. DOI: 10.1186/1472-6769-10-6.
- Goldberg, D. (1989) *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, MA. 432 pp.
- Goldberg, D. E.; Holland, J. H. (1988) Genetic Algorithms and Machine Learning, *Machine Learning*, v. 3(2), p. 95–99. DOI: 10.1023/A:1022602019183.
- Hansen, E. (1987) Industrial location choice in Sao Paulo, Brazil: a nested Logit model, *Regional Science and Urban Economics*, v. 17(1), p. 89-108. DOI: 10.1016/0166-0462(87)90070-6.
- Holland, J. (1975) *Adaptation in Natural and Artificial Systems*. 211 pp. ISBN: 9780262275552.
- Hu, X.; Zhang, Z. X. (2008) “Application of genetic algorithm to spatial distribution in urban planning”, *2008 IEEE International Symposium on Knowledge Acquisition and Modeling Workshop Proceedings*, p. 1026–1029. DOI: 10.1109/KAMW.2008.4810667.
- IBGE - Instituto Brasileiro de Geografia e Estatística (2010) *Censo Demográfico 2010*.
- IBM Corp. Released 2016. *IBM SPSS Statistics for Windows, Version 24.0*. Armonk, NY: IBM Corp.

- IBM SPSS *Missing Values 24* (2016), SPSS Inc., Chicago.
- IPLAN (2013) *Plano Diretor de Mobilidade Urbana de Santa Maria*. 1ª Ed. Santa Maria: Prefeitura Municipal de Santa Maria.
- Kalić, M.; Teodorović, D. (2003) Trip distribution modelling using fuzzy logic and a genetic algorithm, *Transportation Planning and Technology*, v. 26(3), p. 213–238. DOI: 10.1080/0308106032000154575.
- Kass, G.V. (1980) An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, v. 29(2), p. 119–127. DOI: 10.2307/2986296.
- Kawamoto, E. (2004) *Análise de sistemas de transporte*. 2. ed. São Carlos - SP: Escola de Engenharia de São Carlos. Universidade de São Paulo. 229 pp.
- Keane, M. P.; Wasi, N. (2012) Estimation of Discrete Choice Models with Many Alternatives Using Random Subsets of the Full Choice Set: With an Application to Demand for Frozen Pizza, *Economics Papers 2012-W13*, Economics Group, Nuffield College, University of Oxford.
- Koppelhuber, F.; Kribernegg, G.; Luger, B.; Aspack, J. (2017) “Application of discrete choice models for mode and destination choice in a large scale demand model”, *IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 469-473. DOI: 10.1109/MTITS.2017.8005718
- Kühlkamp, N. (2003) *Modelo de oportunidades intervenientes de distribuição de viagens com ponderação das posições espaciais relativas das oportunidades*. Tese (Doutorado em Engenharia Civil). Universidade Federal de Santa Catarina. Florianópolis. 197 pp.
- Kühlkamp, N., Ulysséa Neto, I. (2003) Modelos de Oportunidades Intervenientes de Distribuição de Viagens: Um Novo Método de Calibração do Modelo de Schneider. Universidade Federal de Santa Catarina. *Revista Transportes*, v. 11, p. 19-24.
- Lemos, B., Orrico, R.; Vieira da Silva, M. (2017) “Avaliação de uma Nova Abordagem para Distribuição de Viagens na Cidade do Rio de Janeiro”, *XXXI Congresso Nacional de Pesquisa em Transporte da ANPET*, (2008), pp. 2071–2082.
- Lemp, J. D.; Kockelman, K. M. (2012) Strategic sampling for large choice sets in estimation and application, *Transportation Research Part A: Policy and Practice*, v. 46(3), p. 602-613. DOI: 10.1016/j.tra.2011.11.004.
- Lindner, A.; Pitombo, C.S.; Cunha, A.L. (2017) Estimating motorized travel mode choice using classifiers: An application for high-dimensional multicollinear data. *Travel Behaviour and Society*, v. 6, p. 100-109. DOI: 10.1016/j.tbs.2016.08.003.

- Lowry, I. S. (1964) *A model of metropolis*. RM-4125-RC, Rand Corporation, Santa Monica. California. 136 pp.
- Luce, R. D.; Suppes, P. (1965) Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, v. 111. New York: Wiley. pp. 249-410.
- Madan, D.B.; Groenhou, R. (1987) Modelling travel mode choices for the Sydney work trip, *Journal of Transport Economics and Policy*, v. 21, p. 135-149.
- McFadden, D. (1978a) Modelling the choice of residential location. In A. Karquist, L. Lundqvist and F. Snickars (eds.), *Spatial Interaction Theory and Planning Models*. North-Holland, Amsterdam.
- McFadden, D. L. (1978b) "Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments", in D. Hensher and P. Stopher (Eds.), *Behavioural Travel Modelling*, Croom Helm London: London, p. 279-318.
- Michie, D.; Spiegelhalter, D.J.; Taylor, C.C. (1994). *Machine Learning, Neural and Statistical Classification*. Ellis Horwood Limited.
- Mikkonen, K.; Luoma, M., (1999). The parameters of the gravity model are changing - how and why? *Journal Transportation Geography*, v. 7(4), p. 277-283. DOI: 10.1016/S0966-6923(99)00024-1.
- Mishra, S.; Yanli, W.; Xiaoyu, Z.; Rolf, M e Subrat, M. (2013) Comparison between Gravity and Destination Choice Models for Trip Distribution in Maryland, *Transportation Research Board*, p. 1-22.
- Mitchell, Tom M. (1997) *Machine Learning*. McGraw-Hill. ISBN 0070428077.
- Mozolin, M.; Thill, J. C.; Lynn Usery, E. (2000) Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation, *Transportation Research Part B: Methodological*, v. 34(1), p. 53-73. DOI: 10.1016/S0191-2615(99)00014-4.
- Mullainathan, S.; Spiess, J. (2017) *Machine Learning: An Applied Econometric Approach*. *Journal of Economic Perspectives*, v. 31(2): 87-106. DOI: 10.1257/jep.31.2.87.
- Negnevitsky, M. (2005) *Artificial Intelligence: A Guide to Intelligent Systems*. Second Edition. England: Pearson Education. 415 pp.
- Nerella, S.; C.R. Bhat (2004) Numerical Analysis of Effect of Sampling of Alternatives in Discrete Choice Models. *Transportation Research Record*, v. 1894, p. 11-19.
- Novaes, A. G. (1986) *Modelos em Planejamento Urbano, Regional e de Transportes*. São Paulo: Edgard Blücher. 290 pp.

- Ortúzar, J. D.; Willumsen, L. G. (2011) *Modelling Transport*. Wiley, London.
- Paredes, M.; Hemberg, E.; O'Reilly, U.; Zegras, C. (2017) "Machine learning or discrete choice models for car ownership demand estimation and prediction?", *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, Naples, 2017, pp. 780-785. DOI: 10.1109/MTITS.2017.8005618.
- Pitombo, C. S.; Guimarães, H. S. (2016) Uso de técnica de mineração de dados no auxílio à modelagem de distribuição de viagens intermunicipais, *Engenharia Civil UM*, v. 52, p. 45–56.
- Pitombo, C. S.; Kawamoto, E. ; Sousa, A. J. (2011) An exploratory analysis of relationships between socioeconomic, land use, activity participation variables and travel patterns. *Transport Policy (Oxford)*, v. 18, p. 347-357.
- Pitombo, C. S.; de Souza, A. D.; Lindner, A. (2017) Comparing decision tree algorithms to estimate intercity trip distribution, *Transportation Research Part C: Emerging Technologies*, v. 77, p. 16–32. DOI: 10.1016/j.trc.2017.01.009.
- Plaza, C. V., Rodrigues da Silva, A. N. (2015) "Intrazonal trip distances: an estimation approach and application to a case study", in: *Transportation Research Board 94th Annual Meeting*.
- Pulugurta, S., Arun, A.; Errampalli, M. (2013) 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. DOI: 10.1016/j.sbspro.2013.11.152.
- Quinlan, R. (1983) Learning efficient classification procedures and their application to chess end-games. *Machine Learning: An Artificial Intelligence Approach*, Tioga, Palo Alto, p. 463–482. DOI: 10.1016/B978-0-08-051054-5.50019-4.
- Rocha, S. S. (2019) Uso de Algoritmos Genéticos para otimização de modelagem geoestatística aplicada à demanda por transportes. Tese de Doutorado, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos. DOI:10.11606/T.18.2019.
- Rocha, S.S; Pitombo, C.S.; Costa, L.H.M; Marques, S.F (2021) Applying optimization algorithms for spatial estimation of travel demand variables. *Transportation Research Interdisciplinary Perspectives*. Artigo aceito
- Rodrigues, F. L.; Leite, H. G.; Santos, H. N; Souza, A. L. e Silva, G. F. (2004) Metaheurística algoritmo genético para solução de problemas de planejamento florestal com restrições de integridade. *Revista Árvore*, v. 28(2), p. 233-245. DOI: 10.1590/S0100-67622004000200010.

- Roma, A. D. de S., Souza Pitombo, C., Guimarães, H. S., & Costa, L. H. M. (2018). Análise de desempenho de algoritmos de aprendizagem de máquinas para análise desagregada de viagens intermunicipais. *Transportes*, v. 26(3), p. 159–175. DOI: 10.14295/transportes.v26i3.1614
- Russel, S.J.; Norvig, P. (2002) *Artificial Intelligence: A Modern Approach*. 2. Prentice-Hall. 1152 pp. ISBN10 0137903952.
- Schneider, M. (1959) Gravity model and trip distribution theory, *Papers in Regional Science*. DOI: 10.1111/j.1435-5597.1959.tb01665.x.
- Shmueli, D., Salomon, I.; Shefer, D. (1996) Neural network analysis of travel behavior: Evaluating tools for prediction, *Transportation Research Part C: Emerging Technologies*, 4(3), p. 151–166. DOI: 10.1016/S0968-090X(96)00007-1.
- Souza, A. D.; Pitombo, C. S.; Guimarães, H. S.; Costa, L. H. M. (2018) Análise de desempenho de algoritmos de aprendizagem de máquinas para análise desagregada de viagens intermunicipais. *Transportes* (Rio de Janeiro), v. 26, p. 159-175.
- Stouffer, S. A. (1940) Intervening Opportunities: A Theory Relating Mobility and Distance, *American Sociological Review*, p. 845-867. DOI: 10.1016/S1003-6326(10)60620-6.
- Sung, K.; Rakha, H. (2009) A Genetic algorithm for Trip Distribution and Traffic Assignment from Traffic Counts in a Stochastic User Equilibrium. *International Journal of Management Science*, 15.
- Thomas, R. W.; Hugget, R. J. (1980) *Modeling in Geography: A Mathematical Approach*. Barnes and Noble, Totowa.
- Train, K. (2009) *Discrete Choice Methods with Simulation*. Cambridge University Press. 398 pp.
- Williams, I. (1976) A comparison of some calibration techniques for doubly constrained models with an exponential cost function, *Transportation Research*, v. 10(2), p. 91-104. DOI: 10.1016/0041-1647(76)90045-9.
- Wilson, A. G. (1967) A statistical theory of spatial distribution models, *Transportation Research*, v. 1(3), p. 253–269. DOI: 10.1016/0041-1647(67)90035-4.
- Wilson, A. G. (1970) Advances and problems in distribution modelling, *Transportation Research*, v. 4(1), p. 1–18. DOI: 10.1016/J.BMCL.2012.01.082.
- Winston, C. (1983) The demand for freight transportation: models and applications, *Transportation Research Part A: General*, v. 17(6), p. 419-427. DOI: 10.1016/0191-2607(83)90162-0.

- Xie, C.; Lu, J.; Parkany, E. (2003) Work Travel Mode Choice Modeling with Data Mining: Decision Trees and Neural Networks, *Transportation Research Record: Journal of the Transportation Research Board*, 1854(03), p. 50–61. DOI: 10.3141/1854-06.
- Zhong, M.; Lingras, P. (2008) Evolving parameters of Logit model using genetic algorithm.

**APÊNDICE A - Calibração sem o procedimento por “macro Bairros” para nível de significância de 5%**

**Estimation report**

**Number of estimated parameters:** 97

**Sample size:** 2196

**Excluded observations:** 0

**Init log likelihood:** -6465.988

**Final log likelihood:** -4186.032

**Likelihood ratio test for the init. model:** 4559.912

**Rho-square for the init. model:** 0.353

**Rho-square-bar for the init. model:** 0.337

**Akaike Information Criterion:** 8570.064

**Bayesian Information Criterion:** 9133.809

**Final gradient norm:** 1.5869E-01

**Diagnostic:** b'CONVERGENCE:  
Norm\_of\_projected\_gradient<=PGTOL'

**Database readings:** 305

**Iterations:** 287

**Data processing time:** 0:00:00

**Optimization time:** 0:00:48.000001

**Nbr of threads:** 8

**Estimated parameters**

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
Beta10_Distancia	-2.94	0.82	-3.59	0.000329	0.797	-3.7	0.000218
Beta10_FrequenciaEscolar	-2.64	0.618	-4.27	1.96e-05	0.641	-4.12	3.86e-05
Beta10_GraudeInstrucao	-3.65	0.699	-5.22	1.75e-07	0.634	-5.76	8.49e-09
Beta10_Idade	-4.29	1.05	-4.1	4.14e-05	0.942	-4.55	5.33e-06
Beta10_Moradores	-2.03	0.953	-2.13	0.0329	0.872	-2.33	0.0197
Beta10_MotivoLazer	5.47	1.15	4.78	1.77e-06	1.07	5.1	3.43e-07
Beta10_MotivoResidencia	4.78	1.01	4.74	2.13e-06	0.885	5.4	6.55e-08
Beta10_MotivoSaude	6.13	1.16	5.28	1.32e-07	1.1	5.58	2.38e-08
Beta10_MotivoTrabalho	4.31	1.01	4.27	1.97e-05	0.864	4.98	6.31e-07
Beta10_PopulacaoOrigem	2.28	0.687	3.32	0.000909	0.805	2.83	0.00464
Beta11_Distancia	-5.35	0.617	-8.67	0	0.658	-8.14	4.44e-16
Beta11_GraudeInstrucao	1.19	0.304	3.91	9.42e-05	0.288	4.13	3.66e-05
Beta11_Moradores	-1.91	0.502	-3.8	0.000147	0.502	-3.79	0.000148
Beta11_MotivoIntegracao	5.04	9.21	0.547	0.584	0.477	10.6	0

Beta11_MotivoResidencia	-1.38	0.179	-7.73	1.09e-14	0.178	-7.78	7.33e-15
Beta11_PopulacaoOrigem	1.16	0.338	3.44	0.000579	0.323	3.6	0.000324
Beta12_Distancia	-2.55	0.341	-7.48	7.24e-14	0.342	-7.45	9.3e-14
Beta12_GraudeInstrucao	1.39	0.224	6.2	5.53e-10	0.223	6.21	5.13e-10
Beta12_Moradores	-2.03	0.389	-5.22	1.81e-07	0.41	-4.96	7.11e-07
Beta12_MotivoIntegracao	5.7	9.21	0.619	0.536	0.256	22.3	0
Beta12_MotivoResidencia	-2.61	0.146	-17.9	0	0.15	-17.4	0
Beta12_PopulacaoOrigem	1.07	0.281	3.81	0.000138	0.283	3.78	0.000157
Beta13_Distancia	-6.12	0.729	-8.4	0	0.947	-6.47	9.96e-11
Beta13_MotivoIntegracao	-4.05	38	-0.107	0.915	0.228	-17.7	0
Beta13_MotivoResidencia	-1.6	0.19	-8.44	0	0.193	-8.28	2.22e-16
Beta13_PopulacaoOrigem	2.58	0.262	9.83	0	0.271	9.52	0
Beta14_Distancia	-4.57	1.36	-3.35	0.000798	1.03	-4.44	9.19e-06
Beta14_MotivoEstudo	2.33	0.806	2.89	0.00389	0.738	3.15	0.00161
Beta14_MotivoSaude	-3.12	13.7	-0.228	0.82	0.342	-9.12	0
Beta14_MotivoTrabalho	-5.2	14.7	-0.355	0.723	0.282	-18.4	0
Beta15_FrequenciaEscolar	6.43	7.51	0.857	0.392	0.788	8.16	2.22e-16
Beta15_GraudeInstrucao	8.59	4.13	2.08	0.0374	3.17	2.71	0.00676
Beta15_MotivoLazer	6.07	5.68	1.07	0.286	1.03	5.89	3.84e-09
Beta15_MotivoResidencia	-6.18	51.2	-0.121	0.904	0.526	-11.7	0
Beta15_MotivoTrabalho	6.11	5.83	1.05	0.294	1.04	5.87	4.25e-09
Beta15_Renda	-4.39	3.33	-1.32	0.187	1.29	-3.4	0.000686
Beta15_Trabalha	-2.53	1.62	-1.56	0.118	1.04	-2.44	0.0148
Beta15_VeiculosMotorizados	8.29	5.07	1.64	0.102	1.66	5.01	5.5e-07
Beta16_Distancia	-16	2.84	-5.61	1.98e-08	4.46	-3.58	0.000346
Beta16_MotivoIntegracao	-1.38	33.6	-0.0411	0.967	0.423	-3.27	0.00109
Beta16_MotivoLazer	-5.32	16	-0.333	0.739	0.301	-17.6	0
Beta16_MotivoSaude	-5.16	17	-0.302	0.762	0.351	-14.7	0
Beta17_Distancia	-4.22	0.576	-7.32	2.44e-13	0.62	-6.8	1.03e-11
Beta17_Moradores	-2.55	0.772	-3.3	0.000978	0.759	-3.35	0.000798
Beta17_MotivoIntegracao	-2.49	34.9	-0.0712	0.943	0.243	-10.2	0
Beta17_PopulacaoOrigem	1.61	0.498	3.22	0.00127	0.436	3.68	0.000229
Beta18_Distancia	-3.97	0.444	-8.94	0	0.493	-8.05	8.88e-16
Beta18_Moradores	-2.69	0.517	-5.21	1.85e-07	0.473	-5.7	1.18e-08
Beta18_MotivoEstudo	0.701	0.333	2.1	0.0354	0.294	2.39	0.017
Beta18_MotivoIntegracao	-3.48	36.2	-0.0961	0.923	0.236	-14.8	0
Beta18_PopulacaoOrigem	2.28	0.286	7.99	1.33e-15	0.291	7.84	4.44e-15
Beta18_Renda	-1.77	0.343	-5.15	2.6e-07	0.327	-5.4	6.67e-08
Beta19_FrequenciaEscolar	-6.61	17.1	-0.386	0.699	0.219	-30.2	0
Beta19_MotivoResidencia	-8.7	17.2	-0.507	0.612	0.209	-41.6	0
Beta19_MotivoSaude	-5.1	17.9	-0.284	0.776	0.275	-18.6	0
Beta1_Distancia	-3.46	0.8	-4.33	1.5e-05	0.816	-4.25	2.18e-05



Beta3_Distancia	-7.44	1.37	-5.43	5.66e-08	1.62	-4.6	4.25e-06
Beta3_MotivoIntegracao	5.25	9.26	0.567	0.571	1.04	5.03	4.82e-07
Beta3_PopulacaoOrigem	1.45	0.397	3.66	0.000249	0.415	3.51	0.000453
Beta2_Moradores	-5.89	1.49	-3.96	7.56e-05	1.73	-3.41	0.000649
Beta2_MotivoLazer	2.6	0.581	4.47	7.89e-06	0.563	4.62	3.92e-06
Beta2_MotivoTrabalho	2.09	0.508	4.12	3.8e-05	0.498	4.2	2.67e-05
Beta4_GraudeInstrucao	4.77	2.9	1.64	0.1	1.6	2.98	0.00288
Beta4_MotivoEstudo	-4.08	15.5	-0.263	0.793	0.477	-8.56	0
Beta4_MotivoResidencia	-9.06	17.5	-0.518	0.604	0.407	-22.3	0
Beta4_Trabalha	5.42	12	0.453	0.65	0.405	13.4	0
Beta5_Distancia	-4.91	0.806	-6.1	1.06e-09	0.983	-5	5.81e-07
Beta5_Moradores	-2.06	0.685	-3.01	0.00264	0.632	-3.26	0.00111
Beta5_MotivoIntegracao	-2.01	34.7	-0.0578	0.954	0.225	-8.94	0
Beta5_MotivoSaude	-5.68	16.3	-0.349	0.727	0.203	-28	0
Beta6_Distancia	-3.1	0.353	-8.78	0	0.375	-8.26	2.22e-16
Beta6_Moradores	-1.86	0.585	-3.18	0.00147	0.621	-3	0.00271
Beta6_MotivoEstudo	1.66	0.269	6.19	6.18e-10	0.273	6.08	1.21e-09
Beta6_MotivoIntegracao	-4	38.5	-0.104	0.917	0.32	-12.5	0
Beta6_MotivoLazer	-0.709	0.362	-1.96	0.0503	0.319	-2.23	0.026
Beta6_MotivoResidencia	-2.46	0.223	-11	0	0.224	-11	0
Beta6_PopulacaoOrigem	3.16	0.347	9.1	0	0.365	8.66	0
Beta7_Distancia	-5.41	0.796	-6.8	1.03e-11	0.836	-6.48	9.39e-11
Beta7_MotivoIntegracao	5.24	9.23	0.568	0.57	0.723	7.24	4.35e-13
Beta7_Renda	-2.64	0.502	-5.27	1.4e-07	0.413	-6.4	1.55e-10
Beta8_Distancia	-5.3	0.501	-10.6	0	0.515	-10.3	0
Beta8_Moradores	-1.63	0.519	-3.14	0.00168	0.496	-3.28	0.00102
Beta8_MotivoIntegracao	4.92	9.21	0.534	0.593	0.498	9.89	0
Beta8_MotivoResidencia	-2.08	0.182	-11.4	0	0.186	-11.2	0
Beta8_PopulacaoOrigem	0.79	0.376	2.1	0.0355	0.382	2.07	0.0384
Beta9_Distancia	-3.36	1.52	-2.21	0.027	1.02	-3.3	0.000972
Beta9_Genero	-1.49	0.64	-2.33	0.0198	0.616	-2.42	0.0155
Beta9_MotivoEstudo	-6.64	18.1	-0.366	0.714	0.288	-23	0
Beta9_MotivoResidencia	-3.73	1.05	-3.55	0.000386	1.01	-3.69	0.000226
Beta9_PopulacaoOrigem	3.02	0.649	4.66	3.22e-06	0.512	5.9	3.57e-09
Beta_PtosOnibus	3.62	0.398	9.11	0	0.373	9.72	0
CTE15	-16.6	10.2	-1.62	0.104	2.87	-5.8	6.68e-09
CTE16	2.48	0.422	5.86	4.62e-09	0.53	4.67	2.94e-06
CTE17	1.9	0.441	4.3	1.7e-05	0.401	4.73	2.22e-06
CTE4	-14.3	15.6	-0.918	0.359	1.4	-10.2	0
CTE7	3.14	0.322	9.75	0	0.301	10.4	0
CTE8	2.03	0.335	6.07	1.28e-09	0.342	5.94	2.92e-09



**APÊNDICE B - Calibração sem o procedimento por “macro Bairros” para nível de significância de 10%**

**Estimation report**

**Number of estimated parameters:** 112

**Sample size:** 2196

**Excluded observations:** 0

**Init log likelihood:** -6465.988

**Final log likelihood:** -4172.127

**Likelihood ratio test for the init. model:** 4587.722

**Rho-square for the init. model:** 0.355

**Rho-square-bar for the init. model:** 0.337

**Akaike Information Criterion:** 8572.254

**Bayesian Information Criterion:** 9221.414

**Final gradient norm:** 2.1696E-01

**Diagnostic:** b'CONVERGENCE:  
Norm\_of\_projected\_gradient\_<=\_PGTOL'

**Database readings:** 284

**Iterations:** 262

**Data processing time:** 0:00:00

**Optimization time:** 0:00:58.734376

**Nbr of threads:** 8

**Estimated parameters**

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
Beta10_Distancia	-2.93	0.817	-3.59	0.000333	0.791	-3.7	0.000212
Beta10_FrequenciaEscolar	-2.65	0.619	-4.28	1.89e-05	0.644	-4.12	3.86e-05
Beta10_GraudeInstrucao	-3.7	0.7	-5.28	1.28e-07	0.637	-5.8	6.51e-09
Beta10_Idade	-4.32	1.05	-4.12	3.8e-05	0.944	-4.57	4.79e-06
Beta10_Moradores	-2.69	0.986	-2.73	0.00639	0.909	-2.96	0.0031
Beta10_MotivoLazer	5.42	1.15	4.7	2.64e-06	1.08	5.01	5.31e-07
Beta10_MotivoResidencia	4.71	1.02	4.61	3.96e-06	0.898	5.24	1.59e-07
Beta10_MotivoSaude	6.07	1.17	5.17	2.32e-07	1.11	5.47	4.45e-08
Beta10_MotivoTrabalho	4.24	1.02	4.15	3.33e-05	0.876	4.84	1.32e-06
Beta10_PopulacaoOrigem	2.25	0.686	3.28	0.00105	0.801	2.81	0.00502
Beta11_Distancia	-5.36	0.616	-8.71	0	0.656	-8.18	2.22e-16
Beta11_GraudeInstrucao	1.18	0.304	3.87	0.000107	0.284	4.14	3.41e-05
Beta11_Moradores	-2.66	0.569	-4.68	2.9e-06	0.571	-4.66	3.13e-06

Beta11_MotivoEstudo	2.85	4.05	0.705	0.481	0.359	7.94	2e-15
Beta11_MotivoIntegracao	4.86	8.57	0.567	0.571	0.482	10.1	0
Beta11_MotivoResidencia	-1.35	0.184	-7.35	1.96e-13	0.183	-7.38	1.53e-13
Beta11_PopulacaoOrigem	1.12	0.339	3.3	0.000966	0.321	3.49	0.000489
Beta12_Distancia	-2.52	0.345	-7.31	2.7e-13	0.347	-7.26	3.95e-13
Beta12_GraudeInstrucao	1.37	0.226	6.03	1.64e-09	0.228	6	1.96e-09
Beta12_Moradores	-2.62	0.466	-5.62	1.86e-08	0.493	-5.32	1.02e-07
Beta12_MotivoEstudo	2.13	4.04	0.526	0.599	0.3	7.1	1.26e-12
Beta12_MotivoIntegracao	5.45	8.56	0.637	0.524	0.267	20.4	0
Beta12_MotivoResidencia	-2.64	0.149	-17.7	0	0.155	-17	0
Beta12_PopulacaoOrigem	1.1	0.282	3.91	9.16e-05	0.285	3.87	0.000107
Beta13_Distancia	-6.08	0.734	-8.28	2.22e-16	0.963	-6.31	2.7e-10
Beta13_Moradores	-1.29	0.552	-2.33	0.0197	0.545	-2.36	0.0181
Beta13_MotivoEstudo	2.78	4.05	0.687	0.492	0.357	7.79	6.88e-15
Beta13_MotivoIntegracao	-3.98	33.3	-0.119	0.905	0.242	-16.4	0
Beta13_MotivoResidencia	-1.57	0.195	-8.06	6.66e-16	0.2	-7.88	3.33e-15
Beta13_PopulacaoOrigem	2.68	0.291	9.22	0	0.295	9.08	0
Beta14_Distancia	-3.37	1.71	-1.97	0.0485	0.969	-3.47	0.000515
Beta14_Moradores	-2.31	1.54	-1.51	0.132	1.22	-1.89	0.0583
Beta14_MotivoEstudo	4.86	4.12	1.18	0.238	0.734	6.61	3.74e-11
Beta14_MotivoSaude	-2.86	12.4	-0.231	0.818	0.339	-8.46	0
Beta14_MotivoTrabalho	-4.95	13.3	-0.373	0.709	0.298	-16.6	0
Beta15_FrequenciaEscolar	6.11	6.55	0.932	0.351	0.789	7.73	1.04e-14
Beta15_GraudeInstrucao	8.72	4.09	2.13	0.033	3.16	2.76	0.00578
Beta15_MotivoLazer	5.37	4.85	1.11	0.268	1.1	4.9	9.49e-07
Beta15_MotivoResidencia	-6.22	37.4	-0.166	0.868	0.637	-9.77	0
Beta15_MotivoTrabalho	5.25	5	1.05	0.294	1.12	4.7	2.56e-06
Beta15_Renda	-4.4	3.32	-1.32	0.185	1.25	-3.53	0.000419
Beta15_Trabalha	-2.39	1.59	-1.51	0.132	1.02	-2.34	0.0193
Beta15_VeiculosMotorizados	7.84	5.06	1.55	0.121	1.65	4.74	2.14e-06
Beta16_Distancia	-15.9	2.9	-5.48	4.31e-08	4.67	-3.41	0.000661
Beta16_Moradores	-2.05	1.3	-1.58	0.114	1.17	-1.75	0.0801
Beta16_MotivoEstudo	3.7	4.09	0.904	0.366	0.679	5.45	5.01e-08
Beta16_MotivoIntegracao	-1.32	29.5	-0.0448	0.964	0.501	-2.64	0.0084
Beta16_MotivoLazer	-5.15	14.5	-0.354	0.723	0.328	-15.7	0
Beta16_MotivoSaude	-4.92	15	-0.327	0.743	0.373	-13.2	0
Beta16_Renda	-1.47	1.01	-1.46	0.144	0.661	-2.22	0.0262
Beta17_Distancia	-4.19	0.579	-7.24	4.42e-13	0.617	-6.79	1.11e-11
Beta17_Moradores	-3.29	0.819	-4.02	5.87e-05	0.807	-4.08	4.55e-05
Beta17_MotivoEstudo	2.97	4.07	0.73	0.465	0.496	5.98	2.25e-09
Beta17_MotivoIntegracao	-2.41	30.6	-0.0787	0.937	0.262	-9.21	0

Beta17_PopulacaoOrigem	1.61	0.5	3.23	0.00125	0.438	3.69	0.000226
Beta18_Distancia	-4	0.445	-8.98	0	0.496	-8.06	8.88e-16
Beta18_Moradores	-3.36	0.577	-5.82	6.01e-09	0.543	-6.18	6.59e-10
Beta18_MotivoEstudo	3.22	4.05	0.796	0.426	0.341	9.44	0
Beta18_MotivoIntegracao	-3.44	31.9	-0.108	0.914	0.253	-13.6	0
Beta18_PopulacaoOrigem	2.27	0.286	7.94	2e-15	0.291	7.8	6.44e-15
Beta18_Renda	-1.79	0.343	-5.22	1.76e-07	0.328	-5.47	4.5e-08
Beta19_Distancia	-2.3	1.46	-1.58	0.115	1.16	-1.99	0.047
Beta19_FrequenciaEscolar	-5.28	14.7	-0.359	0.719	0.432	-12.2	0
Beta19_MotivoResidencia	-8.16	15.7	-0.518	0.604	0.353	-23.1	0
Beta19_MotivoSaude	-3.91	14	-0.279	0.781	0.501	-7.81	5.55e-15
Beta1_Distancia	-4.23	0.881	-4.8	1.58e-06	0.935	-4.52	6.15e-06
Beta3_Distancia	-7.71	1.34	-5.76	8.37e-09	1.55	-4.97	6.79e-07
Beta3_MotivoEstudo	2.97	4.08	0.727	0.467	0.629	4.72	2.4e-06
Beta3_MotivoIntegracao	5.04	8.62	0.585	0.559	1.04	4.84	1.3e-06
Beta3_PopulacaoOrigem	1.16	0.403	2.88	0.00403	0.402	2.88	0.00396
Beta2_Moradores	-7.22	1.57	-4.59	4.33e-06	1.86	-3.88	0.000105
Beta2_MotivoEstudo	4.52	4.18	1.08	0.279	1.13	3.99	6.55e-05
Beta2_MotivoLazer	2.59	0.587	4.42	1e-05	0.559	4.64	3.51e-06
Beta2_MotivoTrabalho	2.09	0.516	4.05	5.19e-05	0.492	4.24	2.19e-05
Beta4_GraudeInstrucao	4.99	2.91	1.71	0.0866	1.61	3.1	0.00193
Beta4_MotivoEstudo	-2.78	28	-0.0991	0.921	0.516	-5.38	7.36e-08
Beta4_MotivoResidencia	-8.92	16.1	-0.555	0.579	0.407	-21.9	0
Beta4_Trabalha	5.12	10.3	0.499	0.618	0.405	12.6	0
Beta5_Distancia	-5	0.808	-6.18	6.27e-10	0.989	-5.05	4.33e-07
Beta5_Moradores	-2.71	0.734	-3.69	0.000223	0.683	-3.97	7.31e-05
Beta5_MotivoEstudo	1.75	4.16	0.42	0.674	1.02	1.71	0.0879
Beta5_MotivoIntegracao	-1.99	30.8	-0.0648	0.948	0.238	-8.37	0
Beta5_MotivoSaude	-5.46	14.6	-0.375	0.708	0.204	-26.7	0
Beta6_Distancia	-3.07	0.352	-8.71	0	0.374	-8.21	2.22e-16
Beta6_Moradores	-2.54	0.64	-3.96	7.39e-05	0.681	-3.73	0.000195
Beta6_MotivoEstudo	4.09	4.05	1.01	0.312	0.334	12.3	0
Beta6_MotivoIntegracao	-3.95	33.7	-0.117	0.907	0.327	-12.1	0
Beta6_MotivoLazer	-0.71	0.362	-1.96	0.0497	0.318	-2.23	0.0255
Beta6_MotivoResidencia	-2.48	0.225	-11	0	0.227	-10.9	0
Beta6_PopulacaoOrigem	3.1	0.349	8.88	0	0.368	8.42	0
Beta7_Distancia	-5.47	0.805	-6.79	1.11e-11	0.849	-6.44	1.18e-10
Beta7_MotivoEstudo	1.96	4.08	0.481	0.63	0.616	3.19	0.00143
Beta7_MotivoIntegracao	5.05	8.59	0.588	0.557	0.724	6.97	3.16e-12
Beta7_Renda	-2.68	0.508	-5.28	1.32e-07	0.424	-6.32	2.59e-10
Beta8_Distancia	-5.3	0.501	-10.6	0	0.516	-10.3	0

Beta8_Moradores	-2.27	0.575	-3.95	7.77e-05	0.555	-4.1	4.19e-05
Beta8_MotivoEstudo	2.56	4.05	0.631	0.528	0.352	7.26	3.77e-13
Beta8_MotivoIntegracao	4.71	8.57	0.55	0.582	0.504	9.35	0
Beta8_MotivoResidencia	-2.09	0.187	-11.2	0	0.192	-10.8	0
Beta8_PopulacaoOrigem	0.784	0.376	2.08	0.0371	0.382	2.05	0.0399
Beta9_Distancia	-3.53	1.53	-2.31	0.0211	1.04	-3.41	0.000638
Beta9_Genero	-1.53	0.637	-2.4	0.0165	0.612	-2.5	0.0126
Beta9_MotivoEstudo	-5.56	35.1	-0.159	0.874	0.338	-16.5	0
Beta9_MotivoResidencia	-3.76	1.06	-3.56	0.000367	1.02	-3.68	0.000236
Beta9_PopulacaoOrigem	2.72	0.669	4.06	4.86e-05	0.543	5.01	5.58e-07
Beta_PtosOnibus	3.49	0.441	7.91	2.44e-15	0.422	8.27	2.22e-16
CTE15	-15.9	9.07	-1.75	0.0795	2.92	-5.45	4.98e-08
CTE16	3.3	0.731	4.51	6.39e-06	0.598	5.52	3.47e-08
CTE17	1.76	0.458	3.85	0.000117	0.421	4.19	2.79e-05
CTE4	-14.3	13.7	-1.05	0.295	1.41	-10.2	0
CTE7	2.85	0.356	8.03	8.88e-16	0.342	8.35	0
CTE8	1.98	0.336	5.89	3.84e-09	0.343	5.77	8.12e-09

## APÊNDICE C – Detalhamento das calibrações

### Sem procedimento

#### Macrobairro

		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Variáveis explicativas**	A									-2,64					6,43				-6,61	
	B								-1,49											
	C			4,77							-3,65	1,19	1,39			8,59				
	D										-4,29									
	E	-5,89			-2,06	-1,86			-1,63		-2,03	-1,91	-2,03					-2,55	-2,69	
	F			4,08		1,66				-6,64					2,33					0,70
	G		5,25		-2,01	-4,00	5,24	4,92				5,04	5,70	-4,05				-1,38	-2,49	-3,48
	H	2,60				0,71					5,47					6,07	-5,32			
	I			-9,06		-2,46			-2,08	-3,73	4,78	-1,38	-2,61	1,60		-6,18				-8,70
	J				-5,68						6,13					-3,12		-5,16		-5,10
	K	2,09									4,31					-5,20	6,11			
	L						-2,64										-4,39			-1,77
	M			5,42													-2,53			
	N															8,29				
	O		1,45			3,16			0,79	3,02	2,28	1,16	1,07	2,58				1,61	2,28	

**Porcentagem adotada\*: 10%**

#### Macrobairro

		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Variáveis explicativas**	A		0,88							-2,61					7,35				-6,47	
	B								-1,54											
	C			3,25	-1,23			-1,20			-4,52							-1,97	-1,30	
	D		1,90								-4,29									
	E				-2,55	-1,60			-1,56		-2,54	-1,38	-2,28	-1,35				-3,01	-2,71	
	F			-4,53		1,32				-4,70			-0,71			-3,22				
	G	-3,52		-4,64	-6,51	-7,96				-3,92				-7,86	-5,64		-6,64	-7,31	-7,53	-5,06
	H	3,33		1,81		-0,64					8,34					6,17	-6,38			
	I			-7,67	1,29	-1,35	1,32	-0,96			8,75		-1,57			-4,42			1,21	-7,44
	J	-2,88		-3,14	-6,10					3,19	9,08					-5,22		-6,29		-4,89
	K	2,51								2,87	7,09					-6,93	5,28		-1,22	
	L				-0,93			-2,71									-1,44			-1,61
	M			6,62																
	N																			
	O	-2,10				2,10											-2,32		1,46	

**Porcentagem adotada\*: 20%**

#### Macrobairro

		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Variáveis explicativas**	A									-2,83					7,07				-6,63	
	B								-1,99											
	C				-1,70	-1,19	-1,41				-4,10							-2,33	-1,61	
	D										-4,29									
	E		2,41					1,98												4,53
	F			-4,38		1,90				-5,24	-6,28		-0,82			-3,31				
	G	-4,23		-4,36	-6,49	-7,59				-4,61	-5,86			-7,76	-5,57		-6,69	-6,69	-7,64	-4,60
	H	2,64		1,77												6,12	-6,32			
	I		0,96	-7,27	1,69			1,83			2,18		-1,04			-4,08		1,64	1,66	-7,63
	J	-3,71		-3,46	-6,19					2,52	1,71					-5,18		-5,78		-5,07
	K	1,96					1,03			2,14						-7,03	5,34			
	L				-1,15			-2,76												-1,77
	M			5,80																
	N																			
	O	-2,80				1,69											-2,73		1,13	

**Porcentagem adotada\*: 30%**

Macrobairro

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A									-0,97					7,16				-6,61
B								-1,53										
C			3,82						-2,38	1,67	1,66							
D																		
E		3,21				2,03												4,26
F			-4,24		1,69			-4,80	-5,74	-0,59				-3,66				
G	-3,07		-3,49	-6,19	-7,47			-3,64	-5,71		0,70	-8,23	-5,15		-6,55	-6,51	-7,47	-3,84
H	2,83		1,74											6,05	-6,32			
I		1,31	-7,18	1,80		1,89			2,23		-0,85			-4,28		1,69	1,78	-7,38
J	-3,30		-3,03	-6,06				2,75	1,73					-5,20		-5,99		-5,04
K	2,14							2,45						-7,26	5,27			
L				-1,27		-2,86												-1,82
M			6,05						0,94									
N																		
O	-2,39				1,85										-2,47		1,29	

**Porcentagem adotada\*: 40%**

Macrobairro

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A									-2,13					6,22				-6,33
B								-1,66										
C				-1,62		-1,38			-2,93					6,17		-1,96	-1,55	
D																		
E		2,78				1,95												5,01
F			-3,48		1,77			-4,64	-5,76					-3,97				
G	-2,90		-3,10	-5,99	-7,98			-3,57	-5,22		0,74	-7,80	-4,83		-6,88	-6,61	-7,60	-3,22
H	2,84		1,82											5,54	-6,70			
I		1,16	-7,13	1,79		1,90			2,22		-0,87			-4,72		1,73	1,75	-7,15
J	-3,07		-2,83	-5,99				2,40	1,33					-4,86		-6,26		-4,78
K	2,23							1,95						-7,08	5,76			
L		-0,99		-1,32		-2,99			-1,87					-3,90	-2,01		-1,96	-2,38
M			5,96											-2,51				
N			-6,93											8,10				
O	-2,17				1,37										-2,73		1,19	

**Porcentagem adotada\*: 50%**

Macrobairro

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A									-1,23					1,64				-6,22
B								-1,53										
C			3,46			-1,18			-2,72					2,53		-1,74	-1,25	
D														-13,00				
E		3,08				1,95												4,64
F			-3,72		1,76			-3,96	-5,58					-6,52				
G	-2,50		-3,19	-6,18	-7,32			-3,04	-5,34		0,73	-7,68	-5,07	-5,12	-6,41	-6,28	-7,10	-3,60
H	3,34		1,88												-6,28			
I		1,17	-6,61	1,59		1,90			2,22		-0,80			-7,15		1,73	1,79	-7,21
J	-2,62			-6,07				3,03	1,34					-4,96	-3,08	-5,91		-4,71
K	2,76					-2,98		2,61						-6,94				
L				-2,25					-1,85						-1,87		-1,93	
M			5,74															
N	-6,44																	
O	-1,64				1,74					0,56					-2,50		1,39	



**Porcentagem adotada\*: 60%**

**Macrobairro**

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A									-1,05					1,47				-5,87
B								-1,56										
C			2,79						-2,79					1,95		-1,86		
D														-12,30				
E		3,39				2,17												5,20
F			-3,54		1,59			-4,16	-5,21					-6,56				
G	-2,53		-2,90	-5,77	-7,45			-3,13	-4,61		0,92	-7,76	-5,01	-4,48	-6,20	-5,80	-6,64	
H	3,07		1,75												-6,28			
I			-6,35	1,80		2,08			2,40					-6,60		1,95	2,16	-6,33
J	-2,78			-5,88				2,62	1,46				-5,14	-2,83	-5,81			-4,55
K	2,48							2,22					-7,32					
L				-2,31		-3,35			-2,15						-1,82		-1,89	-2,01
M			5,35															
N	-6,69																	
O	-1,91														-2,86			

**Porcentagem adotada\*: 70%**

**Macrobairro**

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A														7,92				-6,29
B								-1,58										
C			3,15			-1,05			-2,68									
D																		
E		3,19																4,50
F			-4,88					-5,01	-7,17					-6,93				-2,99
G	-4,03		-4,57	-6,63	-8,63			-4,68	-6,44			-8,63	-6,37	-5,82	-7,13	-7,94	-8,65	-4,88
H	3,27		1,70												-6,45			
I			-7,24	1,82		1,67			1,51					-7,34				-7,57
J	-3,52		-3,42	-6,13				2,71					-5,84	-3,51	-6,10			-5,36
K	2,67							2,41					-7,71					
L						-2,42			-1,77									
M			-6,15															
N	-6,96																	
O	-2,00														-2,71			

**Porcentagem adotada\*: 80%**

**Macrobairro**

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A														8,98				-7,62
B								-1,41										
C			3,00						-3,11									
D																		
E		3,22																4,67
F			-6,93					-7,15	-8,52					-8,27				-4,54
G	-7,19		-6,41	-8,84	-9,44			-6,89	-8,10			-9,36	-7,76	-7,53	-7,97	-8,65	-9,37	-6,75
H	1,91		1,67			-2,10									-7,74			
I			-8,55	0,86										-8,91				-9,19
J	-6,89		-5,33	-8,52									-7,47	-5,76	-7,59			-7,22
K								2,29					-8,82					
L						-2,59												
M			7,85															
N	-6,19																	
O															-2,65			

**Porcentagem adotada\*: 90%**

Macrobairro

		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Variáveis explicativas**	A														7,38				-5,61	
	B																			
	C									-3,08										
	D																			
	E																			
	F			-4,34						-6,06	-7,69					-6,09				-2,38
	G	-4,99		-4,31	-7,16	-8,56				-5,78	-6,56			-8,47	-5,10	-4,89	-6,10	-7,60	-8,51	-3,87
	H																-6,09			
	I			-7,46						-2,40						-6,63				-6,67
	J	-4,52		-3,38	-6,57										-4,59		-5,42			-3,74
	K														-6,66					
	L																			-2,13
	M			4,86																
	N	-5,91							-2,02											4,84
	O																			

**Porcentagem adotada\*: 100%**

Macrobairro

		2	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Variáveis explicativas**	A																		
	B																		
	C																		
	D																		
	E																		
	F																		
	G																		
	H																		
	I																		
	J																		
	K																		
	L																		
	M																		
	N																		
	O																		

\* Porcentagem adotada como limite da variação proporcional entre os nós filhos obtidos na árvore de decisão para cada alternativa de destino.

\*\* A: Frequência escolar; B: Gênero; C: Grau de instrução; D: Idade; E: Moradores; F: Motivo estudo; G: Motivo integração; H: Motivo lazer; I: Motivo residência; J: Motivo saúde; K: Motivo trabalho; L: Renda; M: Trabalha; N: Veículos motorizados; O: População origem.

## APÊNDICE D - Calibração clássica por bairros

### Estimation report

**Number of estimated parameters:** 200

**Sample size:** 2196

**Excluded observations:** 0

**Init log likelihood:** -7807.544

**Final log likelihood:** -5404.475

**Likelihood ratio test for the init. model:** 4806.138

**Rho-square for the init. model:** 0.308

**Rho-square-bar for the init. model:** 0.282

**Akaike Information Criterion:** 11208.95

**Bayesian Information Criterion:** 12347.83

**Final gradient norm:** 2.9056E-01

**Diagnostic:** b'CONVERGENCE:  
NORM\_OF\_PROJECTED\_GRADIENT\_<=\_PGTOL'

**Database readings:** 316

**Iterations:** 295

**Data processing time:** 0:00:00

**Optimization time:** 0:03:07.935499

**Nbr of threads:** 8

### Estimated parameters

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
Beta10_Distancia	-2.91	1.2	-2.42	0.0153	0.954	-3.05	0.00226
Beta10_GraudeInstrucao	-2.06	0.671	-3.07	0.00217	0.724	-2.84	0.00446
Beta10_Idade	2.05	0.666	3.07	0.00214	0.558	3.67	0.000246
Beta10_Moradores	2.88	0.834	3.45	0.000563	0.809	3.56	0.000373
Beta10_MotivoLazer	-5.69	16.5	-0.346	0.73	0.364	-15.6	0
Beta10_MotivoResidencia	1.8	0.38	4.74	2.1e-06	0.373	4.83	1.35e-06
Beta10_Renda	-1.98	0.801	-2.48	0.0133	0.746	-2.66	0.00789
Beta11_Distancia	-5.34	1.38	-3.88	0.000104	1.5	-3.56	0.000368
Beta11_Moradores	2.7	0.807	3.34	0.000834	0.833	3.23	0.00122
Beta11_MotivoIntegracao	-5.59	17.2	-0.324	0.746	0.313	-17.9	0
Beta12_FrequenciaEscolar	1.43	0.617	2.31	0.0208	0.654	2.18	0.0292
Beta12_Moradores	-3.87	2.01	-1.92	0.0547	1.24	-3.12	0.0018
Beta12_MotivoIntegracao	-5.18	16.8	-0.309	0.757	0.402	-12.9	0
Beta12_MotivoResidencia	-7.36	16.2	-0.456	0.649	0.314	-23.5	0
Beta12_MotivoSaude	-4.73	16.3	-0.29	0.772	0.315	-15	0

Beta12_Renda	-3.52	1.27	-2.78	0.00545	1.25	-2.82	0.00478
Beta12_VeiculosMotorizados	2.83	1.89	1.49	0.136	1.15	2.45	0.0143
Beta13_Distancia	-2.97	1.27	-2.35	0.0188	0.966	-3.08	0.00208
Beta13_Idade	2.56	0.894	2.86	0.00428	0.961	2.66	0.00782
Beta13_Moradores	3.47	0.968	3.58	0.000345	1.04	3.32	0.000885
Beta13_MotivoEstudo	-5.4	15.9	-0.339	0.735	0.422	-12.8	0
Beta13_MotivoResidencia	1.26	0.363	3.46	0.000541	0.357	3.52	0.000429
Beta13_Renda	-0.855	0.756	-1.13	0.259	0.429	-1.99	0.0465
Beta14_FrequenciaEscolar	7.5	11.9	0.63	0.528	0.607	12.4	0
Beta14_MotivoEstudo	-5.91	17.3	-0.341	0.733	0.627	-9.42	0
Beta14_MotivoIntegracao	-4.14	16.5	-0.251	0.802	0.695	-5.97	2.44e-09
Beta14_MotivoResidencia	-6.33	16.5	-0.383	0.702	0.613	-10.3	0
Beta14_PopulacaoOrigem	-2.39	2.41	-0.991	0.322	0.876	-2.73	0.00626
Beta14_VeiculosMotorizados	3.64	3.95	0.923	0.356	1.58	2.31	0.0208
Beta15_Distancia	-5.78	0.665	-8.7	0	0.826	-7	2.53e-12
Beta15_GraudeInstrucao	-0.934	0.415	-2.25	0.0245	0.387	-2.41	0.016
Beta15_MotivoIntegracao	-6.72	17	-0.395	0.693	0.377	-17.8	0
Beta15_MotivoResidencia	1.98	0.248	8	1.33e-15	0.228	8.69	0
Beta15_MotivoTrabalho	-1.24	0.39	-3.17	0.00154	0.394	-3.14	0.00169
Beta15_PopulacaoOrigem	1.87	0.382	4.9	9.61e-07	0.38	4.92	8.59e-07
Beta15_Renda	-1.86	0.446	-4.17	3.05e-05	0.422	-4.41	1.06e-05
Beta15_VeiculosMotorizados	1.27	0.666	1.91	0.0565	0.558	2.28	0.0229
Beta16_Distancia	-7.17	1.64	-4.39	1.15e-05	1.39	-5.16	2.46e-07
Beta16_MotivoEstudo	-5.96	16.2	-0.368	0.713	0.243	-24.5	0
Beta17_Distancia	-4.44	1.6	-2.78	0.00539	2.01	-2.21	0.0274
Beta17_Moradores	5.17	1.11	4.65	3.38e-06	1.19	4.34	1.41e-05
Beta17_MotivoEstudo	-3.88	15	-0.259	0.796	0.767	-5.06	4.12e-07
Beta17_MotivoIntegracao	-2.78	14.7	-0.189	0.85	0.829	-3.36	0.000788
Beta17_MotivoLazer	2.23	1.05	2.12	0.034	1.09	2.06	0.0398
Beta17_MotivoResidencia	3.24	0.801	4.04	5.4e-05	0.809	4	6.37e-05
Beta17_MotivoSaude	-3.63	15.2	-0.238	0.812	0.832	-4.36	1.3e-05
Beta18_Distancia	-4.72	0.921	-5.13	2.95e-07	0.972	-4.86	1.18e-06
Beta18_PopulacaoOrigem	-0.928	0.413	-2.24	0.0248	0.4	-2.32	0.0202
Beta18_Renda	1.41	0.445	3.16	0.00157	0.495	2.84	0.00448
Beta18_Trabalha	-0.604	0.22	-2.75	0.00594	0.219	-2.76	0.00581
Beta19_Distancia	-8.77	2.05	-4.27	1.91e-05	3.18	-2.76	0.0058
Beta19_Moradores	3.34	0.759	4.4	1.09e-05	1.02	3.26	0.00113
Beta19_MotivoIntegracao	-6.09	17	-0.358	0.72	0.333	-18.3	0
Beta19_MotivoTrabalho	-1.69	0.608	-2.79	0.00533	0.635	-2.67	0.00758
Beta1_Distancia	-3.47	0.857	-4.05	5.02e-05	0.743	-4.67	2.97e-06
Beta20_Distancia	-5.33	0.763	-6.98	2.99e-12	0.858	-6.2	5.48e-10
Beta20_GraudeInstrucao	1.55	0.257	6.03	1.67e-09	0.262	5.9	3.61e-09

Beta20_Moradores	1.66	0.508	3.26	0.00111	0.531	3.12	0.00181
Beta20_MotivoResidencia	0.854	0.188	4.54	5.7e-06	0.185	4.63	3.74e-06
Beta21_Distancia	-3.51	0.766	-4.59	4.54e-06	0.753	-4.66	3.11e-06
Beta21_MotivoLazer	1.02	0.342	2.98	0.00287	0.341	2.99	0.00283
Beta21_MotivoResidencia	-0.895	0.405	-2.21	0.0271	0.378	-2.37	0.0178
Beta22_Distancia	-4.24	0.722	-5.88	4.11e-09	0.714	-5.94	2.77e-09
Beta22_GraudeInstrucao	2.02	0.191	10.6	0	0.191	10.6	0
Beta22_MotivoTrabalho	-0.984	0.253	-3.88	0.000103	0.255	-3.86	0.000115
Beta23_Distancia	-6.89	1.33	-5.19	2.07e-07	1.56	-4.42	9.88e-06
Beta23_FrequenciaEscolar	1.08	0.342	3.17	0.00155	0.328	3.3	0.000973
Beta23_MotivoIntegracao	-5.84	17.4	-0.335	0.738	0.402	-14.5	0
Beta23_MotivoLazer	-6.06	15.9	-0.382	0.702	0.315	-19.3	0
Beta23_MotivoResidencia	1	0.335	2.99	0.0028	0.348	2.87	0.00407
Beta23_MotivoSaude	-5.17	16.3	-0.317	0.751	0.291	-17.7	0
Beta24_Distancia	-8.62	2.09	-4.13	3.59e-05	2.39	-3.61	0.000308
Beta24_Idade	2	0.535	3.73	0.00019	0.519	3.85	0.000119
Beta24_Moradores	1.82	0.813	2.24	0.0252	0.875	2.08	0.0376
Beta24_MotivoIntegracao	-5.99	17.2	-0.348	0.728	0.31	-19.3	0
Beta24_VeiculosMotorizados	-2.94	1.3	-2.26	0.0237	1.2	-2.46	0.0141
Beta25_Distancia	-5.8	1.09	-5.33	9.69e-08	1.36	-4.25	2.13e-05
Beta25_MotivoIntegracao	-5.84	17.2	-0.34	0.734	0.326	-17.9	0
Beta25_MotivoLazer	-6.5	15.9	-0.408	0.683	0.21	-30.9	0
Beta25_MotivoResidencia	0.864	0.267	3.24	0.00121	0.254	3.4	0.000677
Beta25_MotivoSaude	-5.56	16.2	-0.344	0.731	0.259	-21.5	0
Beta26_Distancia	-4.35	0.956	-4.55	5.25e-06	1.02	-4.29	1.82e-05
Beta26_GraudeInstrucao	-2.68	0.67	-4	6.36e-05	0.622	-4.31	1.67e-05
Beta26_MotivoEstudo	-3.1	47.2	-0.0657	0.948	0.221	-14	0
Beta26_MotivoIntegracao	-1.85	40.8	-0.0454	0.964	0.328	-5.65	1.64e-08
Beta26_MotivoLazer	6.09	7.71	0.789	0.43	0.589	10.3	0
Beta26_MotivoResidencia	7.35	7.69	0.955	0.339	0.249	29.5	0
Beta26_MotivoSaude	6.6	7.71	0.856	0.392	0.544	12.1	0
Beta26_MotivoTrabalho	4.72	7.7	0.613	0.54	0.493	9.57	0
Beta26_PopulacaoOrigem	1.88	0.679	2.77	0.00563	0.79	2.38	0.0173
Beta26_Renda	-1.69	0.73	-2.32	0.0205	0.714	-2.37	0.0179
Beta27_Distancia	-4.95	1.12	-4.42	9.93e-06	1.48	-3.34	0.000838
Beta27_MotivoIntegracao	-5.46	17.4	-0.313	0.754	0.336	-16.2	0
Beta27_MotivoResidencia	1.67	0.273	6.13	8.89e-10	0.325	5.15	2.54e-07
Beta28_Distancia	-8.14	1.54	-5.29	1.21e-07	1.39	-5.86	4.75e-09
Beta28_MotivoSaude	-4.87	15.7	-0.31	0.757	0.272	-17.9	0
Beta29_FrequenciaEscolar	-6.32	16.7	-0.379	0.705	0.544	-11.6	0
Beta29_Moradores	5.19	2.39	2.17	0.03	1.64	3.16	0.00158
Beta29_MotivoIntegracao	-2.88	14.9	-0.192	0.847	0.806	-3.57	0.00036

Beta29_MotivoLazer	2.48	1.04	2.39	0.0167	1.03	2.4	0.0163
Beta29_MotivoResidencia	-5.37	15.5	-0.346	0.729	0.734	-7.31	2.64e-13
Beta29_MotivoSaude	-3.75	15.7	-0.238	0.812	0.761	-4.93	8.32e-07
Beta2_Distancia	-2.3	2.24	-1.03	0.303	1.1	-2.09	0.0369
Beta2_GraudeInstrucao	5.23	3.01	1.74	0.0821	1.62	3.23	0.00126
Beta2_MotivoEstudo	-3.41	15.3	-0.223	0.824	0.527	-6.47	1.01e-10
Beta2_MotivoIntegracao	-3.8	15.8	-0.24	0.811	0.589	-6.44	1.2e-10
Beta2_MotivoLazer	2.05	0.786	2.61	0.00919	0.752	2.72	0.00648
Beta2_MotivoResidencia	-6.08	15.7	-0.388	0.698	0.496	-12.3	0
Beta2_MotivoSaude	-2.98	15.3	-0.195	0.845	0.575	-5.19	2.16e-07
Beta2_Trabalha	6.21	14.1	0.441	0.659	0.463	13.4	0
Beta30_Distancia	-3.34	0.924	-3.61	0.000302	1.09	-3.05	0.00225
Beta30_MotivoIntegracao	-6.32	17.2	-0.369	0.712	0.331	-19.1	0
Beta30_MotivoTrabalho	-0.927	0.393	-2.36	0.0181	0.417	-2.22	0.0262
Beta31_Distancia	-12.1	1.9	-6.4	1.58e-10	2.74	-4.43	9.55e-06
Beta31_FrequenciaEscolar	-0.642	0.333	-1.93	0.0542	0.294	-2.18	0.029
Beta31_Moradores	1.46	0.83	1.76	0.078	0.687	2.13	0.0331
Beta31_MotivoIntegracao	-6.06	17.2	-0.353	0.724	0.408	-14.9	0
Beta31_MotivoResidencia	1.12	0.327	3.43	0.000609	0.32	3.5	0.000459
Beta31_MotivoSaude	-6.55	16.9	-0.387	0.699	0.398	-16.5	0
Beta31_MotivoTrabalho	-1.54	0.655	-2.36	0.0185	0.63	-2.45	0.0144
Beta31_VeiculosMotorizados	-3.23	1.16	-2.79	0.00529	1.13	-2.85	0.00439
Beta32_Distancia	-5.65	1.23	-4.6	4.13e-06	1.52	-3.71	0.000208
Beta32_Moradores	4.01	0.599	6.7	2.11e-11	0.598	6.71	2e-11
Beta32_MotivoResidencia	2.07	0.306	6.78	1.23e-11	0.314	6.6	4.15e-11
Beta32_Renda	-3.21	0.654	-4.9	9.47e-07	0.626	-5.12	3.08e-07
Beta33_Distancia	-2.92	0.951	-3.08	0.0021	0.849	-3.45	0.000569
Beta33_MotivoIntegracao	-5.24	17	-0.308	0.758	0.392	-13.4	0
Beta33_MotivoSaude	-5.01	16.4	-0.305	0.761	0.314	-16	0
Beta33_MotivoTrabalho	-7.34	16.1	-0.457	0.648	0.282	-26	0
Beta34_Distancia	-5.75	0.65	-8.84	0	0.742	-7.74	9.55e-15
Beta34_GraudeInstrucao	-1.53	0.488	-3.15	0.00165	0.483	-3.18	0.0015
Beta34_MotivoIntegracao	-6.21	17.1	-0.363	0.717	0.391	-15.9	0
Beta34_MotivoResidencia	1.64	0.295	5.55	2.9e-08	0.288	5.69	1.25e-08
Beta34_MotivoTrabalho	-1.22	0.438	-2.78	0.00543	0.457	-2.67	0.00769
Beta34_PopulacaoOrigem	2.82	0.617	4.57	4.79e-06	0.672	4.2	2.67e-05
Beta35_MotivoIntegracao	-5.54	17.9	-0.31	0.756	0.319	-17.4	0
Beta35_MotivoLazer	1.33	0.55	2.42	0.0155	0.547	2.44	0.0149
Beta35_MotivoSaude	-5.34	17.1	-0.312	0.755	0.235	-22.7	0
Beta35_VeiculosMotorizados	-11	2.84	-3.87	0.000109	2.97	-3.7	0.000219
Beta5_Distancia	-7.79	1.62	-4.8	1.56e-06	2.48	-3.14	0.00168
Beta5_GraudeInstrucao	-1.59	0.711	-2.24	0.0252	0.703	-2.26	0.0237

Beta5_Idade	-2.18	0.892	-2.45	0.0144	0.652	-3.35	0.000816
Beta5_Moradores	-3.05	1.19	-2.56	0.0104	0.902	-3.39	0.00071
Beta5_MotivoEstudo	-2.23	43.6	-0.0511	0.959	0.322	-6.91	4.85e-12
Beta5_MotivoIntegracao	-1.75	36.4	-0.0479	0.962	0.393	-4.44	8.85e-06
Beta5_MotivoLazer	6.6	8.55	0.772	0.44	0.506	13.1	0
Beta5_MotivoResidencia	7.48	8.54	0.877	0.381	0.208	36	0
Beta5_MotivoSaude	-1.25	49.1	-0.0255	0.98	0.274	-4.57	4.98e-06
Beta5_MotivoTrabalho	4.33	8.56	0.506	0.613	0.594	7.3	2.97e-13
Beta5_Trabalha	0.942	0.421	2.24	0.0251	0.389	2.42	0.0155
Beta3_Distancia	-4.84	0.924	-5.24	1.61e-07	0.959	-5.05	4.41e-07
Beta3_MotivoEstudo	1.21	0.353	3.42	0.000635	0.355	3.4	0.000676
Beta3_MotivoIntegracao	-6.27	17.3	-0.362	0.717	0.285	-22	0
Beta3_MotivoLazer	1.33	0.31	4.31	1.61e-05	0.321	4.16	3.18e-05
Beta4_Distancia	-1.81	0.255	-7.1	1.26e-12	0.257	-7.04	1.87e-12
Beta4_MotivoEstudo	1.67	0.252	6.64	3.1e-11	0.269	6.21	5.24e-10
Beta4_MotivoIntegracao	-7.53	16.8	-0.449	0.653	0.389	-19.4	0
Beta4_PopulacaoOrigem	1.04	0.238	4.37	1.25e-05	0.241	4.32	1.58e-05
Beta6_Idade	2.96	1.34	2.21	0.0273	1.03	2.88	0.00393
Beta6_MotivoEstudo	-1.42	64.4	-0.0221	0.982	0.338	-4.2	2.62e-05
Beta6_MotivoLazer	-2.07	52.9	-0.039	0.969	0.14	-14.7	0
Beta6_MotivoResidencia	6.91	10.7	0.643	0.52	0.314	22	0
Beta6_MotivoSaude	-1.53	51.7	-0.0296	0.976	0.193	-7.92	2.44e-15
Beta6_MotivoTrabalho	4.37	10.8	0.405	0.685	1.01	4.32	1.56e-05
Beta6_Renda	-1.36	1.19	-1.15	0.251	0.597	-2.28	0.0224
Beta7_Distancia	-7.05	2.55	-2.76	0.00574	1.62	-4.35	1.39e-05
Beta7_MotivoIntegracao	-5.87	17.4	-0.337	0.736	0.348	-16.9	0
Beta7_MotivoLazer	-5.98	16.1	-0.372	0.71	0.262	-22.8	0
Beta7_MotivoSaude	-5.58	16.7	-0.335	0.738	0.306	-18.2	0
Beta7_PopulacaoOrigem	-5.12	1.32	-3.88	0.000104	1.72	-2.97	0.00296
Beta7_Renda	-2.03	0.983	-2.07	0.0389	0.776	-2.62	0.00886
Beta8_Distancia	-1.77	0.342	-5.17	2.36e-07	0.337	-5.25	1.54e-07
Beta8_GraudeInstrucao	1.7	0.222	7.68	1.6e-14	0.23	7.4	1.37e-13
Beta8_Idade	1.2	0.265	4.52	6.24e-06	0.275	4.35	1.34e-05
Beta8_MotivoIntegracao	0.701	0.315	2.23	0.026	0.319	2.2	0.0279
Beta8_MotivoResidencia	-1.06	0.156	-6.79	1.1e-11	0.158	-6.72	1.85e-11
Beta8_MotivoTrabalho	-0.585	0.137	-4.27	1.98e-05	0.141	-4.15	3.36e-05
Beta9_FrequenciaEscolar	-1.56	1.43	-1.09	0.274	0.671	-2.33	0.0199
Beta9_MotivoEstudo	6.82	9.08	0.751	0.453	0.669	10.2	0
Beta9_MotivoLazer	-2.29	51.7	-0.0442	0.965	0.123	-18.6	0
Beta9_MotivoResidencia	4.01	9.04	0.444	0.657	1.02	3.94	8.15e-05
Beta9_MotivoSaude	-1.66	50.6	-0.0328	0.974	0.158	-10.5	0
Beta9_MotivoTrabalho	5.59	8.99	0.621	0.534	0.417	13.4	0

Beta_PontosdeOnibus	1.67	0.283	5.9	3.72e-09	0.281	5.92	3.13e-09
CTE13	-2.41	0.791	-3.04	0.00235	0.774	-3.11	0.00189
CTE14	-8	11.9	-0.671	0.502	0.634	-12.6	0
CTE15	1.35	0.379	3.56	0.000365	0.371	3.64	0.000269
CTE16	0.962	0.398	2.42	0.0157	0.36	2.67	0.00751
CTE17	-3.48	1.03	-3.39	0.000701	1.15	-3.04	0.00237
CTE18	1.54	0.381	4.03	5.55e-05	0.411	3.74	0.000181
CTE2	-11.2	14.3	-0.785	0.433	1.55	-7.24	4.61e-13
CTE26	-3.6	7.71	-0.468	0.64	0.517	-6.97	3.1e-12
CTE29	-4.12	1.28	-3.21	0.00133	1.04	-3.96	7.5e-05
CTE30	0.693	0.312	2.22	0.0264	0.366	1.89	0.0584
CTE31	2.22	0.447	4.97	6.57e-07	0.443	5.02	5.29e-07
CTE34	1.35	0.481	2.82	0.00486	0.441	3.07	0.00216
CTE36	-3.27	8.57	-0.381	0.703	0.769	-4.24	2.19e-05
CTE6	-7.86	10.8	-0.729	0.466	0.559	-14.1	0
CTE7	3.1	0.649	4.77	1.83e-06	0.674	4.6	4.25e-06
CTE9	-6.7	8.98	-0.746	0.456	0.146	-45.8	0



**APÊNDICE E - Calibração por Algoritmos Genéticos que gerou o menor valor para a  
função objetivo no cenário livre**

---


$$V_{1j} = -13,72 \cdot Distância_{1j} + 1,96 \cdot PontosdeOnibus_{1j}$$


---

$$V_{2j} = 33,01 - 38,75 \cdot Moradores_j + 45,71 \cdot VeiculosMotorizados_j - 46,55 \cdot Renda_j - 24,22 \cdot Genero_j + 20,81 \cdot Idade_j - 36,93 \cdot FrequenciaEscolar_j + 26,10 \cdot Trabalha_j - 36,97 \cdot GraudeInstrução_j - 0,92 \cdot MotivoTrabalho_j - 39,44 \cdot MotivoEstudo_j - 11,99 \cdot MotivoLazer_j - 18,34 \cdot MotivoResidência_j - 33,11 \cdot MotivoSaúde_j - 7,07 \cdot MotivoIntegração_j + 36,51 \cdot PopulaçãoOrigem_j + 21,42 \cdot Distância_{2j} + 1,96 \cdot PontosdeOnibus_{2j}$$


---

$$V_{3j} = 44,41 + 35,61 \cdot Moradores_j - 24,75 \cdot VeiculosMotorizados_j - 38,10 \cdot Renda_j + 48,45 \cdot Genero_j + 40,67 \cdot Idade_j + 21,59 \cdot FrequenciaEscolar_j + 28,93 \cdot Trabalha_j + 33,62 \cdot GraudeInstrução_j + 19,38 \cdot MotivoTrabalho_j + 9,82 \cdot MotivoEstudo_j - 10,18 \cdot MotivoLazer_j + 28,61 \cdot MotivoResidência_j - 25,23 \cdot MotivoSaúde_j - 10,44 \cdot MotivoIntegração_j + 4,17 \cdot PopulaçãoOrigem_j - 9,09 \cdot Distância_{3j} + 1,96 \cdot PontosdeOnibus_{3j}$$


---

$$V_{4j} = 50,00 + 19,63 \cdot Moradores_j + 31,67 \cdot VeiculosMotorizados_j - 0,01 \cdot Renda_j + 23,80 \cdot Genero_j + 2,44 \cdot Idade_j + 44,57 \cdot FrequenciaEscolar_j + 2,83 \cdot Trabalha_j + 29,81 \cdot GraudeInstrução_j + 48,10 \cdot MotivoTrabalho_j - 27,29 \cdot MotivoEstudo_j - 27,72 \cdot MotivoLazer_j + 48,23 \cdot MotivoResidência_j + 26,92 \cdot MotivoSaúde_j + 38,53 \cdot MotivoIntegração_j + 40,02 \cdot PopulaçãoOrigem_j + 17,43 \cdot Distância_{4j} + 1,96 \cdot PontosdeOnibus_{4j}$$


---

$$V_{5j} = -6,60 + 24,32 \cdot Moradores_j - 1,95 \cdot VeiculosMotorizados_j - 6,00 \cdot Renda_j + 26,11 \cdot Genero_j + 47,05 \cdot Idade_j + 1,72 \cdot FrequenciaEscolar_j + 45,24 \cdot Trabalha_j + 48,57 \cdot GraudeInstrução_j - 28,40 \cdot MotivoTrabalho_j - 15,67 \cdot MotivoEstudo_j + 28,23 \cdot MotivoLazer_j + 1,22 \cdot MotivoResidência_j + 33,62 \cdot MotivoSaúde_j + 24,17 \cdot MotivoIntegração_j + 23,35 \cdot PopulaçãoOrigem_j + 40,99 \cdot Distância_{5j} + 1,96 \cdot PontosdeOnibus_{5j}$$


---

$$V_{6j} = 50,00 - 11,97 \cdot Moradores_j + 22,36 \cdot VeiculosMotorizados_j + 29,63 \cdot Renda_j - 19,90 \cdot Genero_j + 24,25 \cdot Idade_j + 46,01 \cdot FrequenciaEscolar_j + 44,74 \cdot Trabalha_j - 28,07 \cdot GraudeInstrução_j + 24,89 \cdot MotivoTrabalho_j + 33,40 \cdot MotivoEstudo_j + 44,51 \cdot MotivoLazer_j - 26,67 \cdot MotivoResidência_j - 33,71 \cdot MotivoSaúde_j + 29,09 \cdot MotivoIntegração_j + 46,14 \cdot PopulaçãoOrigem_j + 33,21 \cdot Distância_{6j} + 1,96 \cdot PontosdeOnibus_{6j}$$


---

$$V_{7j} = -6,68 + 39,56 \cdot Moradores_j - 45,07 \cdot VeiculosMotorizados_j - 39,82 \cdot Renda_j - 48,07 \cdot Genero_j - 12,89 \cdot Idade_j - 19,27 \cdot FrequenciaEscolar_j + 34,28 \cdot Trabalha_j - 9,68 \cdot GraudeInstrução_j + 40,07 \cdot MotivoTrabalho_j - 15,18 \cdot MotivoEstudo_j - 29,57 \cdot MotivoLazer_j + 8,66 \cdot MotivoResidência_j - 42,16 \cdot MotivoSaúde_j + 43,77 \cdot MotivoIntegração_j + 9,28 \cdot PopulaçãoOrigem_j + 34,38 \cdot Distância_{7j} + 1,96 \cdot PontosdeOnibus_{7j}$$


---

$$V_{8j} = 38,26 + 5,09 \cdot Moradores_j + 8,09 \cdot VeiculosMotorizados_j + 28,94 \cdot Renda_j + 11,41 \cdot Genero_j + 33,89 \cdot Idade_j + 47,46 \cdot FrequenciaEscolar_j + 31,98 \cdot Trabalha_j + 38,90 \cdot GraudeInstrução_j + 42,40 \cdot MotivoTrabalho_j - 14,26 \cdot MotivoEstudo_j + 32,54 \cdot MotivoLazer_j - 27,64 \cdot MotivoResidência_j + 47,59 \cdot MotivoSaúde_j + 35,75 \cdot MotivoIntegração_j + 42,5 \cdot PopulaçãoOrigem_j - 35,29 \cdot Distância_{8j} + 1,96 \cdot PontosdeOnibus_{8j}$$


---

$$V_{9j} = -6,48 - 37,79 \cdot Moradores_j + 34,62 \cdot VeiculosMotorizados_j + 42,57 \cdot Renda_j + 38,80 \cdot Genero_j + 46,77 \cdot Idade_j - 14,78 \cdot FrequenciaEscolar_j + 38,98 \cdot Trabalha_j + 8,65 \cdot GraudeInstrução_j - 29,38 \cdot MotivoTrabalho_j - 23,10 \cdot MotivoEstudo_j - 41,77 \cdot MotivoLazer_j - 20,56 \cdot MotivoResidência_j + 19,81 \cdot MotivoSaúde_j + 36,47 \cdot MotivoIntegração_j + 10,30 \cdot PopulaçãoOrigem_j - 18,53 \cdot Distância_{9j} + 1,96 \cdot PontosdeOnibus_{9j}$$


---

---


$$V_{10j} = -15,09 - 49,04 \cdot \text{Moradores}_j + 42,43 \cdot \text{VeiculosMotorizados}_j + 9,28 \cdot \text{Renda}_j + 43,08$$

- $\text{Genero}_j + 8,41 \cdot \text{Idade}_j - 21,27 \cdot \text{FrequenciaEscolar}_j - 49,86 \cdot \text{Trabalha}_j - 9,14$
- $\text{GraudeInstrução}_j + 49,78 \cdot \text{MotivoTrabalho}_j - 6,05 \cdot \text{MotivoEstudo}_j - 5,83$
- $\text{MotivoLazer}_j - 23,57 \cdot \text{MotivoResidência}_j - 4,65 \cdot \text{MotivoSaúde}_j + 16,79$
- $\text{MotivoIntegração}_j - 24,56 \cdot \text{PopulaçãoOrigem}_j - 12,32 \cdot \text{Distância}_{10j} + 1,96$
- $\text{PontosdeOnibus}_{10j}$

---


$$V_{11j} = -28,79 - 31,60 \cdot \text{Moradores}_j - 11,50 \cdot \text{VeiculosMotorizados}_j + 2,36 \cdot \text{Renda}_j + 6,24$$

- $\text{Genero}_j + 33,18 \cdot \text{Idade}_j + 6,70 \cdot \text{FrequenciaEscolar}_j + 4,46 \cdot \text{Trabalha}_j + 6,09$
- $\text{GraudeInstrução}_j - 44,95 \cdot \text{MotivoTrabalho}_j + 12,62 \cdot \text{MotivoEstudo}_j - 13,82$
- $\text{MotivoLazer}_j - 8,09 \cdot \text{MotivoResidência}_j + 3,55 \cdot \text{MotivoSaúde}_j - 41,72$
- $\text{MotivoIntegração}_j - 20,08 \cdot \text{PopulaçãoOrigem}_j + 23,31 \cdot \text{Distância}_{11j} + 1,96$
- $\text{PontosdeOnibus}_{11j}$

---


$$V_{12j} = 6,75 - 33,25 \cdot \text{Moradores}_j - 37,43 \cdot \text{VeiculosMotorizados}_j + 38,74 \cdot \text{Renda}_j - 24,58$$

- $\text{Genero}_j + 28,84 \cdot \text{Idade}_j - 28,54 \cdot \text{FrequenciaEscolar}_j + 39,54 \cdot \text{Trabalha}_j$
- $- 37,61 \cdot \text{GraudeInstrução}_j + 3,31 \cdot \text{MotivoTrabalho}_j - 15,71 \cdot \text{MotivoEstudo}_j$
- $- 33,65 \cdot \text{MotivoLazer}_j - 43,21 \cdot \text{MotivoResidência}_j - 7,44 \cdot \text{MotivoSaúde}_j + 40,87$
- $\text{MotivoIntegração}_j + 43,11 \cdot \text{PopulaçãoOrigem}_j + 12,23 \cdot \text{Distância}_{12j} + 1,96$
- $\text{PontosdeOnibus}_{12j}$

---


$$V_{13j} = 30,50 + 37,52 \cdot \text{Moradores}_j + 20,92 \cdot \text{VeiculosMotorizados}_j + 8,99 \cdot \text{Renda}_j + 20,89$$

- $\text{Genero}_j + 40,66 \cdot \text{Idade}_j + 40,76 \cdot \text{FrequenciaEscolar}_j + 30,22 \cdot \text{Trabalha}_j$
- $- 33,64 \cdot \text{GraudeInstrução}_j + 20,37 \cdot \text{MotivoTrabalho}_j - 8,88 \cdot \text{MotivoEstudo}_j$
- $+ 31,88 \cdot \text{MotivoLazer}_j + 27,43 \cdot \text{MotivoResidência}_j - 34,35 \cdot \text{MotivoSaúde}_j$
- $- 41,83 \cdot \text{MotivoIntegração}_j - 14,98 \cdot \text{PopulaçãoOrigem}_j - 4,40 \cdot \text{Distância}_{13j}$
- $+ 1,96 \cdot \text{PontosdeOnibus}_{13j}$

---


$$V_{14j} = -44,27 + 37,18 \cdot \text{Moradores}_j - 46,28 \cdot \text{VeiculosMotorizados}_j - 21,67 \cdot \text{Renda}_j + 24,65$$

- $\text{Genero}_j + 18,67 \cdot \text{Idade}_j - 0,02 \cdot \text{FrequenciaEscolar}_j - 21,12 \cdot \text{Trabalha}_j$
- $- 32,39 \cdot \text{GraudeInstrução}_j + 45,60 \cdot \text{MotivoTrabalho}_j + 1,12 \cdot \text{MotivoEstudo}_j$
- $+ 19,02 \cdot \text{MotivoLazer}_j + 19,98 \cdot \text{MotivoResidência}_j - 1,42 \cdot \text{MotivoSaúde}_j + 45,11$
- $\text{MotivoIntegração}_j + 13,08 \cdot \text{PopulaçãoOrigem}_j - 6,98 \cdot \text{Distância}_{14j} + 1,96$
- $\text{PontosdeOnibus}_{14j}$

---


$$V_{15j} = 46,77 + 27,95 \cdot \text{Moradores}_j - 35,24 \cdot \text{VeiculosMotorizados}_j + 29,32 \cdot \text{Renda}_j + 31,93$$

- $\text{Genero}_j + 22,24 \cdot \text{Idade}_j + 12,11 \cdot \text{FrequenciaEscolar}_j + 22,95 \cdot \text{Trabalha}_j + 0,79$
- $\text{GraudeInstrução}_j + 24,02 \cdot \text{MotivoTrabalho}_j + 15,07 \cdot \text{MotivoEstudo}_j + 45,58$
- $\text{MotivoLazer}_j + 29,08 \cdot \text{MotivoResidência}_j - 42,16 \cdot \text{MotivoSaúde}_j + 19,14$
- $\text{MotivoIntegração}_j + 28,58 \cdot \text{PopulaçãoOrigem}_j - 23,79 \cdot \text{Distância}_{15j} + 1,96$
- $\text{PontosdeOnibus}_{15j}$

---


$$V_{16j} = -41,90 + 39,36 \cdot \text{Moradores}_j - 38,39 \cdot \text{VeiculosMotorizados}_j - 12,84 \cdot \text{Renda}_j - 45,75$$

- $\text{Genero}_j - 47,48 \cdot \text{Idade}_j + 15,52 \cdot \text{FrequenciaEscolar}_j - 18,54 \cdot \text{Trabalha}_j$
- $- 18,81 \cdot \text{GraudeInstrução}_j + 33,06 \cdot \text{MotivoTrabalho}_j + 21,78 \cdot \text{MotivoEstudo}_j$
- $- 19,46 \cdot \text{MotivoLazer}_j + 5,85 \cdot \text{MotivoResidência}_j + 21,47 \cdot \text{MotivoSaúde}_j + 35,10$
- $\text{MotivoIntegração}_j - 42,74 \cdot \text{PopulaçãoOrigem}_j + 27,74 \cdot \text{Distância}_{16j} + 1,96$
- $\text{PontosdeOnibus}_{16j}$

---


$$V_{17j} = -13,23 + 32,18 \cdot \text{Moradores}_j + 43,87 \cdot \text{VeiculosMotorizados}_j + 9,42 \cdot \text{Renda}_j - 3,21$$

- $\text{Genero}_j - 24,44 \cdot \text{Idade}_j - 0,31 \cdot \text{FrequenciaEscolar}_j + 18,88 \cdot \text{Trabalha}_j + 48,34$
- $\text{GraudeInstrução}_j + 0,31 \cdot \text{MotivoTrabalho}_j + 35,77 \cdot \text{MotivoEstudo}_j - 38,37$
- $\text{MotivoLazer}_j - 34,16 \cdot \text{MotivoResidência}_j + 30,31 \cdot \text{MotivoSaúde}_j - 40,35$
- $\text{MotivoIntegração}_j - 45,19 \cdot \text{PopulaçãoOrigem}_j - 9,22 \cdot \text{Distância}_{17j} + 1,96$
- $\text{PontosdeOnibus}_{17j}$

---

$$V_{18j} = 42,23 - 2,50 \cdot \text{Moradores}_j + 34,49 \cdot \text{VeiculosMotorizados}_j + 26,76 \cdot \text{Renda}_j + 17,33 \\ \cdot \text{Genero}_j + 27,75 \cdot \text{Idade}_j + 23,89 \cdot \text{FrequenciaEscolar}_j + 40,71 \cdot \text{Trabalha}_j \\ + 36,43 \cdot \text{GraudeInstrução}_j + 4,27 \cdot \text{MotivoTrabalho}_j - 36,66 \cdot \text{MotivoEstudo}_j \\ - 23,43 \cdot \text{MotivoLazer}_j + 43,70 \cdot \text{MotivoResidência}_j + 25,08 \cdot \text{MotivoSaúde}_j - 4,09 \\ \cdot \text{MotivoIntegração}_j + 49,04 \cdot \text{PopulaçãoOrigem}_j - 3,73 \cdot \text{Distância}_{18j} + 1,96 \\ \cdot \text{PontosdeOnibus}_{18j}$$

$$V_{19j} = -8,26 - 30,01 \cdot \text{Moradores}_j + 3,31 \cdot \text{VeiculosMotorizados}_j - 43,95 \cdot \text{Renda}_j - 28,64 \\ \cdot \text{Genero}_j - 0,74 \cdot \text{Idade}_j - 5,59 \cdot \text{FrequenciaEscolar}_j - 12,84 \cdot \text{Trabalha}_j + 17,79 \\ \cdot \text{GraudeInstrução}_j - 6,78 \cdot \text{MotivoTrabalho}_j - 13,04 \cdot \text{MotivoEstudo}_j + 6,96 \\ \cdot \text{MotivoLazer}_j + 44,78 \cdot \text{MotivoResidência}_j - 32,08 \cdot \text{MotivoSaúde}_j - 3,12 \\ \cdot \text{MotivoIntegração}_j + 11,83 \cdot \text{PopulaçãoOrigem}_j + 33,25 \cdot \text{Distância}_{19j} + 1,96 \\ \cdot \text{PontosdeOnibus}_{19j}$$

$$V_{20j} = 6,74 - 5,48 \cdot \text{Moradores}_j + 19,89 \cdot \text{VeiculosMotorizados}_j + 38,03 \cdot \text{Renda}_j + 20,64 \cdot \text{Genero}_j \\ + 34,02 \cdot \text{Idade}_j + 50,00 \cdot \text{FrequenciaEscolar}_j + 26,38 \cdot \text{Trabalha}_j + 30,28 \\ \cdot \text{GraudeInstrução}_j - 25,58 \cdot \text{MotivoTrabalho}_j + 11,40 \cdot \text{MotivoEstudo}_j + 41,41 \\ \cdot \text{MotivoLazer}_j + 34,65 \cdot \text{MotivoResidência}_j - 14,97 \cdot \text{MotivoSaúde}_j + 28,67 \\ \cdot \text{MotivoIntegração}_j - 1,19 \cdot \text{PopulaçãoOrigem}_j + 22,62 \cdot \text{Distância}_{20j} + 1,96 \\ \cdot \text{PontosdeOnibus}_{20j}$$

$$V_{21j} = 34,37 + 49,52 \cdot \text{Moradores}_j + 41,51 \cdot \text{VeiculosMotorizados}_j + 27,26 \cdot \text{Renda}_j + 16,81 \\ \cdot \text{Genero}_j + 38,85 \cdot \text{Idade}_j + 29,78 \cdot \text{FrequenciaEscolar}_j + 41,16 \cdot \text{Trabalha}_j \\ + 48,93 \cdot \text{GraudeInstrução}_j + 39,20 \cdot \text{MotivoTrabalho}_j - 31,50 \cdot \text{MotivoEstudo}_j \\ + 18,49 \cdot \text{MotivoLazer}_j + 47,35 \cdot \text{MotivoResidência}_j - 37,59 \cdot \text{MotivoSaúde}_j \\ - 36,29 \cdot \text{MotivoIntegração}_j + 48,76 \cdot \text{PopulaçãoOrigem}_j + 16,45 \cdot \text{Distância}_{21j} \\ + 1,96 \cdot \text{PontosdeOnibus}_{21j}$$

$$V_{22j} = 50,00 - 44,74 \cdot \text{Moradores}_j - 22,25 \cdot \text{VeiculosMotorizados}_j + 15,45 \cdot \text{Renda}_j + 34,48 \\ \cdot \text{Genero}_j - 45,26 \cdot \text{Idade}_j - 3,29 \cdot \text{FrequenciaEscolar}_j + 46,76 \cdot \text{Trabalha}_j + 44,18 \\ \cdot \text{GraudeInstrução}_j + 22,20 \cdot \text{MotivoTrabalho}_j - 45,28 \cdot \text{MotivoEstudo}_j + 29,12 \\ \cdot \text{MotivoLazer}_j + 47,25 \cdot \text{MotivoResidência}_j - 40,70 \cdot \text{MotivoSaúde}_j - 41,61 \\ \cdot \text{MotivoIntegração}_j - 2,76 \cdot \text{PopulaçãoOrigem}_j - 28,68 \cdot \text{Distância}_{22j} + 1,96 \\ \cdot \text{PontosdeOnibus}_{22j}$$

$$V_{23j} = 21,30 + 20,07 \cdot \text{Moradores}_j + 28,45 \cdot \text{VeiculosMotorizados}_j - 13,02 \cdot \text{Renda}_j + 48,49 \\ \cdot \text{Genero}_j + 17,63 \cdot \text{Idade}_j + 41,79 \cdot \text{FrequenciaEscolar}_j + 24,25 \cdot \text{Trabalha}_j \\ - 13,63 \cdot \text{GraudeInstrução}_j - 28,54 \cdot \text{MotivoTrabalho}_j + 13,60 \cdot \text{MotivoEstudo}_j \\ - 10,84 \cdot \text{MotivoLazer}_j + 20,81 \cdot \text{MotivoResidência}_j - 48,11 \cdot \text{MotivoSaúde}_j \\ - 37,22 \cdot \text{MotivoIntegração}_j + 43,32 \cdot \text{PopulaçãoOrigem}_j + 25,04 \cdot \text{Distância}_{23j} \\ + 1,96 \cdot \text{PontosdeOnibus}_{23j}$$

$$V_{24j} = 7,63 - 22,82 \cdot \text{Moradores}_j - 39,07 \cdot \text{VeiculosMotorizados}_j - 44,86 \cdot \text{Renda}_j + 7,33 \cdot \text{Genero}_j \\ - 45,90 \cdot \text{Idade}_j - 44,77 \cdot \text{FrequenciaEscolar}_j - 6,32 \cdot \text{Trabalha}_j + 16,47 \\ \cdot \text{GraudeInstrução}_j + 27,22 \cdot \text{MotivoTrabalho}_j + 45,16 \cdot \text{MotivoEstudo}_j + 23,86 \\ \cdot \text{MotivoLazer}_j + 30,55 \cdot \text{MotivoResidência}_j + 34,42 \cdot \text{MotivoSaúde}_j - 10,39 \\ \cdot \text{MotivoIntegração}_j + 17,48 \cdot \text{PopulaçãoOrigem}_j + 27,93 \cdot \text{Distância}_{24j} + 1,96 \\ \cdot \text{PontosdeOnibus}_{24j}$$

$$V_{25j} = -42,26 + 16,27 \cdot \text{Moradores}_j - 6,97 \cdot \text{VeiculosMotorizados}_j - 18,06 \cdot \text{Renda}_j + 41,26 \\ \cdot \text{Genero}_j + 45,75 \cdot \text{Idade}_j - 25,27 \cdot \text{FrequenciaEscolar}_j - 42,83 \cdot \text{Trabalha}_j - 4,16 \\ \cdot \text{GraudeInstrução}_j + 3,31 \cdot \text{MotivoTrabalho}_j + 48,36 \cdot \text{MotivoEstudo}_j + 40,88 \\ \cdot \text{MotivoLazer}_j + 34,73 \cdot \text{MotivoResidência}_j - 2,69 \cdot \text{MotivoSaúde}_j + 27,37 \\ \cdot \text{MotivoIntegração}_j + 9,75 \cdot \text{PopulaçãoOrigem}_j - 12,84 \cdot \text{Distância}_{25j} + 1,96 \\ \cdot \text{PontosdeOnibus}_{25j}$$

---


$$V_{26j} = -0,85 + 3,74 \cdot \text{Moradores}_j + 19,90 \cdot \text{VeiculosMotorizados}_j + 13,16 \cdot \text{Renda}_j - 21,27$$

- $\text{Genero}_j - 13,79 \cdot \text{Idade}_j - 11,61 \cdot \text{FrequenciaEscolar}_j - 7,25 \cdot \text{Trabalha}_j - 20,58$
- $\text{GraudeInstrução}_j - 24,61 \cdot \text{MotivoTrabalho}_j - 22,52 \cdot \text{MotivoEstudo}_j + 1,73$
- $\text{MotivoLazer}_j + 36,80 \cdot \text{MotivoResidência}_j - 9,15 \cdot \text{MotivoSaúde}_j - 2,06$
- $\text{MotivoIntegração}_j - 49,27 \cdot \text{PopulaçãoOrigem}_j - 40,41 \cdot \text{Distância}_{26j} + 1,96$
- $\text{PontosdeOnibus}_{26j}$

---


$$V_{27j} = -48,15 - 22,68 \cdot \text{Moradores}_j - 34,46 \cdot \text{VeiculosMotorizados}_j + 42,52 \cdot \text{Renda}_j - 18,40$$

- $\text{Genero}_j - 48,48 \cdot \text{Idade}_j + 8,18 \cdot \text{FrequenciaEscolar}_j - 36,11 \cdot \text{Trabalha}_j + 47,33$
- $\text{GraudeInstrução}_j - 33,79 \cdot \text{MotivoTrabalho}_j + 13,55 \cdot \text{MotivoEstudo}_j - 16,42$
- $\text{MotivoLazer}_j - 12,16 \cdot \text{MotivoResidência}_j + 42,72 \cdot \text{MotivoSaúde}_j + 30,01$
- $\text{MotivoIntegração}_j - 3,81 \cdot \text{PopulaçãoOrigem}_j + 36,76 \cdot \text{Distância}_{27j} + 1,96$
- $\text{PontosdeOnibus}_{27j}$

---


$$V_{28j} = -15,10 + 37,66 \cdot \text{Moradores}_j - 25,06 \cdot \text{VeiculosMotorizados}_j + 40,49 \cdot \text{Renda}_j - 4,91$$

- $\text{Genero}_j - 4,15 \cdot \text{Idade}_j + 15,19 \cdot \text{FrequenciaEscolar}_j - 24,37 \cdot \text{Trabalha}_j + 29,59$
- $\text{GraudeInstrução}_j - 25,16 \cdot \text{MotivoTrabalho}_j + 17,43 \cdot \text{MotivoEstudo}_j + 29,61$
- $\text{MotivoLazer}_j - 49,29 \cdot \text{MotivoResidência}_j - 40,26 \cdot \text{MotivoSaúde}_j - 48,43$
- $\text{MotivoIntegração}_j - 5,08 \cdot \text{PopulaçãoOrigem}_j - 44,02 \cdot \text{Distância}_{28j} + 1,96$
- $\text{PontosdeOnibus}_{28j}$

---


$$V_{29j} = -38,31 + 22,17 \cdot \text{Moradores}_j + 27,25 \cdot \text{VeiculosMotorizados}_j - 8,94 \cdot \text{Renda}_j - 8,04$$

- $\text{Genero}_j - 32,73 \cdot \text{Idade}_j + 44,84 \cdot \text{FrequenciaEscolar}_j - 38,76 \cdot \text{Trabalha}_j$
- $- 20,98 \cdot \text{GraudeInstrução}_j - 37,41 \cdot \text{MotivoTrabalho}_j - 7,66 \cdot \text{MotivoEstudo}_j$
- $- 42,80 \cdot \text{MotivoLazer}_j - 2,01 \cdot \text{MotivoResidência}_j + 31,27 \cdot \text{MotivoSaúde}_j - 16,61$
- $\text{MotivoIntegração}_j + 47,92 \cdot \text{PopulaçãoOrigem}_j + 8,96 \cdot \text{Distância}_{29j} + 1,96$
- $\text{PontosdeOnibus}_{29j}$

---


$$V_{30j} = -20,14 + 2,66 \cdot \text{Moradores}_j + 10,53 \cdot \text{VeiculosMotorizados}_j - 44,73 \cdot \text{Renda}_j + 30,50$$

- $\text{Genero}_j - 38,33 \cdot \text{Idade}_j - 41,25 \cdot \text{FrequenciaEscolar}_j + 46,69 \cdot \text{Trabalha}_j$
- $+ 25,99 \cdot \text{GraudeInstrução}_j + 18,44 \cdot \text{MotivoTrabalho}_j + 45,85 \cdot \text{MotivoEstudo}_j$
- $+ 41,29 \cdot \text{MotivoLazer}_j - 35,08 \cdot \text{MotivoResidência}_j + 13,80 \cdot \text{MotivoSaúde}_j$
- $- 46,10 \cdot \text{MotivoIntegração}_j + 8,88 \cdot \text{PopulaçãoOrigem}_j - 31,00 \cdot \text{Distância}_{30j}$
- $+ 1,96 \cdot \text{PontosdeOnibus}_{30j}$

---


$$V_{31j} = 25,50 + 15,31 \cdot \text{Moradores}_j + 31,92 \cdot \text{VeiculosMotorizados}_j + 40,51 \cdot \text{Renda}_j - 8,12$$

- $\text{Genero}_j - 11,43 \cdot \text{Idade}_j - 31,97 \cdot \text{FrequenciaEscolar}_j - 26,92 \cdot \text{Trabalha}_j + 2,72$
- $\text{GraudeInstrução}_j + 44,80 \cdot \text{MotivoTrabalho}_j - 18,61 \cdot \text{MotivoEstudo}_j + 11,15$
- $\text{MotivoLazer}_j + 41,27 \cdot \text{MotivoResidência}_j - 8,24 \cdot \text{MotivoSaúde}_j + 38,43$
- $\text{MotivoIntegração}_j + 17,43 \cdot \text{PopulaçãoOrigem}_j + 21,29 \cdot \text{Distância}_{31j} + 1,96$
- $\text{PontosdeOnibus}_{31j}$

---


$$V_{32j} = 30,32 - 12,65 \cdot \text{Moradores}_j - 2,79 \cdot \text{VeiculosMotorizados}_j - 2,30 \cdot \text{Renda}_j + 24,20 \cdot \text{Genero}_j$$

- $+ 40,70 \cdot \text{Idade}_j + 19,15 \cdot \text{FrequenciaEscolar}_j + 0,42 \cdot \text{Trabalha}_j + 45,48$
- $\text{GraudeInstrução}_j - 23,86 \cdot \text{MotivoTrabalho}_j + 4,90 \cdot \text{MotivoEstudo}_j + 37,59$
- $\text{MotivoLazer}_j - 2,61 \cdot \text{MotivoResidência}_j + 39,37 \cdot \text{MotivoSaúde}_j + 38,99$
- $\text{MotivoIntegração}_j + 30,85 \cdot \text{PopulaçãoOrigem}_j + 31,53 \cdot \text{Distância}_{32j} + 1,96$
- $\text{PontosdeOnibus}_{32j}$

---


$$V_{33j} = -37,52 + 17,68 \cdot \text{Moradores}_j + 2,64 \cdot \text{VeiculosMotorizados}_j - 30,69 \cdot \text{Renda}_j - 11,57$$

- $\text{Genero}_j + 20,69 \cdot \text{Idade}_j - 15,02 \cdot \text{FrequenciaEscolar}_j + 5,56 \cdot \text{Trabalha}_j - 0,84$
- $\text{GraudeInstrução}_j - 27,95 \cdot \text{MotivoTrabalho}_j - 9,64 \cdot \text{MotivoEstudo}_j - 38,52$
- $\text{MotivoLazer}_j - 19,84 \cdot \text{MotivoResidência}_j - 5,00 \cdot \text{MotivoSaúde}_j - 17,21$
- $\text{MotivoIntegração}_j + 13,06 \cdot \text{PopulaçãoOrigem}_j + 41,07 \cdot \text{Distância}_{33j} + 1,96$
- $\text{PontosdeOnibus}_{33j}$

---

---

$$V_{34j} = 38,89 + 37,91 \cdot \text{Moradores}_j + 18,09 \cdot \text{VeiculosMotorizados}_j - 34,29 \cdot \text{Renda}_j + 10,07$$
$$\cdot \text{Genero}_j - 47,95 \cdot \text{Idade}_j + 28,98 \cdot \text{FrequenciaEscolar}_j + 27,60 \cdot \text{Trabalha}_j + 4,69$$
$$\cdot \text{GraudeInstrução}_j + 19,62 \cdot \text{MotivoTrabalho}_j - 6,59 \cdot \text{MotivoEstudo}_j - 11,40$$
$$\cdot \text{MotivoLazer}_j + 49,16 \cdot \text{MotivoResidência}_j + 38,15 \cdot \text{MotivoSaúde}_j - 47,45$$
$$\cdot \text{MotivoIntegração}_j + 43,15 \cdot \text{PopulaçãoOrigem}_j + 12,87 \cdot \text{Distância}_{34j} + 1,96$$
$$\cdot \text{PontosdeOnibus}_{34j}$$

---

$$V_{35j} = -41,16 + 31,48 \cdot \text{Moradores}_j - 344,93 \cdot \text{VeiculosMotorizados}_j - 41,59 \cdot \text{Renda}_j + 19,98$$
$$\cdot \text{Genero}_j + 10,09 \cdot \text{Idade}_j - 45,05 \cdot \text{FrequenciaEscolar}_j + 23,47 \cdot \text{Trabalha}_j$$
$$+ 22,88 \cdot \text{GraudeInstrução}_j + 46,55 \cdot \text{MotivoTrabalho}_j + 15,60 \cdot \text{MotivoEstudo}_j$$
$$- 29,57 \cdot \text{MotivoLazer}_j + 42,45 \cdot \text{MotivoResidência}_j + 23,50 \cdot \text{MotivoSaúde}_j$$
$$+ 48,70 \cdot \text{MotivoIntegração}_j + 27,08 \cdot \text{PopulaçãoOrigem}_j + 30,83 \cdot \text{Distância}_{35j}$$
$$+ 1,96 \cdot \text{PontosdeOnibus}_{35j}$$

---



**APÊNDICE F - Calibração por Algoritmos Genéticos que gerou o menor valor para a função objetivo no cenário semi-restrito**

---


$$V_{1j} = -14,23 \cdot Distância_{1j} + 3,13 \cdot PontosdeOnibus_{1j}$$


---


$$V_{2j} = -17,54 + 35,65 \cdot Moradores_j + 2,79 \cdot VeiculosMotorizados_j - 13,59 \cdot Renda_j + 16,32$$

$$\cdot Genero_j - 1,15 \cdot Idade_j - 7,56 \cdot FrequenciaEscolar_j + 19,85 \cdot Trabalha_j - 46,75$$

$$\cdot GraudeInstrução_j + 35,59 \cdot MotivoTrabalho_j + 47,46 \cdot MotivoEstudo_j + 48,16$$

$$\cdot MotivoLazer_j + 35,73 \cdot MotivoResidência_j - 4,05 \cdot MotivoSaúde_j + 10,40$$

$$\cdot MotivoIntegração_j + 27,51 \cdot PopulaçãoOrigem_j + 8,87 \cdot Distância_{2j} + 3,13$$

$$\cdot PontosdeOnibus_{2j}$$


---


$$V_{3j} = 41,65 + 43,20 \cdot Moradores_j + 47,28 \cdot VeiculosMotorizados_j + 25,81 \cdot Renda_j - 16,02$$

$$\cdot Genero_j + 38,34 \cdot Trabalha_j + 31,80 \cdot GraudeInstrução_j + 34,64$$

$$\cdot MotivoTrabalho_j - 1,90 \cdot MotivoEstudo_j + 35,28 \cdot MotivoLazer_j - 1,20$$

$$\cdot MotivoResidência_j + 13,84 \cdot MotivoIntegração_j + 21,46 \cdot PopulaçãoOrigem_j$$

$$- 37,92 \cdot Distância_{3j} + 3,13 \cdot PontosdeOnibus_{3j}$$


---


$$V_{4j} = 35,29 - 12,52 \cdot VeiculosMotorizados_j + 38,07 \cdot Idade_j + 32,85 \cdot FrequenciaEscolar_j + 49,48$$

$$\cdot GraudeInstrução_j - 5,67 \cdot MotivoEstudo_j - 33,55 \cdot MotivoLazer_j + 19,21$$

$$\cdot MotivoResidência_j + 13,13 \cdot MotivoSaúde_j + 44,55 \cdot MotivoIntegração_j + 10,60$$

$$\cdot PopulaçãoOrigem_j + 26,98 \cdot Distância_{4j} + 3,13 \cdot PontosdeOnibus_{4j}$$


---


$$V_{5j} = 3,57 - 4,54 \cdot Moradores_j + 44,83 \cdot VeiculosMotorizados_j + 43,18 \cdot Renda_j + 11,12 \cdot Idade_j$$

$$+ 25,96 \cdot Trabalha_j - 2,61 \cdot GraudeInstrução_j + 11,43 \cdot MotivoTrabalho_j + 15,41$$

$$\cdot MotivoEstudo_j + 2,97 \cdot MotivoLazer_j + 26,60 \cdot MotivoResidência_j + 32,08$$

$$\cdot MotivoSaúde_j + 10,09 \cdot MotivoIntegração_j + 21,33 \cdot PopulaçãoOrigem_j - 4,80$$

$$\cdot Distância_{5j} + 3,13 \cdot PontosdeOnibus_{5j}$$


---


$$V_{6j} = -0,33 + 24,80 \cdot Moradores_j - 37,79 \cdot VeiculosMotorizados_j + 8,25 \cdot Renda_j - 45,53$$

$$\cdot Genero_j - 21,35 \cdot Idade_j + 28,79 \cdot Trabalha_j + 47,88 \cdot GraudeInstrução_j + 2,59$$

$$\cdot MotivoTrabalho_j + 27,79 \cdot MotivoEstudo_j - 48,12 \cdot MotivoLazer_j + 15,31$$

$$\cdot MotivoResidência_j - 42,66 \cdot MotivoSaúde_j + 20,13 \cdot MotivoIntegração_j + 48,53$$

$$\cdot PopulaçãoOrigem_j + 13,82 \cdot Distância_{6j} + 3,13 \cdot PontosdeOnibus_{6j}$$


---


$$V_{7j} = -17,16 + 45,74 \cdot Moradores_j - 7,11 \cdot VeiculosMotorizados_j + 33,95 \cdot Renda_j + 13,68$$

$$\cdot Genero_j - 11,57 \cdot Idade_j + 6,24 \cdot FrequenciaEscolar_j + 4,15 \cdot GraudeInstrução_j$$

$$- 27,65 \cdot MotivoTrabalho_j + 41,35 \cdot MotivoEstudo_j - 40,85 \cdot MotivoLazer_j - 18,05$$

$$\cdot MotivoSaúde_j + 24,97 \cdot MotivoIntegração_j + 30,63 \cdot PopulaçãoOrigem_j + 46,46$$

$$\cdot Distância_{7j} + 3,13 \cdot PontosdeOnibus_{7j}$$


---


$$V_{8j} = 49,20 + 27,33 \cdot Moradores_j + 25,30 \cdot Renda_j + 38,69 \cdot Idade_j + 35,84 \cdot GraudeInstrução_j$$

$$+ 29,57 \cdot MotivoTrabalho_j + 21,20 \cdot MotivoLazer_j - 31,79 \cdot MotivoResidência_j$$

$$+ 35,72 \cdot MotivoSaúde_j - 24,25 \cdot MotivoIntegração_j + 25,03 \cdot PopulaçãoOrigem_j$$

$$+ 40,21 \cdot Distância_{8j} + 3,13 \cdot PontosdeOnibus_{8j}$$


---


$$V_{9j} = 1,16 - 38,32 \cdot Moradores_j - 44,23 \cdot VeiculosMotorizados_j - 36,45 \cdot Renda_j - 25,09 \cdot Genero_j$$

$$+ 7,56 \cdot Idade_j + 33,19 \cdot FrequenciaEscolar_j - 31,46 \cdot Trabalha_j - 37,11$$

$$\cdot GraudeInstrução_j + 31,14 \cdot MotivoTrabalho_j + 49,39 \cdot MotivoEstudo_j - 35,12$$

$$\cdot MotivoLazer_j - 1,94 \cdot MotivoResidência_j + 21,91 \cdot MotivoSaúde_j + 1,84$$

$$\cdot MotivoIntegração_j + 26,63 \cdot PopulaçãoOrigem_j - 29,43 \cdot Distância_{9j} + 3,13$$

$$\cdot PontosdeOnibus_{9j}$$


---


$$V_{10j} = 21,68 + 2,66 \cdot Renda_j + 25,75 \cdot Genero_j + 4,52 \cdot Idade_j + 47,34 \cdot FrequenciaEscolar_j + 28,54$$

$$\cdot Trabalha_j + 16,10 \cdot GraudeInstrução_j + 17,23 \cdot MotivoTrabalho_j + 20,17$$

$$\cdot MotivoEstudo_j + 35,74 \cdot MotivoLazer_j + 15,55 \cdot MotivoResidência_j - 27,17$$

$$\cdot MotivoSaúde_j + 35,84 \cdot MotivoIntegração_j + 16,65 \cdot PopulaçãoOrigem_j - 23,35$$

$$\cdot Distância_{10j} + 3,13 \cdot PontosdeOnibus_{10j}$$


---

---


$$V_{11j} = -43,77 - 21,22 \cdot \text{Moradores}_j - 36,92 \cdot \text{VeiculosMotorizados}_j + 22,36 \cdot \text{Renda}_j + 30,44$$

$$\cdot \text{Genero}_j + 38,97 \cdot \text{FrequenciaEscolar}_j + 32,23 \cdot \text{Trabalha}_j + 30,99$$

$$\cdot \text{GraudeInstrução}_j - 38,86 \cdot \text{MotivoTrabalho}_j + 22,36 \cdot \text{MotivoEstudo}_j + 25,08$$

$$\cdot \text{MotivoResidência}_j - 7,93 \cdot \text{MotivoIntegração}_j + 38,03 \cdot \text{PopulaçãoOrigem}_j$$

$$+ 18,96 \cdot \text{Distância}_{11j} + 3,13 \cdot \text{PontosdeOnibus}_{11j}$$


---


$$V_{12j} = 25,75 - 29,31 \cdot \text{Moradores}_j - 30,77 \cdot \text{VeiculosMotorizados}_j - 2,77 \cdot \text{Renda}_j + 20,22$$

$$\cdot \text{Genero}_j + 12,43 \cdot \text{FrequenciaEscolar}_j - 49,90 \cdot \text{MotivoTrabalho}_j + 31,26$$

$$\cdot \text{MotivoLazer}_j - 7,95 \cdot \text{MotivoResidência}_j + 26,50 \cdot \text{MotivoSaúde}_j - 1,98$$

$$\cdot \text{MotivoIntegração}_j + 31,24 \cdot \text{PopulaçãoOrigem}_j + 33,41 + 3,13$$

$$\cdot \text{PontosdeOnibus}_{12j}$$


---


$$V_{13j} = 10,02 - 34,76 \cdot \text{Moradores}_j + 5,65 \cdot \text{VeiculosMotorizados}_j + 15,08 \cdot \text{Renda}_j + 43,16 \cdot \text{Idade}_j$$

$$+ 21,67 \cdot \text{FrequenciaEscolar}_j + 10,99 \cdot \text{MotivoTrabalho}_j - 13,34 \cdot \text{MotivoEstudo}_j$$

$$+ 42,16 \cdot \text{MotivoLazer}_j + 29,69 \cdot \text{MotivoResidência}_j - 0,10 \cdot \text{MotivoSaúde}_j + 46,88$$

$$\cdot \text{MotivoIntegração}_j + 44,43 \cdot \text{PopulaçãoOrigem}_j + 36,06 \cdot \text{Distância}_{13j} + 3,13$$

$$\cdot \text{PontosdeOnibus}_{13j}$$


---


$$V_{14j} = 22,52 + 27,15 \cdot \text{Moradores}_j + 34,37 \cdot \text{VeiculosMotorizados}_j + 5,17 \cdot \text{Renda}_j - 20,99 \cdot \text{Idade}_j$$

$$+ 37,17 \cdot \text{FrequenciaEscolar}_j \pm 15,56 \cdot \text{Trabalha}_j - 45,37 \cdot \text{GraudeInstrução}_j$$

$$+ 5,36 \cdot \text{MotivoEstudo}_j - 3,29 \cdot \text{MotivoLazer}_j - 17,61 \cdot \text{MotivoResidência}_j - 42,07$$

$$\cdot \text{MotivoSaúde}_j - 41,21 \cdot \text{MotivoIntegração}_j - 44,29 \cdot \text{PopulaçãoOrigem}_j + 0,37$$

$$\cdot \text{Distância}_{14j} + 3,13 \cdot \text{PontosdeOnibus}_{14j}$$


---


$$V_{15j} = 43,30 + 5,11 \cdot \text{VeiculosMotorizados}_j + 28,89 \cdot \text{Renda}_j + 12,36 \cdot \text{Idade}_j + 36,81$$

$$\cdot \text{GraudeInstrução}_j + 47,39 \cdot \text{MotivoTrabalho}_j - 1,55 \cdot \text{MotivoLazer}_j + 24,35$$

$$\cdot \text{MotivoResidência}_j - 40,10 \cdot \text{MotivoSaúde}_j + 33,51 \cdot \text{MotivoIntegração}_j + 38,71$$

$$\cdot \text{PopulaçãoOrigem}_j - 9,73 \cdot \text{Distância}_{15j} + 3,13 \cdot \text{PontosdeOnibus}_{15j}$$


---


$$V_{16j} = 36,17 + 18,53 \cdot \text{Moradores}_j - 12,20 \cdot \text{Genero}_j - 20,60 \cdot \text{FrequenciaEscolar}_j + 23,15$$

$$\cdot \text{Trabalha}_j + 31,30 \cdot \text{GraudeInstrução}_j - 15,78 \cdot \text{MotivoTrabalho}_j - 6,64$$

$$\cdot \text{MotivoEstudo}_j - 10,01 \cdot \text{MotivoResidência}_j + 7,54 \cdot \text{MotivoIntegração}_j - 11,89$$

$$\cdot \text{PopulaçãoOrigem}_j + 29,11 \cdot \text{Distância}_{16j} + 3,13 \cdot \text{PontosdeOnibus}_{16j}$$


---


$$V_{17j} = -4,79 - 29,30 \cdot \text{Moradores}_j + 28,26 \cdot \text{VeiculosMotorizados}_j + 8,70 \cdot \text{Renda}_j - 27,97 \cdot \text{Idade}_j$$

$$- 25,66 \cdot \text{GraudeInstrução}_j + 35,78 \cdot \text{MotivoTrabalho}_j - 49,46 \cdot \text{MotivoEstudo}_j$$

$$+ 41,85 \cdot \text{MotivoLazer}_j + 44,56 \cdot \text{MotivoResidência}_j - 4,24 \cdot \text{MotivoSaúde}_j + 45,65$$

$$\cdot \text{MotivoIntegração}_j - 39,78 \cdot \text{PopulaçãoOrigem}_j + 21,82 \cdot \text{Distância}_{17j} + 3,13$$

$$\cdot \text{PontosdeOnibus}_{17j}$$


---


$$V_{18j} = 43,71 + 44,86 \cdot \text{Moradores}_j - 6,36 \cdot \text{VeiculosMotorizados}_j + 27,47 \cdot \text{Renda}_j + 8,85$$

$$\cdot \text{Trabalha}_j + 48,42 \cdot \text{GraudeInstrução}_j - 7,86 \cdot \text{MotivoTrabalho}_j - 42,21$$

$$\cdot \text{MotivoLazer}_j + 2,09 \cdot \text{MotivoResidência}_j + 31,50 \cdot \text{MotivoSaúde}_j + 37,42$$

$$\cdot \text{MotivoIntegração}_j + 19,71 \cdot \text{PopulaçãoOrigem}_j - 35,72 \text{Distância}_{18j} + 3,13$$

$$\cdot \text{PontosdeOnibus}_{18j}$$


---


$$V_{19j} = -32,40 - 14,40 \cdot \text{Moradores}_j + 5,51 \cdot \text{Renda}_j - 44,48 \cdot \text{Genero}_j - 9,92 \cdot \text{Trabalha}_j + 32,13$$

$$\cdot \text{MotivoTrabalho}_j + 28,54 \cdot \text{MotivoEstudo}_j + 15,96 \cdot \text{MotivoLazer}_j + 0,35$$

$$\cdot \text{MotivoResidência}_j - 45,20 \cdot \text{MotivoSaúde}_j - 40,16 \cdot \text{MotivoIntegração}_j + 33,97$$

$$\cdot \text{PopulaçãoOrigem}_j - 36,28 \cdot \text{Distância}_{19j} + 3,13 \cdot \text{PontosdeOnibus}_{19j}$$


---


$$V_{20j} = 50,00 + 24,08 \cdot \text{Moradores}_j + 30,21 \cdot \text{VeiculosMotorizados}_j + 43,67 \cdot \text{Idade}_j + 3,19$$

$$\cdot \text{FrequenciaEscolar}_j + 49,13 \cdot \text{Trabalha}_j + 35,65 \cdot \text{GraudeInstrução}_j + 42,22$$

$$\cdot \text{MotivoEstudo}_j - 30,94 \cdot \text{MotivoLazer}_j + 41,80 \cdot \text{MotivoResidência}_j - 25,28$$

$$\cdot \text{MotivoSaúde}_j + 14,12 \cdot \text{MotivoIntegração}_j + 27,69 \cdot \text{PopulaçãoOrigem}_j - 1,49$$

$$\cdot \text{Distância}_{20j} + 3,13 \cdot \text{PontosdeOnibus}_{20j}$$


---


$$V_{21j} = 37,493 + 39,64 \cdot \text{Renda}_j + 5,37 \cdot \text{Genero}_j + 40,18 \cdot \text{Idade}_j + 49,64 \cdot \text{GraudeInstrução}_j$$

$$+ 29,34 \cdot \text{MotivoTrabalho}_j - 18,36 \cdot \text{MotivoEstudo}_j + 11,41 \cdot \text{MotivoLazer}_j - 14,81$$

$$\cdot \text{MotivoResidência}_j - 49,71 \cdot \text{MotivoIntegração}_j + 40,15 \cdot \text{PopulaçãoOrigem}_j$$

$$- 5,80 \cdot \text{Distância}_{21j} + 3,13 \cdot \text{PontosdeOnibus}_{21j}$$


---



$$V_{22j} = 41,73 - 19,28 \cdot \text{Moradores}_j + 19,90 \cdot \text{VeiculosMotorizados}_j + 8,09 \cdot \text{Renda}_j + 44,99 \\ \cdot \text{GraudeInstrução}_j - 2,03 \cdot \text{MotivoTrabalho}_j + 13,70 \cdot \text{MotivoSaúde}_j - 16,73 \\ \cdot \text{MotivoIntegração}_j + 22,36 \cdot \text{PopulaçãoOrigem}_j - 9,24 \cdot \text{Distância}_{22j} + 3,13 \\ \cdot \text{PontosdeOnibus}_{22j}$$

$$V_{23j} = 49,64 + 45,29 \cdot \text{Moradores}_j + 26,41 \cdot \text{VeiculosMotorizados}_j + 33,82 \cdot \text{Renda}_j - 0,41 \\ \cdot \text{Genero}_j - 3,16 \cdot \text{Idade}_j - 15,49 \cdot \text{FrequenciaEscolar}_j - 19,60 \cdot \text{Trabalha}_j + 35,03 \\ \cdot \text{MotivoEstudo}_j + 28,41 \cdot \text{MotivoLazer}_j + 35,44 \cdot \text{MotivoResidência}_j + 46,21 \\ \cdot \text{MotivoSaúde}_j - 29,47 \cdot \text{MotivoIntegração}_j - 47,68 \cdot \text{PopulaçãoOrigem}_j + 42,91 \\ \cdot \text{Distância}_{23j} + 3,13 \cdot \text{PontosdeOnibus}_{23j}$$

$$V_{24j} = -3,92 - 26,67 \cdot \text{Moradores}_j - 44,17 \cdot \text{VeiculosMotorizados}_j - 13,49 \cdot \text{Renda}_j + 34,98 \\ \cdot \text{Idade}_j - 39,07 \cdot \text{FrequenciaEscolar}_j - 38,44 \cdot \text{GraudeInstrução}_j - 8,17 \\ \cdot \text{MotivoLazer}_j + 47,40 \cdot \text{MotivoIntegração}_j - 35,06 \cdot \text{PopulaçãoOrigem}_j - 3,04 \\ \cdot \text{Distância}_{24j} + 3,13 \cdot \text{PontosdeOnibus}_{24j}$$

$$V_{25j} = 32,76 - 0,62 \cdot \text{Moradores}_j - 5,64 \cdot \text{VeiculosMotorizados}_j + 24,69 \cdot \text{Renda}_j + 47,04 \\ \cdot \text{GraudeInstrução}_j - 37,64 \cdot \text{MotivoTrabalho}_j - 40,55 \cdot \text{MotivoEstudo}_j + 43,58 \\ \cdot \text{MotivoLazer}_j - 28,13 \cdot \text{MotivoResidência}_j - 49,36 \cdot \text{MotivoSaúde}_j - 19,78 \\ \cdot \text{MotivoIntegração}_j + 6,12 \cdot \text{Distância}_{25j} + 3,13 \cdot \text{PontosdeOnibus}_{25j}$$

$$V_{26j} = 39,86 + 40,85 \cdot \text{VeiculosMotorizados}_j - 4,39 \cdot \text{Renda}_j + 16,64 \cdot \text{FrequenciaEscolar}_j + 40,41 \\ \cdot \text{Trabalha}_j - 16,70 \cdot \text{GraudeInstrução}_j + 39,83 \cdot \text{MotivoTrabalho}_j - 47,21 \\ \cdot \text{MotivoEstudo}_j - 13,86 \cdot \text{MotivoLazer}_j + 32,97 \cdot \text{MotivoResidência}_j - 1,31 \\ \cdot \text{MotivoSaúde}_j - 28,75 \cdot \text{MotivoIntegração}_j + 14,36 \cdot \text{PopulaçãoOrigem}_j + 38,18 \\ \cdot \text{Distância}_{26j} + 3,13 \cdot \text{PontosdeOnibus}_{26j}$$

$$V_{27j} = 49,47 + 46,25 \cdot \text{Moradores}_j + 48,95 \cdot \text{Renda}_j + 23,33 \cdot \text{Genero}_j - 1,54 \cdot \text{Idade}_j - 10,89 \\ \cdot \text{FrequenciaEscolar}_j - 3,02 \cdot \text{Trabalha}_j - 28,03 \cdot \text{MotivoEstudo}_j - 7,82 \\ \cdot \text{MotivoLazer}_j + 47,55 \cdot \text{MotivoResidência}_j - 14,27 \cdot \text{MotivoSaúde}_j + 23,73 \\ \cdot \text{MotivoIntegração}_j - 20,59 \cdot \text{PopulaçãoOrigem}_j - 18,27 \cdot \text{Distância}_{27j} + 3,53 \\ \cdot \text{PontosdeOnibus}_{27j}$$

$$V_{28j} = 3,53 - 1,47 \cdot \text{Moradores}_j + 31,37 \cdot \text{VeiculosMotorizados}_j + 47,72 \cdot \text{Renda}_j - 34,98 \cdot \text{Genero}_j \\ + 6,89 \cdot \text{FrequenciaEscolar}_j - 2,49 \cdot \text{Trabalha}_j + 35,04 \cdot \text{GraudeInstrução}_j + 14,24 \\ \cdot \text{MotivoTrabalho}_j + 40,96 \cdot \text{MotivoEstudo}_j + 5,66 \cdot \text{MotivoLazer}_j - 36,75 \\ \cdot \text{MotivoResidência}_j + 34,17 \cdot \text{MotivoSaúde}_j + 18,66 \cdot \text{PopulaçãoOrigem}_j + 1,16 \\ \cdot \text{Distância}_{28j} + 33,74 \cdot \text{PontosdeOnibus}_{28j}$$

$$V_{29j} = 33,74 + 17,50 \cdot \text{Moradores}_j - 3,40 \cdot \text{VeiculosMotorizados}_j + 14,76 \cdot \text{Renda}_j - 9,21 \cdot \text{Genero}_j \\ + 3,46 \cdot \text{Idade}_j - 36,76 \cdot \text{FrequenciaEscolar}_j - 21,86 \cdot \text{Trabalha}_j + 6,41 \\ \cdot \text{GraudeInstrução}_j + 33,95 \cdot \text{MotivoTrabalho}_j - 22,08 \cdot \text{MotivoEstudo}_j - 13,21 \\ \cdot \text{MotivoLazer}_j - 25,06 \cdot \text{MotivoResidência}_j + 15,32 \cdot \text{MotivoSaúde}_j - 19,23 \\ \cdot \text{MotivoIntegração}_j - 14,65 \cdot \text{PopulaçãoOrigem}_j + 25,27 \cdot \text{Distância}_{29j} + 3,13 \\ \cdot \text{PontosdeOnibus}_{29j}$$

$$V_{30j} = -20,33 + 2,04 \cdot \text{Moradores}_j - 1,48 \cdot \text{VeiculosMotorizados}_j + 5,35 \cdot \text{Renda}_j - 1,47 \cdot \text{Genero}_j \\ + 30,66 \cdot \text{Idade}_j - 16,06 \cdot \text{FrequenciaEscolar}_j - 24,79 \cdot \text{GraudeInstrução}_j + 38,43 \\ \cdot \text{MotivoTrabalho}_j + 7,38 \cdot \text{MotivoEstudo}_j - 9,97 \cdot \text{MotivoLazer}_j + 28,25 \\ \cdot \text{MotivoSaúde}_j + 26,36 \cdot \text{MotivoIntegração}_j + 19,85 \cdot \text{PopulaçãoOrigem}_j - 8,32 \\ \cdot \text{Distância}_{30j} + 3,13 \cdot \text{PontosdeOnibus}_{30j}$$

$$V_{31j} = 49,72 + 35,85 \cdot \text{Moradores}_j - 23,86 \cdot \text{VeiculosMotorizados}_j - 32,87 \cdot \text{Renda}_j + 30,48 \\ \cdot \text{Idade}_j - 42,17 \cdot \text{FrequenciaEscolar}_j + 20,14 \cdot \text{Trabalha}_j + 24,76 \\ \cdot \text{GraudeInstrução}_j + 12,83 \cdot \text{MotivoTrabalho}_j - 42,21 \cdot \text{MotivoLazer}_j + 38,64 \\ \cdot \text{MotivoResidência}_j - 21,70 \cdot \text{MotivoSaúde}_j + 39,96 \cdot \text{MotivoIntegração}_j + 28,38 \\ \cdot \text{PopulaçãoOrigem}_j + 9,97 \cdot \text{Distância}_{31j} + 3,13 \cdot \text{PontosdeOnibus}_{31j}$$

$$V_{32j} = -40,18 - 13,80 \cdot \text{Moradores}_j + 21,36 \cdot \text{VeiculosMotorizados}_j + 21,03 \cdot \text{Renda}_j - 28,36 \\ \cdot \text{Genero}_j - 20,36 \cdot \text{Idade}_j + 11,20 \cdot \text{FrequenciaEscolar}_j + 40,80 \\ \cdot \text{GraudeInstrução}_j + 6,10 \cdot \text{MotivoTrabalho}_j + 44,99 \cdot \text{MotivoEstudo}_j + 33,75 \\ \cdot \text{MotivoLazer}_j + 3,01 \cdot \text{MotivoResidência}_j - 37,13 \cdot \text{MotivoSaúde}_j - 24,41 \\ \cdot \text{MotivoIntegração}_j - 49,25 \cdot \text{PopulaçãoOrigem}_j + 27,98 \cdot \text{Distância}_{32j} + 3,13 \\ \cdot \text{PontosdeOnibus}_{32j}$$

$$V_{33j} = 6,15 + 17,20 \cdot \text{Moradores}_j + 0,77 \cdot \text{VeiculosMotorizados}_j - 41,38 \cdot \text{Renda}_j + 26,04 \cdot \text{Idade}_j \\ + 23,83 \cdot \text{FrequenciaEscolar}_j + 26,59 \cdot \text{GraudeInstrução}_j + 14,76 \\ \cdot \text{MotivoTrabalho}_j + 14,27 \cdot \text{MotivoEstudo}_j + 14,62 \cdot \text{MotivoLazer}_j - 16,10 \\ \cdot \text{MotivoResidência}_j + 8,80 \cdot \text{MotivoSaúde}_j - 19,99 \cdot \text{MotivoIntegração}_j + 2,50 \\ \cdot \text{PopulaçãoOrigem}_j + 30,54 \cdot \text{Distância}_{33j} + 3,13 \cdot \text{PontosdeOnibus}_{33j}$$

$$V_{34j} = 41,16 + 41,28 \cdot \text{Moradores}_j + 28,97 \cdot \text{VeiculosMotorizados}_j + 8,18 \cdot \text{Renda}_j - 5,11 \cdot \text{Idade}_j \\ + 23,47 \cdot \text{FrequenciaEscolar}_j + 39,21 \cdot \text{Trabalha}_j + 32,64 \cdot \text{GraudeInstrução}_j \\ + 9,00 \cdot \text{MotivoTrabalho}_j - 26,77 \cdot \text{MotivoLazer}_j + 36,51 \cdot \text{MotivoResidência}_j \\ + 1,12 \cdot \text{MotivoSaúde}_j + 43,25 \cdot \text{MotivoIntegração}_j + 3,62 \cdot \text{PopulaçãoOrigem}_j \\ + 36,82 \cdot \text{Distância}_{34j} + 3,13 \cdot \text{PontosdeOnibus}_{34j}$$

$$V_{35j} = 36,83 - 22,69 \cdot \text{Moradores}_j + 1,80 \cdot \text{VeiculosMotorizados}_j - 32,72 \cdot \text{Renda}_j + 19,51 \\ \cdot \text{Genero}_j - 25,57 \cdot \text{Idade}_j - 33,12 \cdot \text{FrequenciaEscolar}_j + 21,74 \cdot \text{Trabalha}_j \\ - 15,08 \cdot \text{GraudeInstrução}_j + 8,65 \cdot \text{MotivoTrabalho}_j - 15,55 \cdot \text{MotivoLazer}_j \\ - 36,62 \cdot \text{MotivoResidência}_j - 18,90 \cdot \text{MotivoSaúde}_j - 28,93 \cdot \text{MotivoIntegração}_j \\ + 46,61 \cdot \text{PopulaçãoOrigem}_j - 33,28 \cdot \text{Distância}_{35j} + 3,13 \cdot \text{PontosdeOnibus}_{35j}$$

**APÊNDICE G - Calibração por Algoritmos Genéticos com o menor valor para a função  
objetivo no cenário restrito**

$V_{1j} = 0,11 \cdot Distância_{1j} + 2,01 \cdot PontosdeOnibus_{1j}$
$V_{2j} = -10,11 + 1,50 \cdot GraudeInstrução_j + 5,08 \cdot Trabalha_j - 0,28 \cdot MotivoEstudo_j - 1,75$ $\cdot MotivoIntegração_j - 2,45 \cdot MotivoLazer_j + 2,52 \cdot MotivoResidência_j - 7,56$ $\cdot MotivoSaúde_j + 4,39 \cdot Distância_{2j} + 2,01 \cdot PontosdeOnibus_{2j}$
$V_{3j} = 2,34 \cdot MotivoEstudo_j - 7,80 \cdot MotivoIntegração_j + 3,20 \cdot MotivoLazer_j - 2,03 \cdot Distância_{3j}$ $+ 2,01 \cdot PontosdeOnibus_{3j}$
$V_{4j} = 2,99 \cdot MotivoEstudo_j - 13,25 \cdot MotivoIntegração_j + 3,93 \cdot PopulaçãoOrigem_j - 3,28$ $\cdot Distância_{4j} + 2,01 \cdot PontosdeOnibus_{4j}$
$V_{5j} = 0,46 - 5,14 \cdot GraudeInstrução_j + 1,31 \cdot Idade_j + 0,54 \cdot Trabalha_j + 4,55 \cdot Moradores_j - 4,79$ $\cdot MotivoEstudo_j - 4,69 \cdot MotivoIntegração_j + 5,26 \cdot MotivoLazer_j + 4,60$ $\cdot MotivoResidência_j - 4,35 \cdot MotivoSaúde_j + 1,66 \cdot MotivoTrabalho_j - 10,86$ $\cdot Distância_{5j} + 2,01 \cdot PontosdeOnibus_{5j}$
$V_{6j} = -6,48 + 2,57 \cdot Idade_j + 3,92 \cdot MotivoTrabalho_j - 5,31 \cdot MotivoEstudo_j - 3,21 \cdot MotivoLazer_j$ $+ 4,30 \cdot MotivoResidência_j + 3,54 \cdot MotivoSaúde_j + 5,11 \cdot Renda_j + 2,01$ $\cdot PontosdeOnibus_{6j}$
$V_{7j} = 1,56 - 9,31 \cdot MotivoIntegração_j - 9,09 \cdot MotivoLazer_j - 7,89 \cdot MotivoSaúde_j - 1,15$ $\cdot PopulaçãoOrigem_j - 5,38 \cdot Renda_j - 2,86 \cdot Distância_{7j} + 2,01 \cdot PontosdeOnibus_{7j}$
$V_{8j} = 4,69 \cdot GraudeInstrução_j + 0,11 \cdot Idade_j + 4,46 \cdot MotivoIntegração_j - 2,00$ $\cdot MotivoResidência_j + 0,045 \cdot MotivoTrabalho_j - 0,63 \cdot Distância_{8j} + 2,01$ $\cdot PontosdeOnibus_{8j}$
$V_{9j} = -11,78 + 2,46 \cdot FrequênciaEscolar_j + 5,54 \cdot MotivoEstudo_j + 3,20 \cdot MotivoTrabalho_j + 3,49$ $\cdot MotivoResidência_j + 4,87 \cdot MotivoSaúde_j - 2,83 \cdot MotivoLazer_j + 2,01$ $\cdot PontosdeOnibus_{9j}$
$V_{10j} = 0,35 \cdot GraudeInstrução_j + 2,94 \cdot Idade_j + 0,35 \cdot Moradores_j + 5,31 \cdot MotivoLazer_j - 0,49$ $\cdot MotivoResidência_j - 2,49 \cdot Renda_j - 4,95 \cdot Distância_{10j} + 2,01$ $\cdot PontosdeOnibus_{10j}$
$V_{11j} = 2,33 \cdot Moradores_j - 10,44 \cdot MotivoIntegração_j - 6,86 \cdot Distância_{11j} + 2,01$ $\cdot PontosdeOnibus_{11j}$
$V_{12j} = 4,20 \cdot FrequênciaEscolar_j - 5,41 \cdot Moradores_j + 5,81 \cdot VeículosMotorizados_j - 1,56$ $\cdot MotivoIntegração_j - 10,24 \cdot MotivoResidência_j + 1,75 \cdot MotivoSaúde_j - 8,36$ $\cdot Renda_j + 2,01 \cdot PontosdeOnibus_{12j}$
$V_{13j} = -4,14 + 3,81 \cdot Idade_j + 1,79 \cdot Moradores_j + 4,42 \cdot MotivoEstudo_j + 1,79$ $\cdot MotivoResidência_j + 1,45 \cdot Renda_j - 3,17 \cdot Distância_{13j} + 2,01$ $\cdot PontosdeOnibus_{13j}$
$V_{14j} = -9,97 + 4,94 \cdot FrequênciaEscolar_j + 2,00 \cdot VeículosMotorizados_j - 2,46 \cdot MotivoEstudo_j$ $+ 3,12 \cdot MotivoIntegração_j - 7,86 \cdot MotivoResidência_j + 5,45 \cdot PopulaçãoOrigem_j$ $+ 2,01 \cdot PontosdeOnibus_{14j}$
$V_{15j} = 1,37 - 1,60 \cdot GraudeInstrução_j + 3,36 \cdot VeículosMotorizados_j - 6,05 \cdot MotivoIntegração_j$ $+ 2,19 \cdot MotivoResidência_j - 0,80 \cdot MotivoTrabalho_j + 4,37 \cdot PopulaçãoOrigem_j$ $+ 0,34 \cdot Renda_j - 5,61 \cdot Distância_{15j} + 2,01 \cdot PontosdeOnibus_{15j}$
$V_{16j} = -0,27 + 2,72 \cdot MotivoEstudo_j + 1,99 \cdot Distância_{16j} + 2,01 \cdot PontosdeOnibus_{16j}$
$V_{17j} = -3,90 + 0,67 \cdot Moradores_j - 8,63 \cdot MotivoEstudo_j + 0,04 \cdot MotivoIntegração_j + 4,63$ $\cdot MotivoLazer_j + 3,68 \cdot MotivoResidência_j - 4,71 \cdot MotivoSaúde_j - 6,38$ $\cdot Distância_{17j} + 2,01 \cdot PontosdeOnibus_{17j}$
$V_{18j} = 4,14 - 0,73 \cdot PopulaçãoOrigem_j - 1,50 \cdot Renda_j - 1,34 \cdot Trabalha_j + 0,15 \cdot Distância_{18j}$ $+ 2,01 \cdot PontosdeOnibus_{18j}$
$V_{19j} = 2,97 \cdot Moradores_j - 7,27 \cdot MotivoIntegração_j - 2,96 \cdot MotivoTrabalho_j - 4,53 \cdot Distância_{19j}$ $+ 2,01 \cdot PontosdeOnibus_{19j}$

---


$$V_{20j} = 3,62 \cdot \text{GraudeInstrução}_j - 1,34 \cdot \text{Moradores}_j + 2,21 \cdot \text{MotivoResidência}_j - 6,28$$

$$\cdot \text{Distância}_{20j} + 2,01 \cdot \text{PontosdeOnibus}_{20j}$$


---


$$V_{21j} = 1,29 \cdot \text{MotivoLazer}_j + 0,35 \cdot \text{MotivoResidência}_j - 6,07 \cdot \text{Distância}_{21j} + 2,01$$

$$\cdot \text{PontosdeOnibus}_{21j}$$


---


$$V_{22j} = 3,60 \cdot \text{GraudeInstrução}_j - 0,19 \cdot \text{MotivoTrabalho}_j - 5,58 \cdot \text{Distância}_{22j} + 2,01$$

$$\cdot \text{PontosdeOnibus}_{22j}$$


---


$$V_{23j} = 4,33 \cdot \text{FrequênciaEscolar}_j + 1,82 \cdot \text{MotivoIntegração}_j - 10,26 \cdot \text{MotivoLazer}_j + 1,02$$

$$\cdot \text{MotivoResidência}_j + 0,19 \cdot \text{MotivoSaúde}_j - 2,64 \cdot \text{Distância}_{23j} + 2,01$$

$$\cdot \text{PontosdeOnibus}_{23j}$$


---


$$V_{24j} = 4,66 \cdot \text{Idade}_j - 0,20 \cdot \text{Moradores}_j + 4,16 \cdot \text{VeículosMotorizados}_j - 11,30$$

$$\cdot \text{MotivoIntegração}_j - 12,33 \cdot \text{Distância}_{24j} + 2,01 \cdot \text{PontosdeOnibus}_{24j}$$


---


$$V_{25j} = -1,55 \cdot \text{MotivoIntegração}_j - 4,46 \cdot \text{MotivoLazer}_j - 2,29 \cdot \text{MotivoResidência}_j - 1,36$$

$$\cdot \text{MotivoSaúde}_j - 6,27 \cdot \text{Distância}_{25j} + 2,01 \cdot \text{PontosdeOnibus}_{25j}$$


---


$$V_{26j} = -0,70 - 3,37 \cdot \text{GraudeInstrução}_j + 2,59 \cdot \text{MotivoEstudo}_j - 1,43 \cdot \text{MotivoIntegração}_j + 3,95$$

$$\cdot \text{MotivoLazer}_j + 5,00 \cdot \text{MotivoResidência}_j + 4,68 \cdot \text{MotivoSaúde}_j + 2,09$$

$$\cdot \text{MotivoTrabalho}_j + 3,70 \cdot \text{PopulaçãoOrigem}_j - 4,44 \cdot \text{Renda}_j - 2,11 \cdot \text{Distância}_{26j}$$

$$+ 2,01 \cdot \text{PontosdeOnibus}_{26j}$$


---


$$V_{27j} = -2,97 \cdot \text{MotivoIntegração}_j + 0,86 \cdot \text{MotivoResidência}_j + 4,50 \cdot \text{Distância}_{27j} + 2,01$$

$$\cdot \text{PontosdeOnibus}_{27j}$$


---


$$V_{28j} = -3,34 \cdot \text{MotivoSaúde}_j + 2,34 \cdot \text{Distância}_{28j} + 2,01 \cdot \text{PontosdeOnibus}_{28j}$$


---


$$V_{29j} = -1,38 - 0,01 \cdot \text{FrequênciaEscolar}_j + 6,53 \cdot \text{Moradores}_j + 2,38 \cdot \text{MotivoIntegração}_j + 0,69$$

$$\cdot \text{MotivoLazer}_j - 5,72 \cdot \text{MotivoResidência}_j + 4,12 \cdot \text{MotivoSaúde}_j + 2,01$$

$$\cdot \text{PontosdeOnibus}_{29j}$$


---


$$V_{30j} = 0,79 - 0,22 \cdot \text{MotivoIntegração}_j + 0,30 \cdot \text{MotivoTrabalho}_j - 7,15 \cdot \text{Distância}_{30j} + 2,01$$

$$\cdot \text{PontosdeOnibus}_{30j}$$


---


$$V_{31j} = 2,59 + 0,26 \cdot \text{FrequênciaEscolar}_j + 2,98 \cdot \text{Moradores}_j - 7,39 \cdot \text{VeículosMotorizados}_j - 8,13$$

$$\cdot \text{MotivoIntegração}_j + 2,29 \cdot \text{MotivoResidência}_j - 8,21 \cdot \text{MotivoSaúde}_j + 1,40$$

$$\cdot \text{MotivoTrabalho}_j - 9,25 \cdot \text{Distância}_{31j} + 2,01 \cdot \text{PontosdeOnibus}_{31j}$$


---


$$V_{32j} = 2,59 \cdot \text{Moradores}_j + 1,39 \cdot \text{MotivoResidência}_j - 4,28 \cdot \text{Renda}_j + 3,58 \cdot \text{Distância}_{32j} + 2,01$$

$$\cdot \text{PontosdeOnibus}_{32j}$$


---


$$V_{33j} = -8,72 \cdot \text{MotivoIntegração}_j + 5,15 \cdot \text{MotivoSaúde}_j - 13,24 \cdot \text{MotivoTrabalho}_j - 5,89$$

$$\cdot \text{Distância}_{33j} + 2,01 \cdot \text{PontosdeOnibus}_{33j}$$


---


$$V_{34j} = 3,13 - 1,32 \cdot \text{GraudeInstrução}_j - 3,55 \cdot \text{MotivoIntegração}_j + 0,49 \cdot \text{MotivoResidência}_j$$

$$+ 0,84 \cdot \text{MotivoTrabalho}_j + 2,10 \cdot \text{PopulaçãoOrigem}_j - 2,72 \cdot \text{Distância}_{34j} + 2,01$$

$$\cdot \text{PontosdeOnibus}_{34j}$$


---


$$V_{35j} = -9,40 \cdot \text{VeículosMotorizados}_j - 6,93 \cdot \text{MotivoIntegração}_j + 0,03 \cdot \text{MotivoLazer}_j - 11,26$$

$$\cdot \text{MotivoSaúde}_j + 2,01 \cdot \text{PontosdeOnibus}_{35j}$$


---

## ANEXO A – Formulários da Pesquisa Domiciliar

## Dados referentes ao domicílio

PESQUISA O/D DOMICILIAR									
Zona:	<input type="text"/>	Nº do Setor Censitário:	<input type="text"/>	Quadra:	<input type="text"/>	Face:	<input type="text"/>	Ordem:	<input type="text"/>
Endereço:									
<b>Abertura de domicílio</b>									
1ª Visita:	<input type="text"/>	Status:	<input type="text"/>	2ª Visita:	<input type="text"/>	Status:	<input type="text"/>	3ª Visita:	<input type="text"/>
Data:	/ /			Data:	/ /			Data:	/ /
Hora:	:			Hora:	:			Hora:	:
1 - Não Encontrado      7 - Ausência 2 - Em obras            8 - Entrevista realizada parcialmente 3 - Desocupado        9 - Entrevista realizada integralmente 4 - Lazer 5 - Recusa 6 - Ninguém em casa									
<b>Recuperação de moradores</b>									
1º agendamento	<input type="text"/>	Resultado	<input type="text"/>	2º agendamento	<input type="text"/>	Resultado	<input type="text"/>	3 - Todos os membros entrevistados	
Data:	/ /			Data:	/ /			4 - Esgotadas as tentativas	
Hora:	:			Hora:	:				
Informante / respondente					Contatos / telefones				
Número de Quartos					Quantidade de pessoas com 7 anos ou mais				
<input type="text"/>					<input type="text"/>				
99 - ns					99 - ns				
Veículos					Preencher: 0- nenhum				
Automóvel uso particular					<input type="text"/>				
Moto uso particular					<input type="text"/>				
99 - ns					99 - ns				
Supervisor nº	<input type="text"/>	Data:	/ /	Codificador nº	<input type="text"/>	Data:	/ /	Verificador nº	<input type="text"/>
Observações:									

## Dados referentes à pessoa

INDICADORES SOCIAIS DOS RESIDENTES										N.º Pessoas		
Zona:	<input type="text"/>	Nº do Setor Censitário:	<input type="text"/>	Quadra:	<input type="text"/>	Face:	<input type="text"/>	Ordem:	<input type="text"/>	<input type="text"/>		
N.º	Nome da pessoa	Telefone	G		F		H		F		I	Data da entrevista
			Sexo	Idade	Freq. Esc.	Grau Inst.	Possui Hab.	Cond. Ativ.				
01												
02												
03												
04												
05												
06												
07												
08												
09												
10												
11												
12												
F			G			H						
Frequência Escolar/ Possui Habilitação			Sexo			Grau de Instrução						
1 - Sim      8 - Não se aplica 0 - Não      9 - NS / NR			1 - Masculino 2 - Feminino			01 - Analfabeto      09 - Superior incompleto 02 - Pré-escolar /alfabetização [até 1º ano]      10 - Superior completo 03 - 1º grau incompleto (1ª a 4ª série) - [2º ao 5º ano]      11 - Menor de 7 anos não estudaria 04 - 1º grau completo (4ª série) - [5º ano]      12 - Nenhum dos anteriores 05 - 1º grau incompleto (5ª a 8ª série) - [6º ao 9º ano]      90 - Não sabe 06 - 1º grau completo (8ª série) - [9º ano] 07 - 2º grau incompleto 08 - 2º grau completo						
Condição de Atividade			01 - Trabalho Formal      03 - Não trabalha 02 - Trabalho informal									

Dados referentes ao deslocamento

DESMOÇAMENTOS REALIZADOS NO DIA ANTERIOR À ENTREVISTA														Zona	Nº do Setor Comibário	Quadra	Face	Ordem					
ORIGEM							DESTINO																
Nº da Pessoa	ENDEREÇO			A Cód. Local	B Código da Zona	C Cód. Motivo	Hora de Saída	Módulo Transp.	Tempo de Origem	Tempo de viagem	ENDEREÇO			A Cód. Local	B Código da Zona	C Cód. Motivo	Hora de Chegada	Tempo de Destino	D Forma Pago	E Valor Estac.	F Forma Estac.	G Valor Estac.	
	RUA, Nº (ponto de referência)	Bairro									RUA, Nº (ponto de referência)	Bairro											
01																							
02																							
03																							
04																							
05																							
06																							
07																							
08																							
09																							
10																							

A	B	C	D	E	
Código do Local	Motivo na Origem e no Destino	Modo de Transporte	Forma de Pagamento	Estacionamento	
	01 - Residência 02 - Trabalho (negócios) 03 - Estudo 04 - Lazer 05 - Saúde 06 - Compras 07 - Integração 08 - Outros	01 - Ônibus municipal 02 - Ônibus Intermunicipal 03 - Ônibus executivo 04 - Ônibus pirata 05 - Van 06 - Barco/Catamarã 07 - Transp. Escolar 08 - Transp. Fretado 09 - Condutor de auto 10 - Passageiro de auto	11 - Táxi 12 - Moto-táxi 13 - Motocicleta 14 - Bicicleta/Ciclomotor 15 - A pé 16 - Trem 17 - Metrô 18 - Bonde 19 - Outros	01 - Dinheiro 03 - Gratuidade (idoso) 04 - Gratuidade (deficiente) 05 - Gratuidade (estudante) 06 - Cartão Vale Transporte 07 - Cartão Eletrônico 08 - Não se Aplica	01 - Via Pública Gratuita 02 - Via Pública Paga 03 - Estac. Gratuito 04 - Estac. Pago 05 - Vaga Própria 06 - Não estacionou 07 - Não se aplica