

SÍLVIA VITALI SANTOS MAUAD

**Análise de abordagens comportamentais para modelagem de escolha
de destinos em viagens a trabalho**

São Paulo

2022

SÍLVIA VITALI SANTOS MAUAD

**Análise de abordagens comportamentais para modelagem de escolha
de destinos em viagens a trabalho**

Versão Corrigida

Dissertação apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção do título de
Mestre em Ciências

Área de Concentração:
Engenharia de Transportes

Orientador:
Prof. Dr. Cassiano Isler

São Paulo
2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 05 de Janeiro de 2023

Assinatura do autor: Sílvia Vitali Santos Mauad

Assinatura do orientador: 

Catálogo-na-publicação

Mauad, Sílvia Vitali Santos

Análise de abordagens comportamentais para modelagem de escolha de destinos em viagens a trabalho / S. V. S. Mauad -- versão corr. -- São Paulo, 2022.

94 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Transportes.

1.Planejamento de transportes 2.Modelo de escolha discreta 3.Modelo de escolha de destinos 4.Minimização do arrependimento aleatório 5.Maximização da utilidade aleatória I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Transportes II.t.

Agradecimentos

Gostaria de iniciar agradecendo ao meu pai, que apesar de não estar mais comigo em presença física, desde sempre me apoiou muito e tinha orgulho de mim e do meu ingresso no mestrado. Agradeço minha mãe pelo companheirismo que só cresce e amor transmitido sempre em tudo o que realizo. Ao meu namorado pelo empréstimo do computador possante para a realização dos modelos, e pela motivação constante. Agradeço a toda a minha família e amigos.

Gostaria de agradecer ao Professor Dr. Cassiano Isler, pela orientação no desenvolvimento da pesquisa e, principalmente, pela sua compreensão com tudo o que passei na pandemia e verdadeiro interesse em auxiliar o desenvolvimento dos trabalhos, não só no meu, como de todos os seus alunos.

Gostaria de agradecer aos colegas de mestrado: Thaís Baracat, que esteve comigo desde o início, juntas no período das aulas presenciais e depois de forma *online*, compartilhando as dificuldades e alívios. Ao Gabriel Caldeira, pelo apoio no desenvolvimento do tema e no fornecimento de materiais úteis a pesquisa.

Agradeço à GPO Sistran por toda a compreensão e apoio desde a entrada no mestrado e durante os momentos em que tinha as aulas presenciais. Por fim, agradeço a Universidade de São Paulo e a Escola Politécnica, por me proporcionar a oportunidade de me formar e dar continuidade aos meus estudos com o mestrado.

Resumo

Mauad, S. V. S. **Análise de abordagens comportamentais para modelagem de escolha de destinos em viagens a trabalho**. 2022. Dissertação (Mestrado em Engenharia de Transportes) – Escola Politécnica da Universidade de São Paulo, São Paulo, 2022.

Os modelos de escolha discreta baseados na abordagem de maximização da utilidade aleatória (*Random Utility Maximization, RUM*) são alternativos aos tradicionais modelos agregados de distribuição de viagens. Por outro lado, a abordagem de minimização do arrependimento aleatório (*Random Regret Minimization, RRM*) surgiu como opção ao *RUM* na análise de fenômenos comportamentais. Nesse sentido, o objetivo desta dissertação é comparar o desempenho de modelos de escolha discreta sob as abordagens *RUM* e *RRM* para estimativa da escolha de destinos em viagens por motivo trabalho. Para atingir este objetivo foram comparados os resultados da aplicação das abordagens *RUM* e *RRM* na modelagem da escolha de destinos em viagens urbanas a trabalho na cidade de São Paulo utilizando dados da pesquisa Origem-Destino de 2017 do tipo Preferência Revelada realizada pela Companhia do Metropolitano de São Paulo. Foram aplicados métodos de redução do conjunto de escolha de alternativas e de amostragem dos conjuntos de escolhas disponíveis aos indivíduos. Inicialmente, foi utilizado o método para definição do número ótimo de alternativas em modelos *RRM*, haja vista a complexidade computacional para estimativa de modelos envolvendo conjuntos com grande número de escolhas possíveis. Para a seleção do conjunto de alternativas disponíveis aos indivíduos foi aplicado o método de amostragem estratificada baseado no índice de Moran. Na especificação dos modelos foram testadas diferentes variáveis relacionadas a empregos de forma quantitativa (empregos totais e por tipo) e qualitativa (variáveis *dummies* espaciais de localização de empregos), e variável de impedância (distância). Os resultados dos modelos em termos de log-verossimilhança final, ρ^2 , Akaike Information Criteria (AIC) e Bayesian Information Criteria (BIC) foram similares para ambas as abordagens no caso estudado. Entretanto, os modelos *RRM* exigiram tempo de processamento médio de 52 minutos para as estimativas, enquanto os modelos *RUM* foram estimados em 32 segundos em média. Ainda, os modelos com os atributos de quantidade total de empregos e quantidade total de empregos por tipo (indústria e serviços) apresentaram melhor desempenho nas duas abordagens, apesar de todos os atributos considerados terem apresentado significância estatística nos modelos testados.

Palavras-chave: Planejamento de transportes. Modelos de escolha discreta. Modelos de escolha de destinos.

Abstract

Mauad, S. V. S. **Analysis of behavioral approaches for destination choice modelling in business trips.** 2022. Thesis (Master's degree in Transportation Engineering) – Escola Politécnica da Universidade de São Paulo, São Paulo, 2022.

Discrete choice models based on the Random Utility Maximization (*RUM*) approach are alternatives to the traditional aggregated trip distribution models. On the other hand, the Random Regret Minimization (*RRM*) approach has emerged as an option to *RUM* in the analysis of behavioral phenomena. The aim of this thesis is to compare the performance of discrete choice models under the *RUM* and *RRM* approaches to estimate destination choice in trips for work reasons in the city of São Paulo. To achieve this objective, this thesis compares the results of the *RUM* and *RRM* approaches applied to model the destination choice of urban business trips in the city of São Paulo using data from the Origin-Destination survey of 2017, Revealed Preference type carried out by Companhia do Metropolitano de São Paulo. Methods for reducing choice set of alternatives and sampling methods to define available alternatives to individuals were applied. Initially, the developed method to define the optimal number of alternatives in *RRM* models was used, given the computational complexity for estimating models involving sets with many possible choices. To select the choice set, the stratified sampling of alternatives method, based on the Moran index, was applied. In specifying the models, different variables related to jobs were tested quantitatively (total jobs and jobs by type) and qualitatively (spatial *dummy* variables of job location), and impedance variable (distance). The results of the models in terms of final log-likelihood, ρ^2 , Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) were similar for both approaches in the case studied. However, the *RRM* models required an average processing time of 52 minutes, while the *RUM* models were estimated at 32 seconds on average. Furthermore, the models with the attributes of total number of jobs and total number of jobs by type (industry and services) showed better performance in both approaches, despite all considered attributes having presented statistical significance in tested models.

Keywords: Transport planning. Discrete choice models. Destination choice models.

Lista de Figuras

Figura 1 – Fluxograma do método aplicado nesta dissertação.....	35
Figura 2 – Histograma das distâncias em linha reta e em rede entre os centroides das zonas.....	39
Figura 3 – Número de empregos nas zonas OD do município de São Paulo.....	41
Figura 4 – Número de empregos do setor secundário nas zonas OD do município de São Paulo.....	42
Figura 5 – Número de empregos do setor terciário nas zonas OD do município de São Paulo.....	43
Figura 6 – Zonas pertencentes à área central e ao distrito Sé	44
Figura 7 – Zonas com maior concentração de Zona de Centralidade	46
Figura 8 – Zonas com maior concentração de uso comércio, serviços e indústrias e armazéns.....	47
Figura 9 – Tempo de execução de modelos <i>RRM</i> comparado aos modelos <i>RUM</i> em função do número de alternativas (J).....	50
Figura 10 – Valores de β EMPLN obtidos para cada J	60
Figura 11 – Valores de σ EMPLN obtidos para cada J	60
Figura 12 – Valores de β DIS obtidos para cada J.....	61
Figura 13 – Valores de σ DIS obtidos para cada J.....	61
Figura 14 – Valores de log-verossimilhança iniciais e finais dos modelos	63
Figura 15 – Tempos de processamento computacional para estimativa dos modelos.....	64
Figura 16 – Percentual de acertos dos modelos	65

Lista de Tabelas

Tabela 1 – Comparação entre as teorias de maximização da utilidade aleatória (<i>RUM</i>) e minimização do arrependimento aleatório (<i>RRM</i>).....	17
Tabela 2 – Modelos de escolha de destinos na literatura estimados por modelos de escolha discreta	25
Tabela 3 – Estatística descritiva das variáveis quantitativas estudadas	40
Tabela 4 – Correlação entre variáveis quantitativas	48
Tabela 5 – Correlação entre variáveis dummies e variáveis selecionadas.....	48
Tabela 6 – Variáveis selecionadas	49
Tabela 7 – Modelos de escolha de destinos e variáveis utilizadas.....	56
Tabela 8 – Medidas de desempenho dos modelos <i>RUM</i> e <i>RRM</i>	62
Tabela 9 – Resultados dos modelos <i>RUM</i> e <i>RRM</i>	66

Sumário

LISTA DE FIGURAS	8
LISTA DE TABELAS	9
SUMÁRIO	10
1 INTRODUÇÃO	9
1.1 OBJETIVO	11
1.2 JUSTIFICATIVA.....	11
2 REVISÃO DA LITERATURA.....	13
2.1 MODELOS DE ESCOLHA DISCRETA	13
2.1.1 Modelos de Maximização da Utilidade Aleatória.....	14
2.1.2 Modelos de Minimização do Arrependimento Aleatório.....	15
2.2 MODELOS DE ESCOLHA DE DESTINOS	17
2.2.1 Motivo das viagens e variáveis	19
2.2.2 Definição do conjunto de escolha de destinos	23
2.3 MÉTODOS DE AMOSTRAGEM.....	27
2.3.1 Amostragens probabilísticas	28
2.3.2 Amostragens não probabilísticas.....	32
3 MÉTODO.....	35
3.1 BASE DE DADOS.....	36
3.1.1 Filtro da base de dados e caracterização das viagens.....	37
3.1.2 Seleção de variáveis	38
3.2 DEFINIÇÃO DO NÚMERO ÓTIMO DE ALTERNATIVAS	49
3.3 REDUÇÃO DO CONJUNTO DE ESCOLHA DE DESTINOS.....	54
3.4 ESPECIFICAÇÃO DOS MODELOS.....	55
3.5 VALIDAÇÃO	58
4 RESULTADOS.....	59
4.1 NÚMERO ÓTIMO DE ALTERNATIVAS E CONJUNTOS DE ESCOLHA DE DESTINOS	59
4.2 MODELOS DE ESCOLHA DE DESTINOS	62
4.2.1 Medidas de desempenho	62
4.2.2 Parâmetros	65
5 CONCLUSÕES.....	69
6 REFERÊNCIAS BIBLIOGRÁFICAS.....	72

1 INTRODUÇÃO

Os modelos de distribuição de viagens no contexto da modelagem tradicional de demanda de transportes, inicialmente, foram propostos com enfoque na estimativa agregada da quantidade de viagens entre origens e destinos, por exemplo, a partir da construção de matrizes por modelos sintéticos do tipo gravitacional (ORTÚZAR; WILLUMSEN, 2011). Por outro lado, modelos desagregados têm sido propostos na literatura como alternativa no contexto da estimativa de viagens (KIM; LEE, 2017b; MISHRA *et al.*, 2013; PITOMBO; DE SOUZA; LINDNER, 2017; YANG *et al.*, 2010). A abordagem dos modelos de escolha discreta para estimativa da probabilidade de escolha de destinos do ponto de vista desagregado segundo a modelagem proposta por McFadden (1972) e Ben-Akiva (1973) é comumente denominada na literatura como *destination choice models* (MOLLOY, 2016), em português, modelos de escolha de destinos. Enquadram-se, portanto, em uma categoria de modelagem de distribuição de viagens a partir de modelos de escolha discreta.

Tais modelos contribuem para melhor explicar o comportamento das viagens ao utilizarem maior quantidade de variáveis relacionadas à impedância, acessibilidade, atributos de destinos e do usuário, e permitem incluir a autocorrelação espacial nas análises (TRAVEL FORECASTING RESOURCE, 2019). Por exemplo, modelos sintéticos agregados como o gravitacional geralmente consideram o tempo de viagem ou distância, enquanto modelos de escolha de destinos consideram tais variáveis como impedância, além de renda e demais características socioeconômicas, proximidade de destinos similares, conveniência dos destinos, caminhabilidade, uso do solo e autocorrelação espacial.

Os modelos de escolha discreta comumente utilizados para modelagem de escolha de destino são os baseados na Maximização da Utilidade Aleatória (*Random Utility Maximization-RUM*). A abordagem assume que quando o usuário possui à sua disposição um conjunto de alternativas, a escolhida será aquela de maior utilidade como função de atributos relativos à respectiva alternativa (JING *et al.*, 2018).

Uma abordagem alternativa ao *RUM* baseia-se na Minimização do Arrependimento Aleatório (*Random Regret Minimization-RRM*) a partir da Teoria do Arrependimento (CHORUS, 2010; CHORUS; ARENTZE; TIMMERMANS, 2008; LOOMES; GRAHAM AND SUGDEN, 1982). Tal abordagem preconiza que o indivíduo busca minimizar o arrependimento quanto à escolha de uma alternativa em comparação com as demais disponíveis no conjunto de escolha, baseando-se na comparação entre os níveis dos atributos da primeira

em relação às outras alternativas.

Aplicações da abordagem *RRM* têm sido recorrentes em várias áreas de análise comportamental, inclusive na Engenharia de Transportes, como segurança viária (KAPLAN; PRATO, 2012), transporte de carga (BOERI; MASIERO, 2014; KEYA; ANOWAR; ELURU, 2019) e demanda recreacional (THIENE; BOERI; CHORUS, 2012). No contexto brasileiro, Caldeira e Isler (2019) compararam as abordagens *RRM* e *RUM* na escolha de modos de transporte em viagens urbanas.

No contexto de escolha de destinos, no entanto, a aplicação da abordagem *RRM* ainda é escassa, com a pesquisa de Chorus (2010) sendo a única identificada até o momento segundo a revisão da literatura realizada nesta pesquisa. O autor propôs nova formulação para o *RRM* considerando que o arrependimento é experimentado em relação a todas as alternativas que possuem bom desempenho, e não somente em relação à melhor alternativa; indica ainda aplicações do *RRM* em modelos de escolha de rota, escolha de estacionamento, obtenção de informações de transporte público e escolha de destino a shopping centers.

Neste contexto, a pergunta de pesquisa estabelecida nesta dissertação é: Qual é o desempenho da abordagem da Minimização do Arrependimento Aleatório (*RRM*) na modelagem de escolha de destinos comparada à Maximização da Utilidade Aleatória (*RUM*)? A hipótese é que a abordagem *RRM* pode representar de forma satisfatória o comportamento dos indivíduos na escolha de destinos e seu desempenho pode ser igual ou superior à abordagem *RUM*.

Para analisar a hipótese formulada, escolhas de destinos foram modeladas no contexto de viagens realizadas por motivo trabalho no município de São Paulo considerando variáveis de impedância, número de empregos nos destinos e localização geográfica quanto a regiões que concentram empregos. Os dados utilizados foram referentes aos deslocamentos observados pela Pesquisa Origem e Destino 2017 do tipo Preferência Revelada na Região Metropolitana de São Paulo (Companhia do Metropolitano de São Paulo, 2017), porém restritos ao município de São Paulo cujo conjunto de escolhas consiste em 342 destinos (zonas Origem-Destino, OD). Assim, dado o elevado número de alternativas de escolha de destino no contexto estudado, esta pesquisa também utiliza métodos de amostragem de destinos na cidade de São Paulo que viabilizaram a aplicação adequada dos modelos de escolha discreta analisados.

1.1 OBJETIVO

O objetivo principal da pesquisa é avaliar o desempenho de modelos de escolha discreta sob a abordagem da Minimização do Arrependimento Aleatório (*RRM*) comparados à abordagem de Maximização da Utilidade Aleatória (*RUM*) para estimativa desagregada de escolha de destinos em viagens urbanas.

Para atingir o objetivo principal são estabelecidos os seguintes objetivos específicos:

- Definir um método de amostragem para reduzir o conjunto de destinos disponíveis para modelagem de escolha discreta.
- Estimar modelos de escolha discreta sob as abordagens *RUM* e *RRM* no problema de escolha de destino em viagens por motivo trabalho no contexto da cidade de São Paulo.
- Comparar o desempenho dos modelos *RUM* e *RRM* considerando os atributos dos destinos e das viagens.

1.2 JUSTIFICATIVA

Os modelos de demanda são parte essencial do processo de planejamento de transportes, tal que as pesquisas sob esta temática buscam explicar de maneira mais adequada os fenômenos acerca do comportamento dos indivíduos em viagens. No que se refere aos modelos de distribuição de viagens, dadas as limitações dos modelos gravitacionais, estudar o potencial dos modelos de escolha discreta e testar o poder preditivo das abordagens existentes apresentam-se como temas relevantes de pesquisa.

Além disso, a estimativa de modelos de escolha de destinos no contexto de viagens urbanas a trabalho na cidade de São Paulo contribuem para discussões sobre a aplicação do *RRM* no contexto de decisões sobre viagens no Sul Global e, em geral, sobre a aplicabilidade da abordagem em modelos de escolha de destinos.

Outro tema abordado nesta pesquisa é a necessidade de redução do conjunto de alternativas de destinos disponíveis para escolha, dada a grande quantidade de alternativas disponibilizadas nas 342 zonas OD de São Paulo. Desta forma, esta pesquisa contribui com extensões da aplicação de métodos de amostragem para reduzir os conjuntos de escolha em modelos de escolha de destinos considerando ambas as abordagens (*RUM* e *RRM*).

2 REVISÃO DA LITERATURA

Segundo Ortúzar e Willumsen (2011), a quantidade de viagens realizadas por indivíduos é geralmente estimada por modelos de geração de viagens. Esses modelos quantificam viagens na origem e no destino e são agregados no nível de zonas de tráfego. Entretanto, o número de viagens na origem e destino não é suficiente para conhecer os padrões de realização dessas viagens, ou seja, de onde e para onde ocorrem, os quais são estimados pelos denominados modelos de distribuição de viagens.

Assim, padrões de deslocamentos podem ser representados de forma agregada por meio de uma matriz Origem Destino (OD) que representa o número de viagens distribuídas por zonas, ou de forma desagregada por meio de modelos de escolha discreta no nível do indivíduo considerando os fatores que motivam a realização da viagem.

A modelagem de distribuição de viagens é uma etapa importante na previsão de demanda de transportes no contexto da técnica 4 etapas, sendo necessárias previsões de destinos com base na produção e atração de viagens. Essas interações entre origens e destinos são complexas, tornando a modelagem relativamente difícil em termos de acurácia das previsões.

Segundo Yang et al. (2010), existem diversas vantagens para a utilização dos modelos desagregados, em especial os de escolha de destino. Nesses modelos o comportamento de escolha em nível individual é de fácil interpretação e permite a utilização de maior quantidade de variáveis como características socioeconômicas dos indivíduos e atributos de destinos.

Os modelos de escolha discreta são apresentados na próxima seção do ponto de vista das diferentes abordagens utilizadas para caracterizar as escolhas dos indivíduos. Em seguida, são apresentados os resultados de pesquisas sobre a modelagem de escolhas de destinos e as técnicas de amostragem para seleção das alternativas disponíveis para as escolhas individuais.

2.1 MODELOS DE ESCOLHA DISCRETA

Os modelos de escolha discreta são desagregados, pois a entidade de tomada de decisão é o indivíduo, ao considerarem as características das alternativas e individuais como idade, gênero, educação e renda em um conjunto de alternativas mutuamente exclusivas (BEN-AKIVA e LERMAN, 1985; BEN-AKIVA e BIERLAIRE, 1999). Tais modelos assumem que a probabilidade de um indivíduo escolher uma determinada alternativa é função de suas características socioeconômicas e da sua atratividade relativa.

Segundo Ben-Akiva e Bierlaire (1999), o processo de análise das alternativas consiste no conhecimento não somente daquelas escolhidas, mas também das não escolhidas, sendo necessário realizar suposições sobre as opções disponíveis que o indivíduo considera durante o processo de escolha. Ao conjunto de alternativas consideradas por um indivíduo durante o processo de escolha é dado o nome de “conjunto de escolha” e cada alternativa possui atributos que as descrevem.

2.1.1 Modelos de Maximização da Utilidade Aleatória

Os modelos de escolha discreta mais recorrentes na literatura são baseados na Maximização da Utilidade Aleatória (*Random Utility Maximization-RUM*), assumindo que quando o usuário possui várias alternativas à sua disposição, escolherá aquela que possui maior utilidade (JING *et al.*, 2018). A tomada de decisão ocorre com a avaliação ponderada dos atributos que caracterizam uma alternativa, resultando em um valor denominado “utilidade”.

A utilidade é definida como um construto que o indivíduo busca maximizar. Para prever se uma alternativa será escolhida, deve-se comparar o valor de sua utilidade com os respectivos valores de utilidade das outras alternativas (ORTÚZAR; WILLUMSEN, 2011).

Segundo Ben-Akiva e Bierlaire (1999), modelos de utilidade aleatória pressupõem que o tomador de decisão tem capacidade de julgamento perfeita, enquanto o analista tem a informação incompleta. Portanto, a incerteza deve ser considerada tal que o indivíduo q tem uma utilidade associada à alternativa i , U_{iq} , conforme a Equação (1).

$$U_{iq} = V_{iq} + \varepsilon_{iq} = \sum_{k=1}^K \beta_{ik} x_{ikq} + \varepsilon_{iq} \quad \forall i \in I, q \in Q \quad (1)$$

em que V_{iq} é a parcela determinística da utilidade e ε_{iq} é o termo aleatório não observável/mensurável. A parcela determinística é uma função dos atributos das alternativas e dos indivíduos, formalmente definida pela ponderação do nível do k -ésimo atributo relativo à alternativa i para o indivíduo q (x_{ikq}) e o peso do k -ésimo atributo da alternativa i (β_{ik}).

Sob a hipótese de que os erros aleatórios são independentes e identicamente distribuídos (*Independent and Identically Distributed - IID*) com distribuição Valor Extremo Tipo I, a probabilidade de escolha da alternativa i por um indivíduo q , P_{iq} , é dada por um modelo Logit Multinomial conforme a Equação (2). A probabilidade da alternativa i no conjunto de escolha C_Q ser escolhida pelo indivíduo q é dada pela Equação (3) (BEN-AKIVA; BIERLAIRE, 1999).

A alternativa com a maior probabilidade é a escolhida.

$$P_{iq} = \frac{e^{V_{iq}}}{\sum_{j \in I} e^{V_{jq}}} \quad (2)$$

$$P(i | C_Q) = P[U_{iq} > U_{jq} \forall j \in C_Q] = P[U_{iq} = \max_{j \in C_Q} U_{jq}] \quad (3)$$

2.1.2 Modelos de Minimização do Arrependimento Aleatório

Segundo Jing *et al.* (2018), uma abordagem alternativa ao *RUM* é a de Minimização do Arrependimento Aleatório (*Random Regret Minimization-RRM*) baseada na Teoria do Arrependimento; quando o indivíduo percebe que a alternativa não escolhida é melhor que a escolhida, tem-se uma emoção chamada arrependimento. É a sensação sentida quando uma alternativa abandonada tem melhor desempenho do que a escolhida (CHORUS, 2010).

Os modelos caracterizados sob essa abordagem baseiam-se na comparação entre os atributos das alternativas para prever a escolha daquela que minimiza o arrependimento. A abordagem de minimização do arrependimento no contexto de modelagem de escolha discreta tem sido aplicada em uma variedade de disciplinas, como marketing, microeconomia, psicologia, ciências de gestão e transporte (ZEELENBERG, 1999; INMAN; DYER; JIANMIN, 1997; SIMONSON, 1992; BELL, 1982; CALDEIRA, 2021).

A teoria do arrependimento considera que um indivíduo baseia sua escolha entre alternativas com o desejo de evitar a situação em que uma alternativa que não foi escolhida seja mais atrativa do que aquela selecionada, causando arrependimento. Assim, considera-se que o indivíduo deseja minimizar o arrependimento antecipado no momento da escolha das alternativas (CHORUS; ARENTZE; TIMMERMANS, 2008).

Chorus, Arentze e Timmermans (2008) propuseram um modelo *RRM* posteriormente modificado por Chorus (2010). O modelo proposto inicialmente tem limitações pois a função de verossimilhança da especificação não é contínua, criando dificuldades à medida que requer implementação personalizada de rotinas computacionais, dificultando a aplicabilidade geral do modelo.

O modelo proposto por Chorus (2010) pressupõe que o arrependimento é antecipado com relação a todas alternativas abandonadas que têm melhor desempenho do que a escolhida em termos de um ou mais atributos. O modelo proposto incorre em uma função de

arrependimento diferenciável que pode ser estimada utilizando softwares existentes.

Segundo Chorus (2010), o arrependimento é a soma de todos os arrependimentos binários da comparação bilateral da alternativa considerada com as outras alternativas no conjunto de escolha. O nível de arrependimento binário pela comparação da alternativa considerada com outra alternativa j é dado pela soma dos arrependimentos associados à comparação das duas alternativas em termos de cada um de seus atributos m conforme a Equação (4).

$$R_{i \leftrightarrow j} = \sum_{m=1 \dots M} R_{i \leftrightarrow j}^m \quad (4)$$

O nível do arrependimento no atributo é igual a zero se a alternativa considerada obtiver desempenho melhor que a outra alternativa, ou é igual à diferença ponderada no desempenho do atributo caso contrário. Cada parcela $R_{i \leftrightarrow j}^m$ é representada pela Equação (5) em que os valores β_m representam pesos associados aos atributos e os erros aleatórios independentes e identicamente distribuídas (IID) v_{0m} e v_{xm} são adicionados para representar a heterogeneidade não observada nos pesos dos atributos.

$$R_{i \leftrightarrow j}^m = \max\{0 + v_{0m}; \beta_m \times (x_{jm} - x_{im}) + v_{xm}\} \quad (5)$$

Para obter as probabilidades de escolha, os erros devem ser integrados conforme a Equação (6) na comparação entre as alternativas i e j em termos do atributo m .

$$\int_v [R_{i \leftrightarrow j}^m \times f(v)] dv = \ln(1 + \exp[\beta_m \times (x_{jm} - x_{im})]) \quad (6)$$

Deste modo, o arrependimento \tilde{R}_i pode ser descrito conforme a Equação (7) e a probabilidade de escolha associada à alternativa i é calculada conforme a Equação (8).

$$\tilde{R}_i = \sum_{j \neq i} \sum_{m=1 \dots M} \ln(1 + \exp[\beta_m \times (x_{jm} - x_{im})]) \quad (7)$$

$$P_i = \frac{e^{-\tilde{R}_i}}{\sum_{j \in J} e^{-\tilde{R}_i}} \quad (8)$$

Os modelos de escolha discreta baseados em arrependimento ganharam popularidade como abordagem alternativa ao *RUM*. O modelo da Equação (8) baseado na abordagem *RRM* pode ter probabilidades e parâmetros estimados com *softwares* convencionais de escolha discreta. A tabela, a seguir, resume as diferenças entre as abordagens *RUM* e *RRM* descritas nesta seção, cujas características são apresentadas no que se refere à teoria, objetivo, critério, estratégia, estrutura e probabilidade calculada.

Tabela 1 – Comparação entre as teorias de maximização da utilidade aleatória (*RUM*) e minimização do arrependimento aleatório (*RRM*)

Característica	<i>RRM</i>	<i>RUM</i>
Teoria	Teoria do arrependimento	Teoria da utilidade
Objetivo	Escolher o menor arrependimento entre alternativas	Escolher a máxima utilidade entre as alternativas
Critério	Função de arrependimento, descrevendo o arrependimento quando o decisor escolhe uma alternativa	Função de utilidade, descrevendo a satisfação quando o decisor realiza uma escolha
Estratégia	Semi-compensatória	Compensatória
Estrutura	$R_i = R_i + \varepsilon_i = \sum_{j \neq i} \sum_m \ln\{1 + \exp[\beta_m \times (x_{jm} - x_{im})]\} + \varepsilon_i$	$U_i = U_i + \varepsilon_i = \sum_m \beta_m \times x_{im} + \varepsilon_i$
Probabilidade calculada	$P_i = P(R_i < R_j), \forall j \neq i$	$P_i = P(U_i > U_j), \forall j \neq i$

Fonte: JING *et al.* (2018).

2.2 MODELOS DE ESCOLHA DE DESTINOS

Os modelos desagregados de escolha de destinos são alternativas às técnicas convencionais de previsão de distribuição de viagens por modelos agregados na modelagem de demanda de transportes.

Segundo Ortúzar e Willunsem (2011), os modelos de demanda agregada são estabelecidos com base nas relações ao nível da zona. Por outro lado, os modelos de demanda desagregada são baseados em escolhas no nível dos indivíduos ou domicílios. Assim, espera-se que esses últimos sejam mais realistas. Apesar disso, os modelos agregados continuaram a ser usados na maioria dos projetos de planejamento de transportes até início dos anos 1980.

Na literatura são apresentadas vantagens de utilização dos modelos desagregados, em especial os de escolha de destino. Neste tipo de modelo, o comportamento de escolha em seu nível individual é de fácil interpretação, viabilizando que o comportamento em viagens possa ser melhor explicado por permitir a utilização de maior quantidade de variáveis. Tais modelos utilizam as variáveis comuns em modelos gravitacionais como tempo de viagem e distância (impedância), mas também permitem considerar características socioeconômicas dos indivíduos, acessibilidade, e atributos de destinos, como uso do solo, barreiras físicas, além da autocorrelação espacial (YANG *et al.*, 2010; TRAVEL FORECASTING RESOURCE, 2019).

Os modelos de distribuição de viagens baseados nas teorias de escolha discreta foram inicialmente propostos e aplicados nos anos 1980 (MISHRA *et al.*, 2013). Introduzidos por McFadden e Daly (1982), os modelos desagregados de escolha de destinos são úteis para descrever o comportamento de um decisor que escolhe o seu destino de viagem baseado em características de atratividade a partir da origem, custo generalizado, e características socioeconômicas de uso do solo e geográficas que são computadas na forma de função utilidade (KIM; LEE, 2017a).

Ortúzar e Willumsen (2011) e Molloy (2016) citam outras vantagens desse tipo de modelo notadas por Spear (1977). Os modelos desagregados são baseados em teorias do comportamento individual e, portanto, são mais estáveis no tempo e no espaço do que os modelos agregados. Podem ser mais eficientes em termos de uso de informações pois cada escolha individual é utilizada como observação, tal que na abordagem agregada, a observação pode ser a média de centenas ou milhares de observações. Além disso, os modelos são menos propensos a sofrer vieses devido à correlação entre unidades agregadas, pois o comportamento individual pode ser ocultado por características não identificadas associadas às zonas.

Além disso, as variáveis explicativas incluídas no modelo podem ter coeficientes estimados pois a função utilidade permite qualquer número e especificação das variáveis explicativas, ao contrário da função custo generalizado em modelos agregados que geralmente é limitada aos custos entre pares OD. Portanto, os modelos desagregados permitem uma representação mais flexível das variáveis consideradas relevantes para o estudo e os coeficientes das variáveis explicativas refletem a importância relativa de cada atributo.

Simma, Schlich e Axhausen (2001) citam algumas questões particulares do uso do modelo logit multinomial baseado na maximização da utilidade para os modelos de escolha de destinos:

- Suposição do IIA (*Independence from Irrelevant Alternatives*): quando duas alternativas possuem probabilidade diferente de zero de serem escolhidas, a razão entre as probabilidades não é afetada pela presença ou ausência de qualquer alternativa adicional (LUCE; SUPPES, 1965);
- Homogeneidade dos viajantes: presume-se que diferentes pessoas reagem de forma homogênea em resposta aos atributos das alternativas. Esta suposição em geral não é realista, já que o comportamento e vivência das pessoas são diferentes;
- Questões espaciais: a demanda de viagens é influenciada por dependência, heterogeneidade e heterocedasticidade espaciais (BHAT; ZHAO, 2002). A dependência espacial descreve fatores não observados que influenciam o comportamento de viagem. A heterogeneidade trata da relação entre variáveis dependentes e independentes que variam entre unidades espaciais distintas. Por fim, a heterocedasticidade reflete que a variância do comportamento não observado também pode ser diferente entre unidades espaciais distintas.

2.2.1 Motivo das viagens e variáveis

Os modelos de escolha de destinos tratam de diversos motivos, em geral modelados principalmente para viagens a trabalho e lazer. As viagens por motivo trabalho são amplamente estudadas (BEKHOR; PRASHKER, 2008; CHOW *et al.*, 2005; KIM *et al.*, 2009; LI *et al.*, 2005; MISHRA *et al.*, 2013; MOLLOY, 2016; SADHU; TIWARI, 2016; YANG *et al.*, 2010) por serem cotidianas, em grande número e que têm impacto no tempo disponível para realização das atividades pelas pessoas.

As viagens por motivo lazer também são relevantes devido, em alguns casos, às variáveis envolvidas não estarem relacionadas à impedância consideradas, por exemplo, em modelos gravitacionais. Molloy (2016) realizou a estimação de modelos de escolha destino de viagens a lazer no Canadá e justificou que, por exemplo, parques nacionais não possuem população e tem poucos empregos, mas atraem muitas viagens. Nesse caso, o modelo gravitacional em geral não permite estimar de maneira acurada o número de viagens atraídas por aqueles locais.

Para estimar os modelos, Van Wee (2002) destacou a importância da utilização de variáveis espaciais na análise do comportamento de viagens. Outros autores como Badoe e

Miller (2000), Stead (2001) e Handy, Cao e Mokhtarian (2005) indicaram uso do solo, ambiente construído, características de densidade e acessibilidade como variáveis espaciais relevantes para estimação de modelos de escolha de destinos (HAMMADOU *et al.*, 2008). Badoe e Miller (2000) demonstraram, inclusive, que quanto maior a concentração de empregos, maior o impacto no comportamento das viagens.

O enfoque desta pesquisa é nos modelos de escolha de destinos por motivo trabalho. Para esses modelos, as variáveis comumente utilizadas são número de empregos por tipo (indústrias, comércio e serviços, por exemplo) e indicadores de localização onde se concentram empregos, renda, distância e custo de viagem. A seguir, são mencionadas algumas variáveis relevantes consideradas por autores em modelos de escolha de destinos.

Hammadou *et al.* (2008) utilizaram diversos tipos de variáveis (uso do solo, atratividade e acessibilidade) na modelagem de escolha de destinos para viagens por motivo não-trabalho. As variáveis de uso do solo foram relativas ao uso residencial (densidade de residências, área construída, áreas verdes existentes), industrial, comercial, áreas portuárias, áreas verdes e abertas, e infraestrutura de transportes (presença de rodovias, estradas, aeroportos e portos). As variáveis de atratividade consideradas foram população nas zonas de origem e destinos, empregos, locais para compras e presença de escolas. Por fim, as variáveis de acessibilidade utilizadas foram a menor distância e menor tempo de viagem entre centroides de origens e destinos, desconsiderando congestionamento, tempos de espera em semáforos e conversões.

Mishra *et al.* (2013) modelaram diversos motivos de viagem, no total seis (trabalho, compras e outros com origem na residência, e trabalho, compras e outros, com origem fora da residência) e utilizaram as seguintes variáveis: escolha modal, faixas de distâncias, faixas de renda, tipos de emprego (comércio, escritório, industrial e outros), número de residências, localização de centros de empregos e das regiões suburbanas da cidade, e presença de barreira física (ponte). Entre essas variáveis, as de escolha modal, de faixas de distância, de empregos resultaram significativas.

Kim e Lee (2017) também modelaram viagens denominadas mandatórias (a trabalho e estudo) e viagens de outros tipos (compras, lazer e outros). Os autores consideraram variáveis de impedância, medidas quantitativas e não quantitativas, e variáveis indicadoras. As variáveis de impedância foram tempos de viagem, distância e custo generalizado. As medidas quantitativas representaram um ou mais aspectos de cada destino e foram número de empregados, estudantes e área urbanizada. Medidas não quantitativas e variáveis indicadoras

incluíram concentração de indústrias, características socioeconômicas, e características de uso do solo de cada destino. Para as viagens mandatórias, foram significativas as variáveis de tempo de viagem, número de empregados e estudantes, e concentração industrial.

Molloy (2016) modelou viagens de longa distância por motivo trabalho, lazer, visita e outros. As variáveis socioeconômicas consideradas foram população e emprego, e as variáveis de impedância foram os tempos de viagem por automóvel entre as zonas. O autor ainda utiliza uma abordagem diferenciada por considerar número de *check-ins* realizados do portal *Foursquare*, o número de pessoas que estiveram em determinados locais, utilizando as categorias aeroporto, hotel, saúde, vida noturna, turismo ao ar livre e esqui. Nos modelos de escolha de destinos do tipo trabalho em específico, considerou as seguintes variáveis: exponencial da distância entre as zonas (utilizado também no modelo gravitacional), logaritmo da distância entre zonas, logaritmo da soma entre população e empregos, variáveis binárias em viagens para zonas metropolitanas, zonas rurais, número de check-ins nas categorias hotéis e turismo ao ar livre. A variável de distância não é significativa para o modelo motivo trabalho.

Kim *et al.* (2009) utilizaram apenas variáveis relacionadas à impedância dentro e fora do veículo em um modelo de escolha de destinos: tempo; acessibilidade para considerar o efeito espacial; distância; e exponencial da função de custo generalizado. O custo foi verificado como significativo no poder preditivo do modelo.

Yang *et al.* (2010) modelaram viagens por motivo trabalho com a divisão por categoria. As variáveis foram consideradas de forma distinta em vários modelos testados, sendo as principais: número de estabelecimentos em cada zona por tipo de trabalho; variável binária de identificação se a zona está localizada no centro ou não; variável binária para verificação se a zona residencial e de trabalho são as mesmas; variável para indicação se houve cruzamento pelo rio Cao; impedância (tempo de viagem por automóvel); gênero; idade; renda; e modo de viagem. A variável referente à zona da residência ser a mesma do trabalho resultou positiva, o que reflete que os viajantes tendem a escolher locais de trabalho próximos das suas casas e a variável negativa referente ao rio indica que as pessoas tendem a escolher menos trabalhos do outro lado de uma barreira física. Obteve-se valor negativo para o parâmetro da variável de impedância, o que sugere que a atração do destino é menor quanto maior a impedância.

Li *et al.* (2005) estimaram modelos de escolha de destinos por motivo trabalho com variáveis de tempo de viagem, renda, número de empregos divididos por categorias e presença de barreiras físicas. As análises incluíram faixas de renda e tipos de empregos, concluindo que

os trabalhadores de média renda são mais propensos a viajar para áreas comerciais, industriais e para onde há empregos oficiais.

Simma, Schlich e Axhausen (2001) modelaram viagens por motivo lazer utilizando um modelo para cada categoria: esqui, escalada e caminhada. Para cada um deles as variáveis consideradas foram distintas, como por exemplo, para a categoria esqui, a altura da municipalidade, número de habitantes, área de floresta e número de oportunidades de esquiar. Os autores identificaram que a variável de distância é a mais significativa nesses modelos.

Sadhu e Tiwari (2016) utilizaram um modelo do tipo Logit aninhado para estimar modos de escolha de destinos de pessoas que vivem em assentamentos informais em Delhi. As variáveis consideradas foram socioeconômicas (idade, gênero e educação), características do domicílio (número de crianças, homens, empregados e bicicletas), características do uso do solo como proximidade de áreas industriais e características da viagem como o custo.

Pozsgay e Bhat (2002) estimaram modelos de escolha de destinos por motivo lazer e selecionaram variáveis de impedância (tempos de viagem, distâncias e custos), de uso do solo (uso comercial, e áreas de lazer com presença de água, por exemplo) e características sociodemográficas (idade, presença de crianças, carros e renda). Nenhum atributo relacionado à viagem (número de indivíduos participando da atividade, modo, hora de partida da viagem) foi estatisticamente significativo. As variáveis de impedância foram as mais significativas e as variáveis sociodemográficas foram importantes para explicar a escolha de destinos em viagens por motivo lazer.

Prashker, Shiftan e Hershkovitch-Sarusi (2008) investigaram fatores que influenciam a escolha de residências em viagens a trabalho utilizando variáveis de distâncias por faixa de renda e segmentadas por gênero, tempo de viagem por carro e por ônibus, população, domicílios, percentual de população que trabalha, número de empregos, número de empregos por tipo, situação socioeconômica da zona, distância segmentada por motivo, idade, e divisão da cidade considerando setores próximos ou não do centro comercial.

Chow *et al.* (2005) aplicaram modelos de escolha de destinos por motivo trabalho para três faixas de renda distintas. As variáveis utilizadas foram tempo de viagem, características de uso do solo e de barreira física, tal que esta não se mostrou significativa.

Carrasco (2008) aplicou modelos de escolha de destino para viagens de compras a supermercados e utilizou variáveis como horários de abertura, distância de casa até os locais

por diferentes modos, variáveis de renda, idade, gênero e acessibilidade. Distância e renda foram as variáveis que mais influenciaram na decisão das viagens.

2.2.2 Definição do conjunto de escolha de destinos

A definição do conjunto de escolha é um dos primeiros problemas que o pesquisador se depara em modelos de escolha discreta, pois deve decidir quais alternativas estão disponíveis para cada indivíduo na amostra. Embora na modelagem de escolha de modo o número de alternativas é pequeno, em modelos de escolha de destinos o problema é mais complexo, pois a quantidade de alternativas disponíveis para escolha é muito grande. Portanto, no caso específico de modelos de escolha de destinos o pesquisador deve definir o tamanho ideal do conjunto de alternativas e em seguida escolher quais deverão estar inseridas no conjunto.

Ben-Akiva e Lerman (1985) afirmam que os modelos de escolha de destinos, apesar de serem modelos desagregados e servirem para modelar decisões individuais, ainda necessitam basear-se em algum nível de agregação. Isso se deve porque os destinos disponíveis para escolha podem ser inúmeros e se todos forem incluídos em nível desagregado demandará grande capacidade computacional para estimativa dos parâmetros. Para superar esses obstáculos é importante agregar as alternativas no nível das zonas de tráfego baseando-se em similaridades espaciais e viabilidade de análise (KIM; LEE, 2017).

Mesmo considerando a zona de tráfego como unidade de análise para origens e destinos, observa-se em geral que as regiões de estudo possuem número elevado de zonas como potenciais destinos, incorrendo novamente no problema de grande capacidade computacional necessária.

Assim, uma das formas de lidar com um conjunto grande de alternativas (destinos) disponíveis é considerar apenas subconjuntos das opções potencialmente a serem escolhidas utilizando técnicas de amostragem ou assumindo que todas as alternativas estão disponíveis. No primeiro caso, ao utilizar subconjuntos, é possível omitir opções realistas e no segundo caso, ao assumir todas as alternativas disponíveis, o modelo pode perder capacidades discriminatórias e há aumento de complexidade computacional (ORTÚZAR; WILLUMSEN, 2011).

Segundo Li *et al.* (2005), uma possibilidade de viabilizar a aplicação de modelos de escolha de destino é amostrar um conjunto de zonas de análise de tráfego não escolhidas e aquela que o indivíduo escolheu, para construir um conjunto de escolha de destinos. Dessa forma, a complexidade computacional é reduzida. Os autores mencionam dois tipos de

estratégias para reduzir o conjunto de escolha. A primeira abordagem é a amostragem aleatória simples, sendo selecionado aleatoriamente o conjunto de alternativas não escolhidas com probabilidades iguais e adicionada a alternativa escolhida. A segunda abordagem é a amostragem de importância estratificada que extrai amostras com probabilidades de seleção distintas determinadas com base em estimativas preliminares para cada alternativa. O método de amostragem estratificada atribui diferentes probabilidades de seleção a alternativas em diferentes estratos.

A primeira abordagem pode ser julgada pouco eficiente porque todas as alternativas possuem a mesma probabilidade de escolha. Li *et al.* (2005) afirmam que a segunda estratégia de amostragem é mais eficaz quanto aos valores de máxima verossimilhança resultantes do processo de estimativa dos parâmetros de modelos de escolha discreta.

Guevara, Chorus e Ben-Akiva (2016) apresentaram um método para amostragem de alternativas em aplicações a modelos de escolha discreta baseados na minimização do arrependimento aleatório. A amostragem é especialmente importante em modelos *RRM* uma vez que a função do arrependimento para cada alternativa depende de todas as outras alternativas. Além disso, uma das principais desvantagens do modelo é o esforço computacional necessário nesta comparação entre alternativas.

A Tabela 2 apresenta os artigos identificados na literatura que tratam de modelos de escolha de destinos com o número de observações consideradas, local de aplicação, motivação para pesquisa, as unidades de análise, definição de conjunto de escolha, método utilizado para diminuição do conjunto de escolha e variáveis utilizadas divididas por tipo. As variáveis são divididas naquelas relacionadas à impedância, tais como custo e tempo de viagem ou distância, e as demais são divididas em quantitativas e não quantitativas.

Nota-se que grande parte dos autores, tais como Bhat, Govindarajan e Puugurta (1998), Shiftan (1998), Bowman e Ben-Akiva (2000), Jonnalagadda *et al.* (2001), Chow *et al.* (2005), Li *et al.* (2005), Molloy (2016), Mishra *et al.* (2013), Yang *et al.* (2010) e Prashker, Shiftan e Hershkovitch-Sarusi (2008) trataram de viagens por motivo trabalho, assim como o foco desta dissertação. Além disso, a maioria dos autores mencionados (73%) trabalharam com número elevado de destinos e utilizaram alguma técnica de amostragem para reduzir o conjunto de escolha, sendo que Shiftan (1998), Bowman e Ben-Akiva (2000), Jonnalagadda *et al.* (2001) e Li *et al.* (2005) utilizaram a amostragem estratificada, tal como abordado nesta pesquisa conforme apresentado na próxima seção.

Tabela 2 – Modelos de escolha de destinos na literatura estimados por modelos de escolha discreta

Autor	Assunto	Número de observações	Tamanho total do conjunto de alternativas	Número de alternativas após amostragem	Método de amostragem alternativa	Impedância	Variáveis	
							Variáveis quantitativas	Medidas não quantitativas e indicadores
Bhat, Govindarajan e Puugurta (1998)	Boston, MA (EUA) casa para trabalho e casa para compras	7.337 viagens a trabalho 7.963 viagens para compras	986	7 (0,7%)	Amostragem aleatória simples	Custo generalizado (tempo de viagem, custo)	Número de trabalhadores, número de trabalhadores em comércios/serviços, área	Indicadores de localização de zonas, indicadores de estrutura zonal espacial, gênero, idade, renda
Shifan (1998)	Boise, ID (EUA) viagens a trabalho/não-trabalho	1.600	250	40 (16,0%)	Amostragem com importância estratificada	Tempo de viagem	Número de residências, número de trabalhadores em comércios e não-comércios	Indicador de área central, indicador de área residencial e de trabalhos
Bowman e Ben-Akiva (2000)	Boston, MA (EUA) viagens a trabalho/não-trabalho (inclui modo)	1.901 viagens a trabalho 1.929 viagens não-trabalho	786	8 (1,0%)	Amostragem com importância estratificada	Custo de viagem (automóvel ou não), tempo de viagem, distância	Número de trabalhadores	Renda residencial, número de veículos
Jonnalagadda <i>et al.</i> (2001)	São Francisco, CA (EUA) viagem casa para trabalho e paradas	1.627 viagens a trabalho 976 paradas	1728	40 (2,3%)	Amostragem com importância estratificada	Distância de viagem	Número de trabalhadores, e número de trabalhadores em indústrias	Renda domiciliar, indicador de urbanização e de área central
Pozsgay e Bhat (2001)	Dallas-Fort Worth, TX (EUA) viagem a lazer	777	919	10 (1,1%)	Amostragem aleatória simples	Custo generalizado (tempo de viagem, custo)	Número de veículos	Idade, indicador de área comercial e não-comercial, indicador de domicílios de baixa renda, existência de domicílios com membros entre 5 e 21 anos
Sivakumar e Bhat (2007)	Karsruhe (Alemanha) viagem não-trabalho	903	69	10 (14,5%)	Amostragem aleatória simples	Distância entre casa, trabalho e escola	Área de oportunidades comerciais, oportunidades de lazer, área de uso do solo misto e em desenvolvimento	Densidade populacional, existência de creches e indicador de centro
Molloy (2016)	Modelo de escolha de destinos para viagens de longa distância em Ontario – Canadá, utilizando dados LBSN	69.328 (6.200 viagens a trabalho)	6671 zonas OD	69 (1,0%)	Modelo de agregação de zonas utilizando um método baseado em “raster gradual”	Distâncias em transformação em logaritmo e exponencial	População, empregos, check-ins Foursquare	Variáveis binárias para consideração de viagens em zonas metropolitanas, rurais
Mishra <i>et al.</i> (2013)	Comparação entre modelo gravitacional de escolha de destinos em Maryland	20.626 (motivo trabalho)	8 regiões	8 (100%)	Não utilizou	Distância	Renda, posse de automóvel, empregos, residências	Localização (centro, cruzamento de ponte), escolha de modo
Kim, C. <i>et al.</i> (2009)	Comparação de modelos agregados e desagregados para distribuição de viagens em Seoul, utilizando viagens de metrô	4.000	200 zonas	11 (5,5%)	Amostragem aleatória simples	Tempos de viagem (dentro e fora dos veículos), custo e acessibilidade (distância e custo generalizado)	-	-

Autor	Assunto	Número de observações	Tamanho total do conjunto de alternativas	Número de alternativas após amostragem	Método de amostragem alternativa	Variáveis		
						Impedância	Variáveis quantitativas	Medidas não quantitativas e indicadores
Yang <i>et al.</i> (2010)	Modelagem do comportamento de escolha de destinos, incorporando fatores espaciais, dados sociodemográficos e modo	1.399	22	22 (100%)	-	Tempo de viagem de automóvel entre as zonas OD	Número de estabelecimentos em cada zona por tipo de trabalho, idade, renda	Variável binária de indicação de centro, variável binária para indicação se a zona de trabalho é igual a de residência, variável indicativa de barreira (rio), gênero, modo de viagem
Li <i>et al.</i> (2005)	Aplicação método para reduzir conjunto de escolhas de modelos de escolha de destinos	1.364	892	30 (3,0 %)	Amostragem estratificada	Tempo de viagem de automóvel	Número de trabalhadores em área comercial, serviços, entre outros	Indicador de conexão entre rodovias, indicador de centro
Hammadou <i>et al.</i> (2008)	Incorporação da dimensão espacial em modelos de escolha de destinos na Antuérpia	30.000	608	33 (5,0%)	Agregação de zonas	Distância, tempo de viagem	População, empregos, escolas, locais para compras	Variáveis de uso do solo (relativas ao uso residencial, industrial, comercial, entre outros)
Pozsgay e Bhat ([s. d.])	Modelos de escolha de destino para viagens a lazer e implicações para uso do solo, transporte e qualidade de ar	777	919	10 (1,0%)	Amostragem aleatória simples	Tempo de viagem, distância e custo	Características sociodemográficas (renda, crianças, carros)	Uso do solo (áreas comerciais, de lazer)
Prashker <i>et al.</i> (2008)	Investigação de fatores que influenciam a escolha de residências no papel de viagens a trabalho em Tel Aviv e análise de gênero	9.389	500	20 (4,0%)	Amostragem aleatória simples	Distâncias de viagem por faixa de renda e gênero, tempo de carro, ônibus	População, domicílios, percentual de população que trabalha, número de empregos, idade	Divisão da cidade por proximidade do centro comercial
Chow (2005)	Desenvolvimento e avaliação de modelos de escolha de destino para a Flórida	1.361 / 1.297 / 410	Variável por municipalidade. Não menciona o número exato.	10 (-)	Amostragem aleatória simples	Tempo de viagem	Número de trabalhadores em área comercial, de serviços, entre outros	Uso do solo (comércio, serviços, órgãos públicos, entre outros), barreira física

Fonte: Adaptado de Kim e Lee (2017).

2.3 MÉTODOS DE AMOSTRAGEM

Segundo Fávero e Belfiore (2017), existem situações em que o estudo com todos os elementos da população é impossível ou indesejável, sendo necessário extrair um subconjunto (amostra) da população, que deve ser representativo dessa população. Os resultados obtidos pela amostra podem ser utilizados para inferir conclusões sobre a população.

O processo de obter amostras compreende a seleção do método de amostragem e a determinação do tamanho da amostra. Se o tamanho da amostra for muito pequeno pode não detectar efeitos ou associações importantes, ou pode estimar esses impactos ou associações de forma imprecisa. Da mesma forma, se o tamanho da amostra for muito grande, o estudo pode tornar-se muito complexo e aumentar os custos. Portanto, o tamanho da amostra é um fator essencial de qualquer pesquisa científica (SINGH; MASUKU, 2014).

As amostras podem ser classificadas pelo procedimento de seleção em probabilísticas e não probabilísticas através de critérios objetivos ou subjetivos (BOLFARINE; BUSSAB, 2005). Segundo Costa Neto (2002) e Fávero e Belfiore (2017), a amostragem é probabilística se todos os elementos da população são obtidos de forma aleatória, ou seja, se tiverem probabilidade conhecida ou diferente de zero de pertencer à amostra, mas é não probabilística caso isso não ocorra, ou seja, a probabilidade de pertencer à amostra é desconhecida.

Segundo Fávero e Belfiore (2017), as vantagens das técnicas probabilísticas são que os critérios de seleção dos elementos estão definidos sem subjetividade na escolha dos elementos. Além disso, há a possibilidade de determinar matematicamente a dimensão da amostra em função da precisão e grau de confiança desejado para os resultados. As desvantagens estão relacionadas à dificuldade em obter listagens completas da população originando uma amostra muito dispersa geograficamente, aumentando custos, tempo e dificuldade de coleta de dados. As vantagens das técnicas não aleatórias são o menor custo, menor tempo de estudo e menor necessidade de mão de obra. As desvantagens são que não há garantia que a amostra seja representativa da população, há unidades que não têm possibilidade de serem escolhidas, pode ocorrer viés de opinião pessoal e não se sabe com que grau de confiança as conclusões obtidas podem ser inferidas para a população.

A escolha da técnica de amostragem deve considerar objetivos da pesquisa, erro aceitável nos resultados, acesso aos elementos da população, representatividade desejada, tempo dispendido e a disponibilidade de recursos financeiros e humanos.

2.3.1 Amostras probabilísticas

2.3.1.1. Amostragem aleatória simples

De acordo com Costa Neto (2002), na amostragem aleatória simples todos os elementos da população têm igual probabilidade de pertencer à amostra e todas as amostras possíveis têm probabilidade igual de ocorrer. Essa técnica fornece a estimativa imparcial e melhor dos parâmetros se a população for homogênea (SINGH; MASUKU, 2014). Entretanto, segundo Taherdoost (2018), as desvantagens são a necessidade de uma lista completa de toda a população, o custo em função da obtenção de dados completos de toda a população pode ser muito alto e o erro padrão dos estimadores pode ser alto.

Considerando uma população ou universo U com N elementos, o procedimento aleatório consiste em sortear um elemento da população até que seja retirada uma amostra com n observações. Quando o elemento sorteado for removido de U antes do próximo sorteio, tem-se o processo de amostragem sem reposição. Caso seja permitido o sorteio de uma unidade mais de uma vez, tem-se o processo com reposição. Segundo Bolfarine e Bussab (2005), a amostragem sem reposição é mais interessante, pois não há ganho de informação fazendo com que a mesma unidade apareça novamente. Por outro lado, a reposição traz vantagens matemáticas e estatísticas, como a independência entre as unidades sorteadas (FÁVERO; BELFIORE, 2017).

Conforme Bolfarine e Bussab (2005), na amostragem sem reposição há $C_{N,n} = \binom{N}{n} = \frac{N!}{n!(N-n)!}$ possíveis amostras de n elementos que podem ser extraídas a partir da população, e cada amostra tem a mesma probabilidade, $\frac{1}{\binom{N}{n}}$, de ser selecionada. Na amostragem com reposição, há N^n amostras possíveis de n elementos que podem ser extraídas a partir da população, e cada amostra tem a mesma probabilidade, $\frac{1}{N^n}$, de ser selecionada. Nesse último caso, os autores definem o tamanho da amostra como:

$$n = \frac{\sigma^2}{(B/z_\alpha)^2} = \frac{\sigma^2}{D} \quad (9)$$

Para determinar o tamanho da amostra, é preciso fixar o erro máximo desejado B , com grau de confiança z_α e possuir algum conhecimento a priori sobre a variância da população

σ^2 . Os dois primeiros são fixados pelo pesquisador e o terceiro é baseado em pesquisas passadas ou amostras piloto, por exemplo.

2.3.1.2. Amostragem sistemática

Segundo Costa Neto (2002) e Fávero e Belfiore (2017), quando os elementos da população se apresentarem ordenados e forem retirados em intervalos de tempo sucessivos, tem-se uma amostragem sistemática. A principal vantagem desse tipo de amostragem é a facilidade de determinação dos elementos da amostra, pois pode ser executada com rapidez e baixo custo e estar menos sujeita a erros do entrevistador durante a pesquisa. A desvantagem é a possibilidade de existirem ciclos de variação da variável de interesse, por exemplo, se o período dos ciclos coincidir com o período de retirada dos elementos da amostra.

Neste método de amostragem, a primeira unidade da amostra é selecionada aleatoriamente e as unidades subsequentes são selecionadas de forma sistemática. Se houver N unidades na população e n unidades forem selecionadas, então o intervalo de amostragem é igual a N/n , ou seja, é o quociente entre o tamanho da população e o tamanho da amostra. O primeiro elemento é selecionado aleatoriamente e, então, novos elementos são selecionados pelo intervalo igual a N/n (SINGH, MASUKU; 2014).

2.3.1.3. Amostragem estratificada

De acordo com Singh e Masuku (2014), a amostragem aleatória estratificada é um método útil para coleta de dados se a população for heterogênea. Neste método, toda a população é dividida em um número de grupos, geralmente conhecidos como estratos, em que cada um desses grupos é homogêneo dentro de si, e então, as unidades são amostradas aleatoriamente de cada um desses estratos. O tamanho da amostra em cada estrato varia de acordo com a importância relativa do estrato na população. A técnica de extração dessa amostra é conhecida como amostragem estratificada. Em outras palavras, a estratificação é a técnica pela qual a população é dividida em subgrupos/estratos e a amostragem é realizada separadamente em cada estrato. Os estratos ou subgrupos são escolhidos porque há evidências disponíveis de que eles estão relacionados ao resultado. A seleção de estratos varia de acordo com a área e condições locais. Após a estratificação, a amostragem é realizada separadamente em cada estrato. Neste caso, o erro de amostragem depende da variância da população dentro do estrato, mas não entre os estratos.

A amostragem estratificada consiste em especificar quantos elementos da amostra serão retirados em cada estrato. Dentro dessa abordagem, Costa Neto (2002) cita três tipos de amostragem estratificada: uniforme, proporcional e ótima. Na amostragem uniforme sorteia-se igual número de elementos em cada estrato. Na proporcional, o número de elementos sorteados em cada estrato é proporcional ao número de elementos existentes no estrato. Na ótima, em cada estrato há um número de elementos proporcional ao número de elementos do estrato e à variação da variável de interesse no estrato, medida pelo desvio-padrão, pois quando a variação é menor, menos elementos são necessários para melhor caracterizar o comportamento da variável. Uma das dificuldades do método é justamente não conseguir avaliar antecipadamente o desvio-padrão da variável nos diversos estratos.

Na execução de um plano de amostragem estratificada, a população é dividida em subpopulações bem definidas (estratos) usualmente independentes e de cada estrato é retirada uma amostra. Então, em cada amostra usam-se estimadores convenientes para os parâmetros do estrato e define-se para a população um estimador combinando os estimadores de cada estrato, determinando-se suas propriedades (BOLFARINE; BUSSAB, 2005).

Segundo Freund (2006), se os elementos selecionados em cada estrato constituírem amostras aleatórias simples, o processo global (estratificação seguida de amostragem aleatória) é chamado de amostragem aleatória estratificada (simples).

A amostragem estratificada, segundo Freund (2006), é executada da seguinte forma: uma população de tamanho N é dividida em k estratos de tamanhos N_1, N_2, \dots, N_k ; para cada estrato, uma amostra aleatória de tamanho n_i ($i = 1, 2, \dots, k$) é selecionada, resultando em k subamostras de tamanhos n_1, n_2, \dots, n_k . Na amostragem estratificada uniforme tem-se que $n_1 = n_2 = \dots = n_k$ de modo que o tamanho da amostra extraída de cada estrato é:

$$n_i = \frac{n}{k}, \text{ para } i = 1, 2, \dots, k \quad (10)$$

$$\text{em que } n = n_1 + n_2 + \dots + n_k \quad (11)$$

Na amostragem estratificada proporcional, tem-se que:

$$n_i = \frac{n}{k}, \text{ para } i = 1, 2, \dots, k \quad (12)$$

tal que
$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \dots = \frac{n_k}{N_k} \quad (13)$$

Na amostragem estratificada proporcional, o tamanho da amostra extraída de cada estrato pode ser obtido de acordo com a seguinte expressão:

$$n_i = \frac{N_i}{N} \times n, \text{ para } i = 1, 2, \dots, k \quad (14)$$

De acordo com Bolfarine e Bussab (2005), uma das formas de realizar uma alocação da amostra pelos vários estratos é através da alocação ótima de Neyman, onde

$$n_h = n \times \frac{N_h \times \sigma_h}{\sum_{h=1}^H N_h \times \sigma_h} \quad (15)$$

tal que $\sum_{h=1}^H N_h \times \sigma_h$ é o desvio padrão médio dentro de cada estrato.

O número de unidades a serem observadas no estrato h é proporcional a $N_h \times \sigma_h$.

2.3.1.4. Amostragem por conglomerados

Segundo Costa Neto (2002) e Fávero e Belfiore (2017), na amostragem por conglomerados, ou *clusters*, a população total é subdividida em grupos denominados conglomerados e a amostragem é realizada a partir dos grupos e não dos indivíduos da população. Dessa forma, é sorteado aleatoriamente um número suficiente de conglomerados para compor a amostra.

De acordo com Taherdoost (2018), a amostragem por conglomerados é vantajosa para pesquisadores que trabalham com assuntos fragmentados em grandes áreas geográficas, economizando custos, de maneira que são utilizados os grupos ao invés de todos os indivíduos da área. Segundo Bolfarine e Bussab (2005), uma das inconveniências da amostragem por conglomerados está no fato de que os elementos dentro de um mesmo conglomerado tendem a apresentar características similares.

Conforme Singh e Masuku (2014), a amostragem por conglomerados é geralmente usada quando o pesquisador não consegue obter uma lista completa de uma população que deseja estudar, mas pode obter uma lista completa de grupos ou *clusters* da população. Esse método de amostragem pode ser mais prático e econômico do que a amostragem aleatória simples ou a amostragem estratificada. Em comparação com a amostragem aleatória simples e

a amostragem estratificada, a amostragem por conglomerados apresenta vantagens e desvantagens. Por exemplo, a amostragem por conglomerados geralmente fornece menos precisão, por outro lado, pode ser mais eficaz do que os outros métodos por ter menores custos.

Segundo Fávero e Belfiore (2017), a amostragem por conglomerados em um estágio adota o seguinte procedimento: a população é dividida em M conglomerados (C_1, \dots, C_M) de tamanhos não necessariamente iguais; segundo um plano amostral, geralmente amostragem aleatória simples, sorteiam-se m conglomerados ($m < M$); todos os elementos de cada conglomerado sorteado constituem a amostra global. Já a amostragem por conglomerados em dois estágios é realizada da seguinte forma: a população é dividida em M conglomerados (C_1, \dots, C_M) de tamanhos não necessariamente iguais. Deve-se sortear m conglomerados no primeiro estágio, segundo algum plano amostral, geralmente por amostragem aleatória simples; de cada conglomerado i sorteado de tamanho n_i , sorteiam-se b_i elementos no segundo estágio, conforme o mesmo ou outro plano amostral.

2.3.1.5. Amostragem múltipla

De acordo com Costa Neto (2002), na amostragem múltipla a amostra é retirada em diversas etapas sucessivas. Dependendo dos resultados, etapas a mais podem ser dispensadas. Esse tipo de amostragem é empregado em inspeção quando é importante a amostragem dupla. Sua finalidade é diminuir o número médio de itens inspecionados a longo prazo, reduzindo custos.

2.3.2 Amostragens não probabilísticas

2.3.2.1 Amostragem por quotas

Segundo Taherdoost (2018), na amostragem por cotas todos os elementos são escolhidos com base em características pré-determinadas de modo que a amostra total tenha a mesma distribuição das características da população maior.

A amostragem por quotas consiste em selecionar variáveis de controle ou as características da população consideradas relevantes para o estudo em questão determinando-se a proporção da população (%) para cada uma das categorias das variáveis relevantes. Dimensionam-se as quotas (número de pessoas a serem entrevistadas que possuem as características determinadas) para cada entrevistador, de modo que a amostra tenha proporções

iguais à da população (FÁVERO, BELFIORE; 2017).

As principais vantagens da amostragem por quotas são o baixo custo, a rapidez e a conveniência ou a facilidade para o entrevistador selecionar elementos. Porém, como a seleção dos elementos não é aleatória, não há garantia de que a amostra seja representativa da população, não sendo possível generalizar os resultados da pesquisa para a população (Fávero, Belfiore; 2017).

2.3.2.2 Amostragem bola de neve

De acordo com Taherdoost (2018), a amostragem de bola de neve é aplicável em populações difíceis de acessar. Neste método são identificados um ou mais indivíduos da população-alvo, e estes identificarão outros indivíduos pertencentes à mesma população. O processo é repetido até que seja alcançado o objetivo proposto de formação da amostra, quando os últimos entrevistados não acrescentam novas informações relevantes à pesquisa. Como vantagens, permite ao pesquisador localizar a característica desejada da população, facilidade de aplicação e baixo custo pois necessita de menos planejamento e menor quantidade de indivíduos pesquisados, e eficiência ao penetrar em populações de difícil acesso (FÁVERO, BELFIORE; 2017).

2.3.2.3 Amostragem por conveniência

Conforme Taherdoost (2018), na amostragem por conveniência a amostra é escolhida quando está facilmente disponível, sendo empregada quando a participação é voluntária ou os elementos da amostra são escolhidos por uma questão de conveniência ou simplicidade, por exemplo, amigos, vizinhos ou estudantes. A vantagem deste método é que possibilita obter informações de maneira rápida e barata, entretanto, não garante que a amostra seja representativa da população (FÁVERO, BELFIORE; 2017).

2.3.2.4 Amostragem por julgamento ou intencional

Segundo Fávero e Belfiore (2017), na amostragem por julgamento ou intencional, a amostra é escolhida segundo a opinião ou julgamento prévio de um especialista, trazendo risco decorrente de possível equívoco por parte do pesquisador em seu pré-julgamento. Portanto, requer conhecimento da população e dos elementos selecionados.

3 MÉTODO

O método utilizado nesta dissertação segue as etapas do fluxograma da Figura 1 que serão detalhadas nos subitens a seguir.

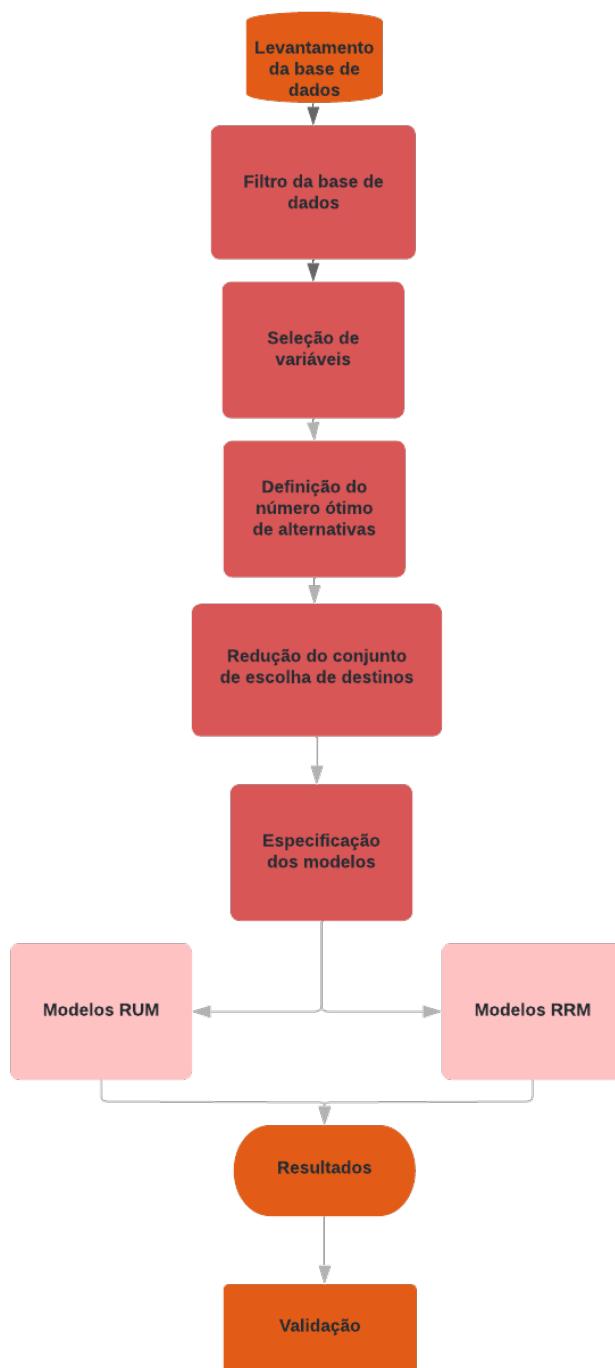


Figura 1 – Fluxograma do método aplicado nesta dissertação
Fonte: Elaboração própria (2022).

3.1 BASE DE DADOS

A base de dados utilizada nesta pesquisa é resultante da Pesquisa Origem e Destino (OD) 2017 da Companhia do Metropolitano de São Paulo-SP, que apresenta informações sobre a mobilidade urbana da Região Metropolitana de São Paulo (RMSP). Os dados levantados referem-se a deslocamentos diários das pessoas e os principais motivos desses deslocamentos, tendo sido realizadas 6 pesquisas em intervalos de 10 anos entre 1967 e 2017 (COMPANHIA DO METROPOLITANO DE SÃO PAULO, 2017).

Na etapa de planejamento da pesquisa de 2017 foi realizada a revisão do zoneamento, que consiste no processo de dividir a RMSP em áreas com características homogêneas no que se refere a critérios urbanísticos, socioeconômicos, de uso e ocupação do solo, de sistema de transportes existente e futuro, e de barreiras físicas. Para definição das zonas, realizou-se a agregação dos setores censitários do Instituto Brasileiro de Geografia e Estatística (IBGE). A Companhia do Metropolitano de São Paulo utiliza essas zonas como base para dimensionamento da amostra a ser pesquisada. Na Pesquisa OD 2017 foram definidas 517 zonas de pesquisa, sendo 342 no município de São Paulo. A partir da definição do zoneamento, foi feito o plano amostral com amostragem estratificada em estratos de renda domiciliar que culminou na definição de uma amostra de 32 mil domicílios (COMPANHIA DO METROPOLITANO DE SÃO PAULO, 2017).

Segundo a Companhia do Metropolitano de São Paulo (2017), 116 mil domicílios foram visitados com o objetivo de contemplar essa amostra domiciliar. As perguntas das entrevistas foram a respeito das viagens diárias realizadas pelos moradores dos domicílios no dia útil anterior à pesquisa e das características socioeconômicas. Os dados foram coletados em *tablets* e tratados para verificação de consistência com base na atribuição de renda familiar para famílias que não quiseram declarar, e aplicação de fatores de expansão para que a amostra representasse as viagens na RMSP.

O resultado foi a obtenção de características de 42 milhões de viagens diárias para a RMSP. Além dos dados socioeconômicos relevantes obtidos, foi possível obter a divisão das viagens por modos e motivos. Os modos pesquisados foram metrô, trem, ônibus, fretado, escolar, automóvel, táxi, moto, bicicleta, a pé e outros. Os motivos foram agrupados em trabalho, educação, compras, saúde, lazer e assuntos pessoais (COMPANHIA DO METROPOLITANO DE SÃO PAULO, 2017).

3.1.1 Filtro da base de dados e caracterização das viagens

Os dados da Pesquisa OD 2017 foram analisados no software R (R Core Team, 2019) e filtrados para se adequarem aos objetivos desta pesquisa conforme descrito a seguir.

- Foram selecionadas apenas as viagens por motivo trabalho no destino. Na Pesquisa OD 2017 representam 44% das viagens, seguida das viagens por motivo estudo (35%), compras (5%), saúde (4%), lazer (4%), assuntos pessoais (8%) (COMPANHIA DO METROPOLITANO DE SÃO PAULO, 2017);
- Foram selecionadas apenas as viagens com início e término no município de São Paulo; filtro aplicado devido à maior disponibilidade de informações sobre os atributos das viagens no município em relação aos demais da região metropolitana;
- Zonas que não possuem empregos registrados foram descartadas. As zonas sem empregos (120-Vista Alegre, 143- Reserva da Cantareira, 146-ETA Guaraú e 179-USP Leste II) foram descartadas pois não podem atrair viagens a trabalho, resultando em 338 zonas de destino possíveis;
- Foram consideradas apenas as viagens pelo modo automóvel (motorista dirigindo);
- Apenas indivíduos com 16 anos ou mais foram incluídos nas análises. Na Constituição Federal, artigo 7º, inciso XXXIII, determina-se que é proibido o trabalho de menores de 16 anos, exceto na condição de aprendiz;
- Apenas observações cujas coordenadas de origem eram iguais às coordenadas da residência e as coordenadas do destino eram iguais ao primeiro ou segundo trabalho mencionado na pesquisa foram selecionadas;
- Observações que apresentaram tempos de viagem iguais a zero ou tempos muito elevados, acima de 3 horas e, portanto, incompatíveis com os percursos realizados, foram eliminadas.

Assim, o processo de tratamento da base original resultou em um banco de dados para modelagem com 4.183 observações, correspondendo a 2,28% do total de observações da Pesquisa OD 2017 ou 633,9 mil viagens.

3.1.2 Seleção de variáveis

Na seleção de variáveis para aplicação dos modelos *RUM* e *RRM* para modelagem desagregada de escolha de destinos considerou-se os apontamentos da revisão da literatura apresentada no capítulo 2.

Considerou-se que as informações presentes na Pesquisa OD 2017 são reportadas pelos indivíduos e desta forma, são percebidas por eles. Assim, não é possível identificar de forma precisa as alternativas disponíveis para escolha e todos os seus atributos. Neste caso, Koppelman e Bhat (2006) recomendam a modelagem de todos os atributos relacionados às alternativas, incluindo as escolhidas.

Inicialmente, foram selecionadas as variáveis tempo e custo para caracterizar a impedância da viagem realizada, e a quantidade de empregos em cada zona de tráfego como variável representativa da característica socioeconômica do destino. Cabe destacar que a modelagem foi realizada exclusivamente para escolha de destino, sem incluir questões de escolha de modo no processo decisório e assumindo-se as escolhas em viagens realizadas por automóvel particular.

O cálculo do tempo de viagem foi realizado pelo pacote *r5r* (PEREIRA *et al.*, 2021) no software estatístico R, que obtém rotas para redes de transporte considerando diferentes modos (a pé, bicicleta, transporte público e automóvel). O método consistiu em gerar centroides para as 338 zonas de tráfego consideradas e calcular os tempos de viagem entre origens e destinos para o modo automóvel. Tal procedimento foi utilizado para viabilizar o cálculo do tempo entre uma dada origem e os destinos que não foram escolhidos pelos indivíduos entrevistados na pesquisa OD 2017. Portanto, no modelo foi necessário estimar o tempo de viagem entre todos os pares OD.

A função utilizada do pacote *r5r* foi a *travel_time_matrix* que realiza o cálculo do tempo de viagem entre pares OD; as origens e destinos são lidos em coordenadas em sistema WGS84 e a função utiliza um algoritmo específico chamado RAPTOR. Este algoritmo calcula caminhos de uma origem para todos os destinos de acordo com o horário de partida definido pelo usuário. O modelo utiliza como rede a base do OpenStreetMap e não considera congestionamentos. Para o modo de transporte automóvel, o *r5r* considera que a velocidade do veículo é o limite legal de velocidade das vias (CONWAY; BYRD; VAN DER LINDEN, 2017).

As distâncias entre os centroides das zonas de tráfego foram obtidas para o cálculo do

custo de viagem pelo modo automóvel conforme a Equação (9) proposta por Gomide e Morato (2011), cujos valores foram atualizados para 2017 pelo Índice de Preços do Consumidor Amplo (IPCA) (BANCO CENTRAL DO BRASIL, 2020).

$$Custo_{Auto} = 2,88 \cdot \frac{2,31 \cdot D}{7000} \quad (16)$$

em que $Custo_{AUTO}$: custo da viagem por automóvel entre uma origem e um destino; e D : distância em rede entre a origem e o destino (metros).

Posteriormente, optou-se por utilizar a variável de distância entre zonas de forma direta, já que as variáveis de tempo e custo estão correlacionadas com a distância conforme explicitado mais adiante. A distância foi calculada em quilômetros entre os centroides das zonas, de forma direta, sem considerar o sistema viário. Cabe ressaltar que a distribuição de distância em linha reta tem a mesma tendência da distribuição de distância em rede, conforme apresentado na Figura 2. A distância média em linha reta é de 15,7 km, enquanto a distância média em rede é de 24 km.; a mediana das distâncias é 14,4 km em linha reta e 21,8 km em rede. Apesar das diferenças, optou-se pela utilização da distância em linha reta pelo menor custo computacional para sua obtenção, útil em aplicações posteriores, e por considerar-se que a significância da variável nos modelos estimados não é afetada dada a similaridade entre as distribuições.

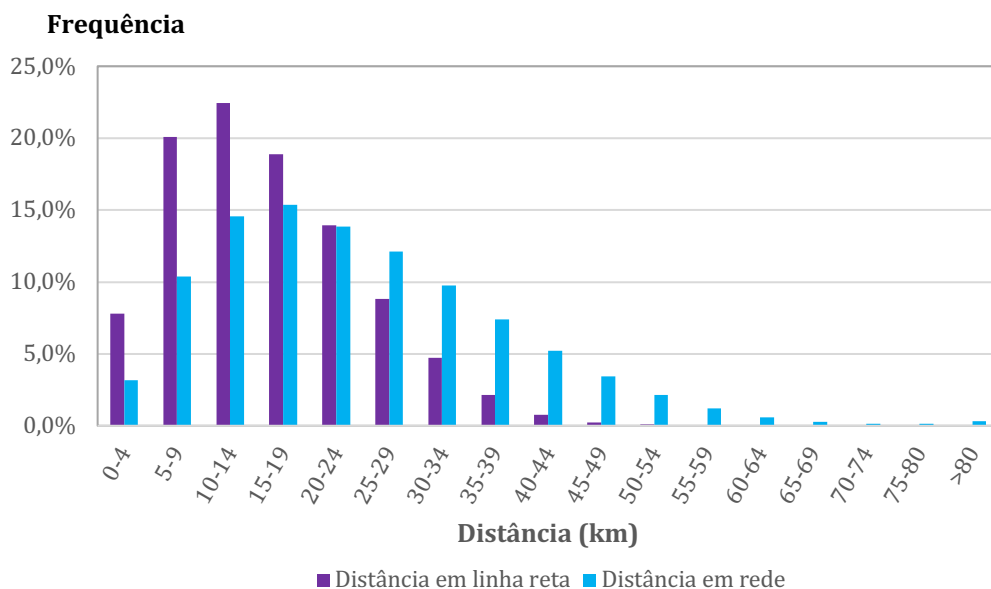


Figura 2 – Histograma das distâncias em linha reta e em rede entre os centroides das zonas

Fonte: Companhia do Metropolitano de São Paulo (2017); Pereira *et al.* (2021).

A variável de número de empregos foi obtida da Pesquisa OD 2017 (COMPANHIA DO METROPOLITANO DE SÃO PAULO, 2017) de acordo com a zona de destino considerada como alternativa conforme ilustrado na Figura 3. Sobre a variável de empregos foi aplicado o logaritmo para que a ordem de grandeza dos respectivos parâmetros estimados pelos modelos fosse próxima daqueles relativos aos demais parâmetros.

Além de considerar a variável de número de empregos totais, foi também incluída a variável de quantidade de empregos por tipo, também obtidas da Pesquisa OD (COMPANHIA DO METROPOLITANO DE SÃO PAULO, 2017) considerando empregos no setor secundário e no setor terciário, distribuídos espacialmente conforme a Figura 4 e Figura 5.

Os empregos no setor “Outros” não foram considerados pela quantidade inferior em relação aos setores secundário e terciário. O setor representa apenas 0,17% dos empregos da Região Metropolitana de São Paulo, enquanto os setores secundário e terciário representam respectivamente 17,27% e 82,56%.

A Tabela 3 apresenta a estatística descritiva das variáveis quantitativas mencionadas anteriormente.

Tabela 3 – Estatística descritiva das variáveis quantitativas estudadas

Estatística	Empregos totais	Empregos setor secundário	Empregos setor terciário	Custo (R\$)	Tempo (minutos)	Distância (km)
Média	17.763	2.349	15.400	7,58	25,1	15,70
Desvio padrão	1.289	1.992	12.043	2,74	11,7	8,93
Mínimo	68	0	0	3,00	1,0	0
Máximo	101.042	14.386	91.125	22,89	108,0	60,63

Fonte: Companhia do Metropolitano de São Paulo (2017); Gomide e Morato (2011); Pereira *et al.* (2021).

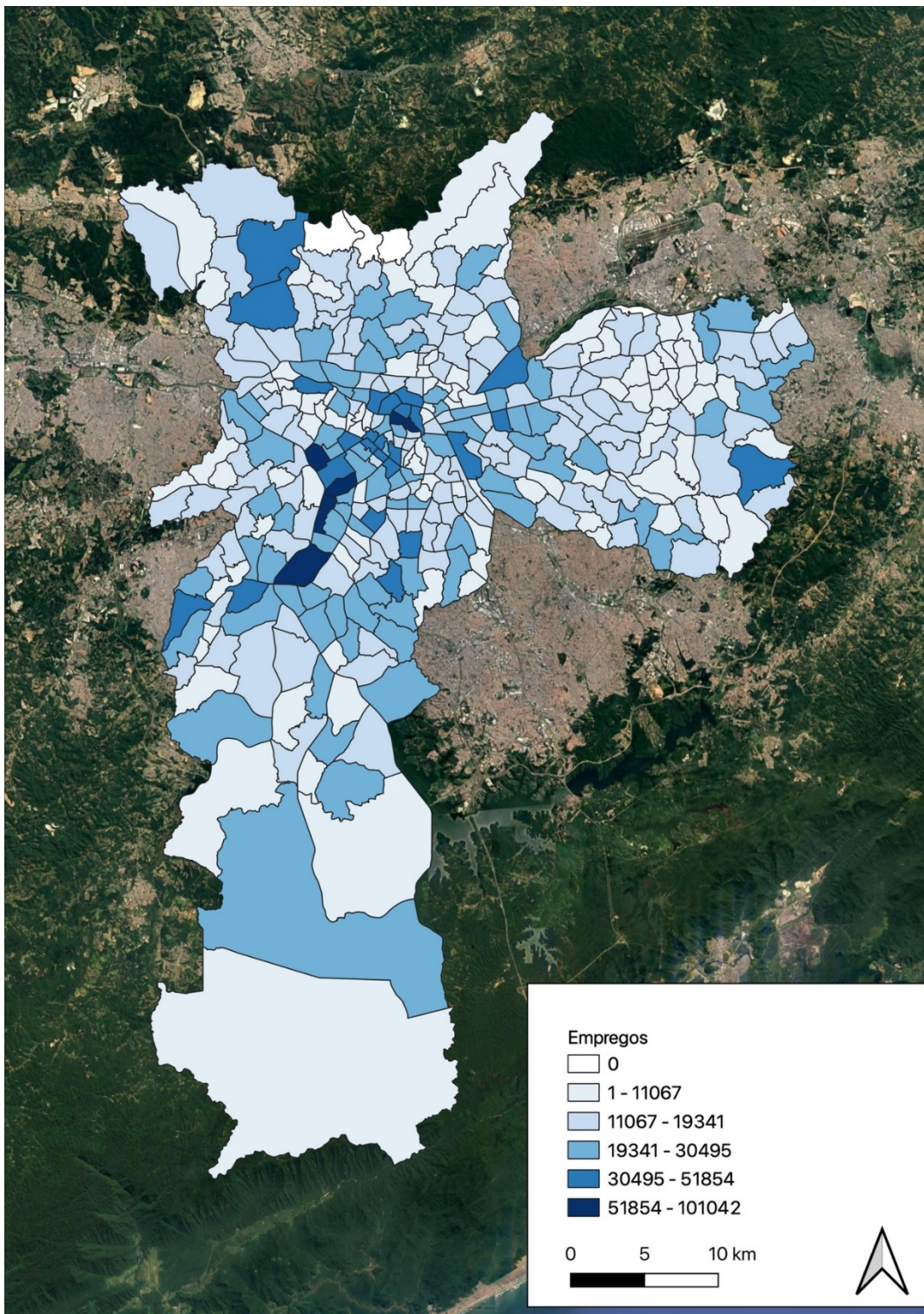


Figura 3 – Número de empregos nas zonas OD do município de São Paulo

Fonte: Companhia do Metropolitano de São Paulo (2017); Prefeitura Municipal de São Paulo (2016).

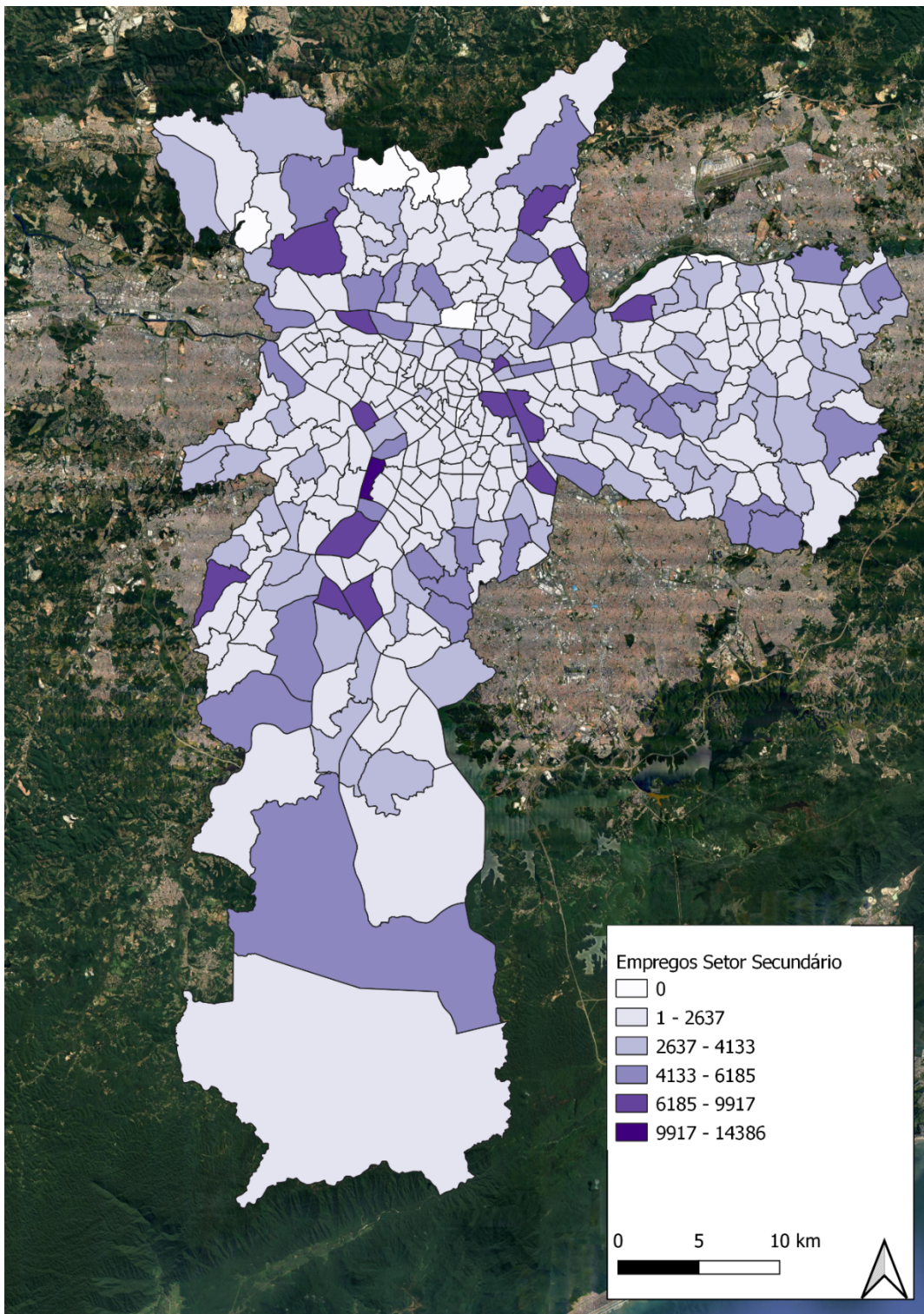


Figura 4 – Número de empregos do setor secundário nas zonas OD do município de São Paulo
 Fonte: Companhia do Metropolitano de São Paulo (2017).

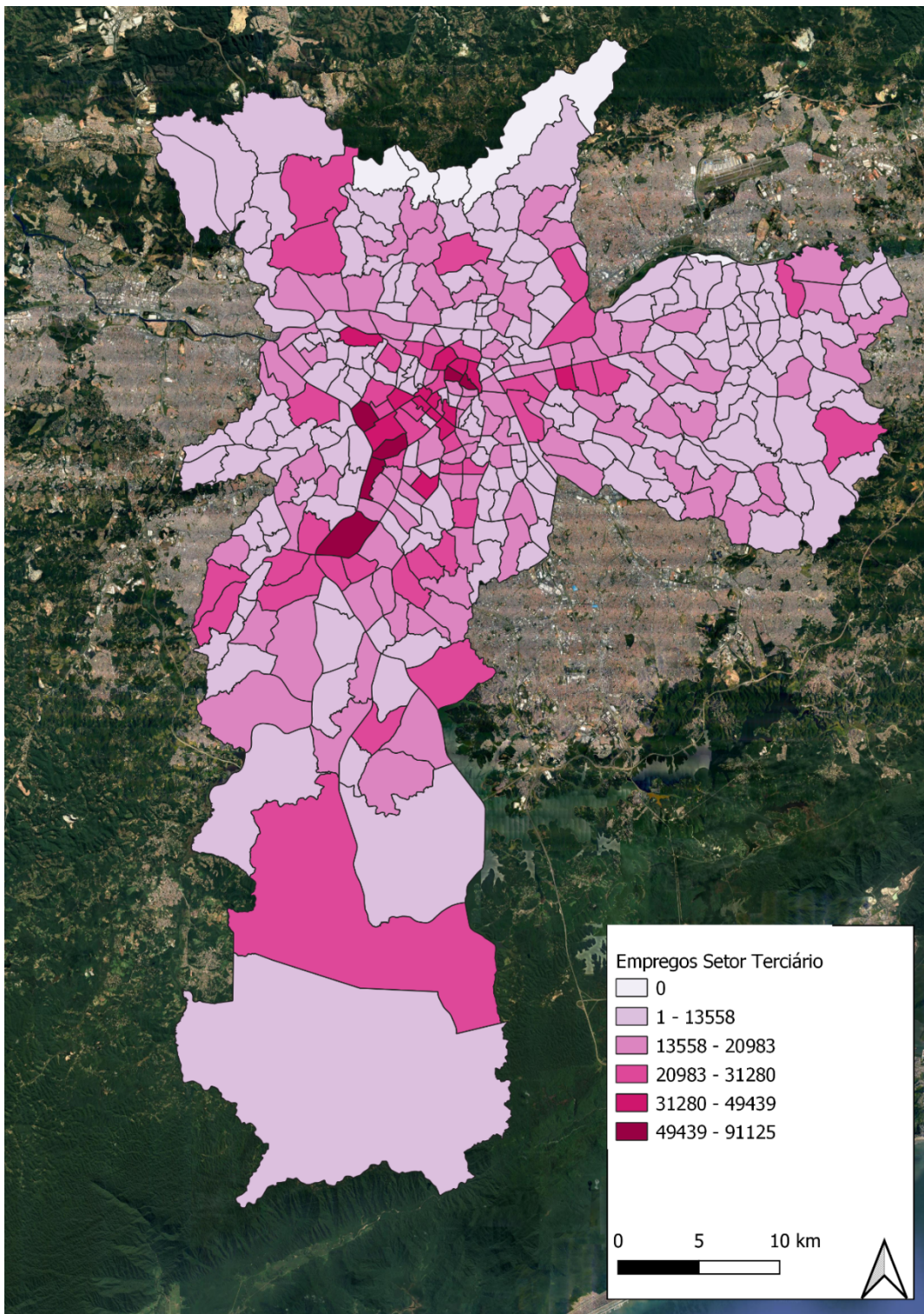


Figura 5 – Número de empregos do setor terciário nas zonas OD do município de São Paulo
 Fonte: Companhia do Metropolitano de São Paulo (2017); Prefeitura Municipal de São Paulo (2016).

Além das variáveis explicativas sociodemográficas de empregos, diretamente relacionadas às viagens à trabalho, buscou-se utilizar variáveis *dummies* que também caracterizassem a concentração espacial de empregos no município.

As duas primeiras variáveis buscaram representar a concentração de empregos na região central do município. A primeira variável foi indicativa de pertencimento da zona de tráfego à área central, com valor 1 para as zonas que pertencessem à área central de São Paulo e 0 para as zonas que não pertencessem à região. As zonas que se enquadram nesta categoria são aquelas numeradas de 1 a 108 da Pesquisa OD, conforme mostra a Figura 6 (COMPANHIA DO METROPOLITANO DE SÃO PAULO, 2017). A segunda variável *dummy* diz respeito às zonas pertencentes ao distrito Sé (zonas 1 – Sé, 2 – Parque Dom Pedro II e 3 – Praça João Mendes) conforme indicado na Figura 6, pertencentes ao distrito mais central do município de São Paulo e responsável por concentrar 142 mil empregos do município (2%).

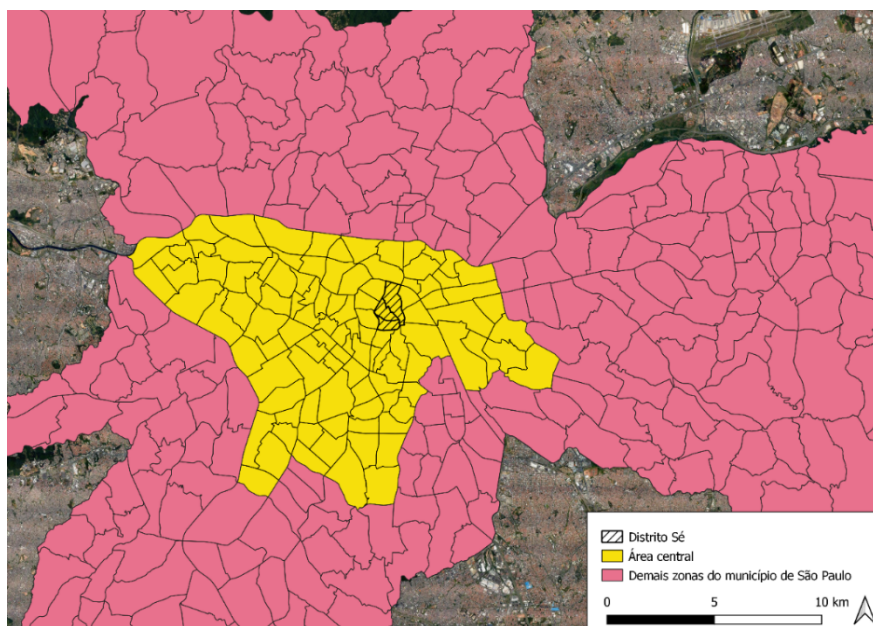


Figura 6 – Zonas pertencentes à área central e ao distrito Sé

Fonte: Companhia do Metropolitano de São Paulo (2017); Prefeitura Municipal de São Paulo (2016).

Para representar a concentração de empregos em São Paulo nas demais regiões do município foram utilizadas outras variáveis *dummies*. Uma delas utilizou como referência as zonas de centralidade (ZC) presentes no zoneamento do município de São Paulo, que refletem áreas das zonas com características centrais em diferentes regiões da cidade. Segundo a Lei nº 16.402, de 22 de março de 2016, Zonas de Centralidade são

“...porções do território destinadas à promoção de atividades típicas de áreas

centrais ou de subcentros regionais ou de bairros, em que se pretende promover majoritariamente os usos não residenciais, com densidades construtiva e demográfica médias e promover a qualificação paisagística e dos espaços públicos (SÃO PAULO, 2016).”

Para incluir a variável foi calculado o percentual de ZC em cada zona de tráfego. A partir disto, foi calculado o percentil desses valores e atribuído o valor 1 apenas para zonas com valor superior ao 85º quinto percentil, que consistiu em aproximadamente 23,5% de área com zona de centralidade conforme ilustrado na Figura 7.

Por fim, a última variável *dummy* considerada utilizou como insumo o uso predominante do solo do ponto de vista fiscal indicado no portal GEOSAMPA (PREFEITURA MUNICIPAL DE SÃO PAULO, 2016). O uso do solo predominante é definido pela predominância de 60% da quadra fiscal. As categorias consideradas foram os usos predominantes “Comércio e serviços”, “Indústria e armazéns” e “Comércio/serviços e Indústrias/armazéns”, de modo a identificar áreas com a maior concentração de empregos desses setores. Da mesma forma, para a variável de Zona de Centralidade foram identificadas predominâncias de áreas com essas características dentro das zonas da Pesquisa OD representadas pelo 85º percentil (Figura 8).

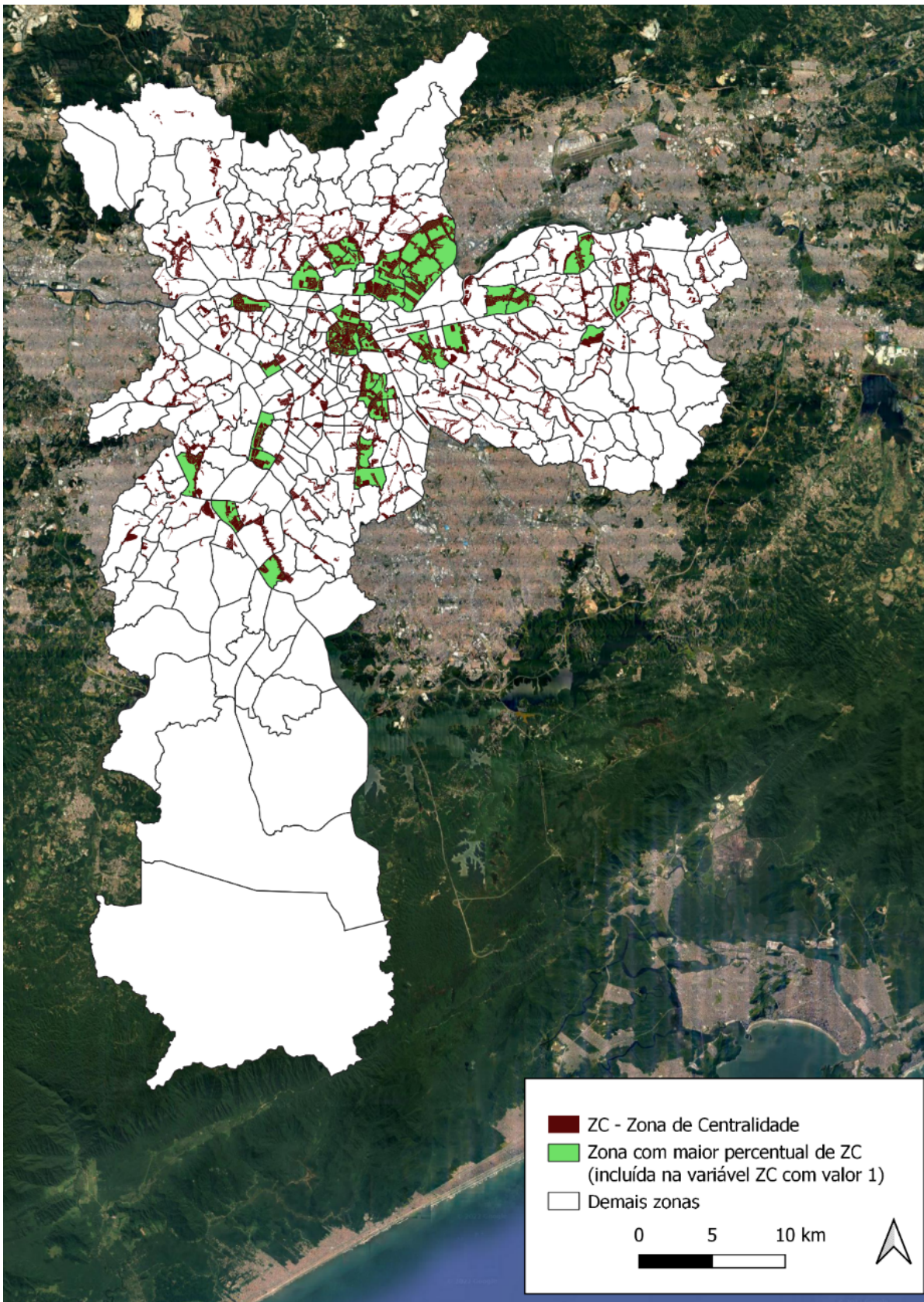


Figura 7 – Zonas com maior concentração de Zona de Centralidade
Fonte: Companhia do Metropolitano de São Paulo (2017); Prefeitura Municipal de São Paulo (2016).

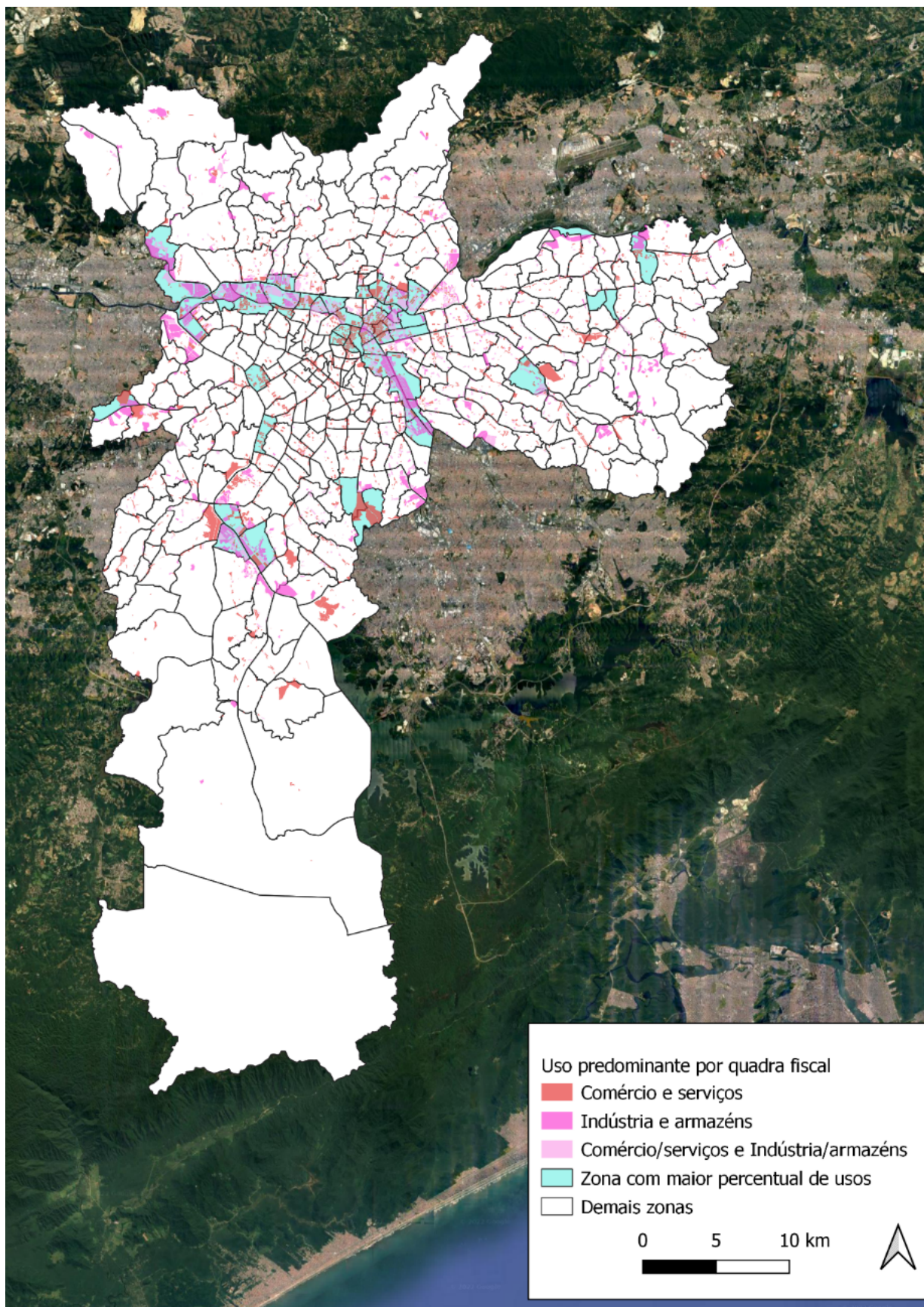


Figura 8 – Zonas com maior concentração de uso comércio, serviços e indústrias e armazéns
Fonte: Companhia do Metropolitano de São Paulo (2017); Prefeitura Municipal de São Paulo (2016).

A correlação entre todas as variáveis quantitativas consideradas foi calculada e os resultados são apresentados na Tabela 4. Verificou-se que a correlação entre as variáveis de tempo (TT) e custo (CO) e distância (DIS) é muito alta (próxima de 1,00) conforme mencionado anteriormente. Portanto, optou-se por utilizar nos modelos finais apenas a variável de distância (DIS). Entre as demais variáveis quantitativas a correlação é baixa, e portanto, foram mantidas nos modelos finais.

Também foi calculada a correlação entre as variáveis *dummies* e as demais variáveis quantitativas selecionadas, conforme Tabela 5. A correlação é do tipo ponto bisserial, utilizada quando uma das variáveis é dicotômica e a outra é métrica (KORNBROT, 2014). Observa-se que a correlação entre as variáveis *dummies* e demais variáveis é mais próxima de zero, com exceção da relação entre as variáveis CENT e DIS, caso isolado.

Tabela 4 – Correlação entre variáveis quantitativas

	EMPLN	EMPSS	EMPST	CO	TT	DIS
EMPLN	1,00	0,55	0,20	-0,23	-0,23	-0,23
EMPSS		1,00	0,07	0,13	0,12	0,13
EMPST			1,00	-0,21	-0,19	-0,21
CO				1,00	0,97	1,00
TT					1,00	0,97
DIS						1,00

Fonte: Companhia do Metropolitano de São Paulo (2017); Gomide e Morato (2011); Pereira *et al.* (2021).

Tabela 5 – Correlação entre variáveis dummies e variáveis selecionadas

	EMPLN	EMPSS	EMPST	DIS	CENT	SE	USO	ZC
EMPLN	1,00	0,55	0,20	-0,23	0,20	0,13	0,09	0,11
EMPSS		1,00	0,07	0,13	-0,15	0,00	0,12	-0,03
EMPST			1,00	-0,21	0,21	0,06	0,07	0,08
DIS				1,00	-0,62	-0,15	-0,26	-0,25
CENT					1,00	0,14	0,26	0,08
SE						1,00	0,14	0,25
USO							1,00	0,21
ZC								1,00

Fonte: Companhia do Metropolitano de São Paulo (2017)

Na tabela a seguir (Tabela 6) estão relacionadas as variáveis utilizadas nos modelos finais especificados no item 3.4.

Tabela 6 – Variáveis selecionadas

Variáveis	Descrição
EMPLN	Logaritmo de empregos totais na zona de destino
EMPSS	Logaritmo de empregos do setor secundário
EMPST	Logaritmo de empregos do setor terciário
DIS	Distância entre zonas
CENT	Variável <i>dummy</i> de centro expandido
SE	Variável <i>dummy</i> de zona Sé
USO	Variável <i>dummy</i> de uso do solo
ZC	Variável <i>dummy</i> de zona de centralidade

Fonte: Companhia do Metropolitano de São Paulo (2017)

3.2 DEFINIÇÃO DO NÚMERO ÓTIMO DE ALTERNATIVAS

Grandes conjuntos de alternativas para modelos de transporte resultam em expressivo esforço computacional para estimativa de modelos de escolha discreta dependendo da abordagem considerada para representação do comportamento individual. No caso dos modelos *RRM* a situação é agravada uma vez que cada atributo de cada alternativa é comparado com todas as outras alternativas.

Em comparação com modelos sob a abordagem *RUM*, o tempo computacional para execução dos modelos *RRM* cresce de exponencialmente conforme é possível verificar na Figura 7. Desta forma, Guevara, Chorus e Ben-Akiva (2016) propuseram um método para definir o número ótimo de alternativas disponíveis aos indivíduos no processo de modelagem em comparação à todas as alternativas J do conjunto de escolha em modelos *RRM*. A seleção do número ótimo de alternativas \tilde{J} disponíveis aos indivíduos depende da distribuição dos dados, do número de atributos, do número de observações N , do procedimento de otimização e de capacidade de processamento computacional.

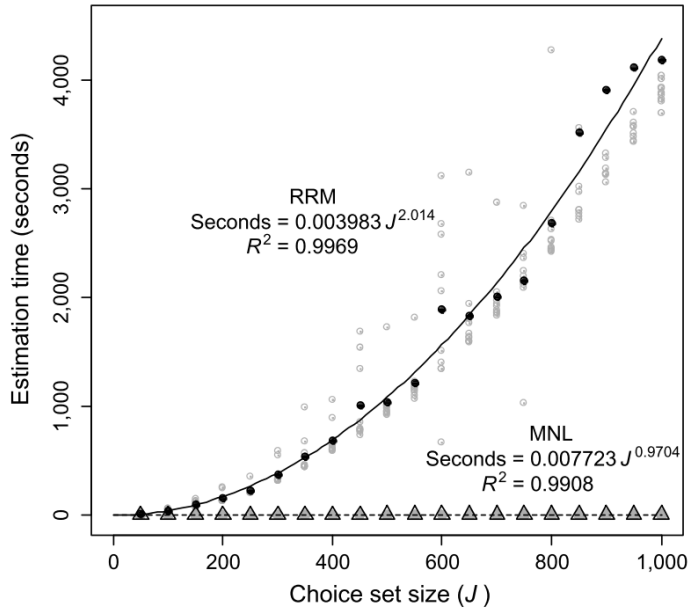


Figura 9 – Tempo de execução de modelos *RRM* comparado aos modelos *RUM* em função do número de alternativas (J)

Fonte: Guevara, Chorus e Ben-Akiva (2016)

Conforme apresentado anteriormente, a probabilidade de um indivíduo escolher a alternativa i nos modelos *RRM* é dada pela Equação 8.

Guevara, Chorus e Ben-Akiva (2016) supõem que, para a seleção das alternativas disponíveis para cada observador, primeiramente, o pesquisador amostra um subconjunto D_n com \tilde{J}_n elementos de um conjunto de escolha C_n . Este subconjunto D_n contém a alternativa escolhida i . Se i não é incluída em D_n , uma medida de probabilidade construída combinando i e os elementos em D_n pode não estar bem definida pois pode ser maior que 1,0 ou a probabilidade pode ser ilimitada, impedindo a estimativa do modelo.

Os autores definem a probabilidade conjunta $\pi(i, D_n)$ da Equação 17 como a probabilidade do indivíduo n escolher a alternativa i e do pesquisador selecionar o conjunto D_n . Pelo teorema de Bayes, a expressão $\pi(i | D_n)$ é a probabilidade condicional de escolher a alternativa i , dado que D_n foi selecionado e $\pi(D_n | i)$ é a probabilidade condicional do pesquisador selecionar D_n , dado que a alternativa i foi escolhida pelo indivíduo.

$$\pi(i, D_n) = \pi(D_n | i)P_n(i) = \pi(i | D_n)\pi(D_n) \quad (17)$$

Dado que os eventos de escolha das alternativas em C_n são mutuamente exclusivos, a probabilidade $\pi(D_n)$ de construir o conjunto D_n é:

$$\pi(D_n) = \sum_{j \in C_n} \pi(D_n | j)P_n(j) = \sum_{j \in D_n} \pi(D_n | j)P_n(j) \quad (18)$$

Ao substituir a Equação (18) e a probabilidade $P(i)$ da Equação (8) na Equação (17), obtém-se a Equação (19) para a probabilidade condicional do indivíduo escolher a alternativa i dado que o subconjunto D_n foi selecionado:

$$\pi(i | D_n) = \frac{e^{-R_{in} + \ln \pi(D_n | i)}}{\sum_{j \in D_n} e^{-R_{jn} + \ln \pi(D_n | j)}} \quad (19)$$

Na Equação (19), apesar do denominador depender apenas de D_n , a necessidade de estimar R_{in} depende do conjunto de escolha completo C_n . Assim, Guevara, Chorus e Ben-Akiva (2016) adaptaram o modelo substituindo R_{in} por um estimador que depende apenas do conjunto D_n . O arrependimento \tilde{R}_{in} com M atributos é aproximado pela Equação (20).

$$\tilde{R}_{in} = \sum_{j \in C_n} \sum_{m=1}^M \ln [1 + \exp(\beta_m \{x_{jmn} - x_{imn}\})] = R_{in} + M \ln (2) \quad (20)$$

Para aproximar \tilde{R}_{in} a uma forma viável, Guevara, Chorus e Ben-Akiva (2016) propuseram o seguinte modelo:

$$\hat{R}_{in} = \sum_{j \in \tilde{D}_n} w_{jn} \sum_{m=1}^M \ln [1 + \exp(\beta_m \{x_{jmn} - x_{imn}\})] \quad (21)$$

$$\text{com } w_{jn} = \frac{\tilde{n}_{jn}}{E(\tilde{n}_{jn})} \quad (22)$$

em que \tilde{n}_{jn} é número de vezes que a alternativa j é incluída na amostra para o indivíduo n e $E(\tilde{n}_{jn})$ é o seu valor esperado.

A estimativa de w_{jn} depende no protocolo de amostragem utilizado e se a amostra D_n utilizada para escrever a correção da amostragem $\ln \pi(D_n | j)$ na Equação (19) é a mesma que o subconjunto \tilde{D}_n usado para construir os fatores de expansão w_{jn} . Portanto, para conseguir aplicar o método proposto por Guevara, Chorus e Ben-Akiva (2016) deve-se determinar valores de w_{jn} e $\ln \pi(D_n | j)$. Os autores exploraram 3 métodos para solucionar o problema.

O primeiro método é quando a reamostragem é possível. O pesquisador tem total controle sobre os dados e consegue amostrar D_n de C_n para determinar $\ln \pi(D_n | i)$ e amostrar

\tilde{D}_n de C_n para obter w_{jn} , e então calcular \hat{R}_{in} . Nesse caso, Guevara, Chorus e Ben-Akiva (2016) definem w_{jn} por:

$$w_{jn} = \frac{1}{\tilde{J}/J} = \frac{J}{\tilde{J}} \quad (23)$$

Para determinação do valor de $\ln \pi (D_n | i)$ inclui-se a alternativa escolhida como disponível para cada indivíduo e as alternativas não escolhidas são amostradas aleatoriamente sem reposição para o total de alternativas \tilde{J} . Como o termo definido pela Equação (24) é constante entre alternativas, ele é cancelado no cálculo.

$$\ln \pi (D_n | i) = \ln \left(\frac{J - 1}{\tilde{J} - 1} \right) \quad (24)$$

Por fim, a probabilidade condicional de escolher a alternativa i , dados os conjuntos D_n e \tilde{D}_n é expressa pela Equação (25), que só depende do valor w_{jn} obtido pela Equação (23).

$$\hat{\pi}_n(i | D_n, \tilde{D}_n) = \frac{e^{-\sum_{j \in \tilde{D}_n} \left(\frac{J}{\tilde{J}}\right) \sum_{m=1}^M [1 + \exp(\beta_m \{x_{jmn} - x_{imn}\})]}}{\sum_{k \in D_n} e^{-\sum_{j \in \tilde{D}_n} \left(\frac{J}{\tilde{J}}\right) \sum_{m=1}^M [1 + \exp(\beta_m \{x_{jmn} - x_{imn}\})]}} \quad (25)$$

Quando a reamostragem não é possível, o protocolo é mais complexo e w_{jn} é definido pela seguinte fórmula, em que há dependência das próprias probabilidades de escolha, que não são conhecidas.

$$w_{jn} = \frac{1}{P_n(j) + ((\tilde{J} - 1)(J - 1))(1 - P_n(j))} \quad (26)$$

Assim, os autores propuseram dois métodos para solução quando a reamostragem não é possível, o método *Pop.Shares* e o método 1_0 . No método *Pop.Shares*, a estimativa das probabilidades de escolha é dada por um fator H_j conforme a Equação (27), definido como função de y_{jn} igual a 1 se o indivíduo n escolhe a alternativa j e 0 caso contrário.

$$H_j \approx \frac{\sum_n y_{jn}}{N} \quad (27)$$

Para o método *Pop.Shares*, a estimativa da probabilidade é dada por:

$$\hat{\pi}_n(i | D_n) = \frac{e^{-Q(i)}}{\sum_{k \in mD_n} e^{-Q(k)}} \quad (28)$$

$$Q(s) = \sum_{j \in D_n} \left[\frac{\sum_n y_{jn}}{N} + \frac{\tilde{J} - 1}{\tilde{J} - 1} \left(1 - \frac{\sum_n y_{jn}}{N} \right) \right] \times \sum_{m=1}^M [1 + \exp(\beta_m \{x_{jmn} - x_{smn}\})] \quad (29)$$

No método 1_0, as probabilidades recebem o valor 1 para a alternativa escolhida e zero para aquelas não escolhidas. Dessa forma, w_{jn} é igual a:

$$w_{in} = \frac{(J_n - 1)}{(\tilde{J}_n - 1)} \quad (30)$$

Para o método 1_0, a estimativa da probabilidade é dada de maneira análoga à Equação (29), em que:

$$Q_1(s) = \sum_{j \in D_n} \left[y_{jn} + (1 - y_{jn}) \frac{J_n - 1}{\tilde{J}_n - 1} \right] \times \sum_{m=1}^M [1 + \exp(\beta_m \{x_{jmn} - x_{smn}\})] \quad (31)$$

Os autores testaram esses métodos variando o valor de J igual a 5, 15, 30 e 50 e repetindo os modelos para cada valor pelo menos 100 vezes. Para cada repetição, anotaram o valor do parâmetro x , as escolhas e os conjuntos D_n e \tilde{D}_n , e calcularam a diferença entre o valor médio obtido e o verdadeiro valor do parâmetro, pois dispunham dessa informação já que executaram os modelos considerando $\tilde{J}=J$. Também calcularam o erro quadrático médio, o valor do teste t , e o número de vezes em que o estimador de cada repetição está em 75% do intervalo de confiança do valor real nas 100 repetições. Ao analisar os resultados, os autores concluíram que os melhores resultados ocorrem quando a reamostragem é possível e o *Pop.Shares* é utilizado.

Para a definição propriamente dita do número de alternativas ideal \tilde{J} , Guevara, Chorus e Ben-Akiva (2016) variaram os valores de \tilde{J} para 5, 15, 30 e 50. Para cada valor de \tilde{J} , amostraram K conjuntos $\tilde{D}_K(\tilde{J})$ e obtiveram valores dos parâmetros $\hat{\beta}_k$ para os quais foram estimados valores médios e desvio-padrão conforme a Equação (32) e Equação (33), respectivamente. Quanto maior \tilde{J} , menor $\hat{\sigma}_{\tilde{\beta}}$, tal que se $\tilde{J} = J$ então $\hat{\sigma}_{\tilde{\beta}}^2 = 0$.

$$\bar{\beta} = \frac{1}{K} \sum_{k=1}^K \hat{\beta}_k(\tilde{J}) \quad (32)$$

$$\hat{\sigma}_{\hat{\beta}} = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (\hat{\beta}_k - \bar{\hat{\beta}})^2} \quad (33)$$

Se o pesquisador não consegue obter $\hat{\beta}$ com o conjunto de escolha completo, então $\bar{\hat{\beta}}$ para \tilde{J} ideal é obtido através da estabilidade do parâmetro. Portanto, a obtenção do valor ideal de \tilde{J} depende da comparação dos valores de $\bar{\hat{\beta}}$ e $\hat{\sigma}_{\hat{\beta}}$ para diversos valores. A partir do momento em que é observada a estabilidade do parâmetro e os desvios se tornam menores, é possível definir o valor de \tilde{J} .

O procedimento para definição do número ótimo de alternativas (\tilde{J}) aplicado nesta dissertação consistiu na aplicação do método descrito, em conjunto com o método de Chow *et al.* (2005), Li *et al.* (2005) e Kim e Lee (2017) apresentado no item 3.3. Foram testados valores de conjuntos de alternativas disponíveis aos indivíduos (J) iguais a 5, 10, 15, 20, 25, 30, 35 e 40. O modelo aplicado para esse procedimento utilizou como variáveis o logaritmo de empregos (EMPLN) e o valor de distância (DIS). As variáveis foram utilizadas porque foram verificadas como significativas a 95% de confiança em testes preliminares e optou-se por utilizá-las pois incluir novas variáveis aumentaria o tempo de processamento computacional, refletido em horas para executar cada uma das iterações. As outras variáveis indicadas na Tabela 6 foram incluídas nos modelos finais após definido o valor de \tilde{J} .

3.3 REDUÇÃO DO CONJUNTO DE ESCOLHA DE DESTINOS

O método para definição do conjunto de alternativas disponíveis para escolha pelos indivíduos foi por amostragem estratificada descrito no item 2.3 e baseou-se nas pesquisas de Chow *et al.* (2005), Li *et al.* (2005) e Kim e Lee (2017), cujas etapas são:

- Para cada origem, todos os possíveis destinos foram ordenados em ordem crescente de proximidade geográfica (distância euclidiana dos respectivos centroides à zona de origem);
- O índice de Moran Global indicado na Equação 34 foi calculado para cada um dos destinos utilizando o pacote *spdep* implementado em R (Bivand *et al.*, 2013; Bivand e Wong, 2018).

$$I = \frac{n \times \sum_{i=1}^n \sum_{j=1}^n w_{ij} \times (y_i - \bar{y}) \times (y_j - \bar{y})}{(\sum_{i=1}^n (y_i - \bar{y})^2) \times \sum_{i=1}^n \sum_{j=1}^n w_{ij}} \quad (34)$$

em que w_{ij} : pesos da matriz de vizinhança calculados pelo pacote *spdep*; n : número de zonas em um estrato, no caso, as 338 zonas de tráfego resultantes da filtragem do banco de dados inicial; y_i : variável de interesse (número de empregos); \bar{y} média da variável de interesse.

- Para cada origem, os destinos foram ordenados pelos respectivos índices de Moran e classificados quanto ao seu sinal (positivo ou negativo);
- Foi definido um estrato para cada mudança de sinal do índice de Moran na sequência de ordenação;
- O tamanho da amostra de cada estrato foi definido pelo método de alocação de Neyman (KIM; LEE, 2017b; LI *et al.*, 2005; NEYMAN, 1934) conforme a Equação (35) e Equação (36).

$$n_h = \frac{n \times N_h \times \sigma_h}{\sum_{k=1}^L N_k \times \sigma_k} \quad (35)$$

$$\sum_{h=1}^L n_h = n \quad (36)$$

em que n_h : tamanho da amostra no estrato h ; n : tamanho total da amostra de alternativas disponíveis ao indivíduo; N_h : número de alternativas no estrato h ; σ_h : desvio padrão para o estrato h quanto à variável de interesse; e L : número de estratos.

3.4 ESPECIFICAÇÃO DOS MODELOS

Foram definidos 6 modelos *RUM* e *RRM* considerando a combinação de variáveis da Tabela 7. Os modelos 1 utilizaram as variáveis quantitativas de logaritmo de empregos totais (EMPLN) e de distância (DIS), os modelos 2 utilizaram as variáveis quantitativas de logaritmo de empregos do setor secundário (EMPSS) e terciário (EMPST) e a variável de distância (DIS). Os modelos de 3 a 6 foram estabelecidos com a variável DIS e as variáveis *dummies*, respectivamente: modelo 3, variável CENT; modelo 4, variável SE; modelo 5, variável USO; e modelo 6, variável ZC. Os modelos que incluíram a variável EMPLN, a variável DIS, e as variáveis *dummies* simultaneamente não foram apresentados pois as variáveis *dummies* não foram significativas e apresentaram sinal oposto ao esperado possivelmente devido à relação

entre elas mesmo não tendo ficado explícitas na análise de correlação.

Tabela 7 – Modelos de escolha de destinos e variáveis utilizadas

Modelos <i>RUM/RRM</i>	Variáveis utilizadas							
	EMPLN	EMPSS	EMPST	DIS	CENT	SE	USO	ZC
1	X			X				
2		X	X	X				
3				X	X			
4				X		X		
5				X			X	
6				X				X

Fonte: Companhia do Metropolitano de São Paulo (2017)

Nos modelos *RUM* as funções utilidade foram especificadas conforme as Equações (37) a (42) enquanto as funções de arrependimento dos modelos *RRM* foram estabelecidas conforme a Equações (43) a (48) equivalentes ao modelo proposto por Chorus (2010).

$$U_{iq}^1 = \beta_{empregos_total} \cdot EMPLN_i + \beta_{distância} \cdot DIS_i \quad (37)$$

$$U_{iq}^2 = \beta_{empregos_setor_secundário} \cdot EMPSS_i + \beta_{empregos_setor_terciário} \cdot EMPST_i + \beta_{distância} \cdot DIS_i \quad (38)$$

$$U_{iq}^3 = \beta_{distância} \cdot DIS_i + \beta_{centro_expandido} \cdot CENT_i \quad (39)$$

$$U_{iq}^4 = \beta_{distância} \cdot DIS_i + \beta_{distrito_sé} \cdot SE_i \quad (40)$$

$$U_{iq}^5 = \beta_{distância} \cdot DIS_i + \beta_{uso_predominante_comércio_serviços} \cdot USO_i \quad (41)$$

$$U_{iq}^6 = \beta_{distância} \cdot DIS_i + \beta_{zona_centralidade} \cdot ZC_i \quad (42)$$

$$R_{iq}^1 = \ln(1 + \exp[\beta_{empregos_total} \cdot (EMPLN_i - EMPLN_j)]) + \ln(1 + \exp[\beta_{distância} \cdot (DIS_i - DIS_j)]) \quad (43)$$

$$R_{iq}^2 = \ln(1 + \exp[\beta_{empregos_setor_secundário} \cdot (EMPSS_i - EMPSS_j)]) + \ln(1 + \exp[\beta_{empregos_setor_terciário} \cdot (EMPST_i - EMPST_j)]) + \ln(1 + \exp[\beta_{distância} \cdot (DIS_i - DIS_j)]) \quad (44)$$

$$R_{iq}^3 = \ln(1 + \exp[\beta_{distância} \cdot (DIS_i - DIS_j)]) + \ln(1 + \exp[\beta_{centro_expandido} \cdot (CENT_i - CENT_j)]) \quad (45)$$

$$R_{iq}^4 = \ln(1 + \exp[\beta_{distância} \cdot (DIS_i - DIS_j)]) + \ln(1 + \exp[\beta_{distrito_sé} \cdot (SE_i - SE_j)]) \quad (46)$$

$$R_{iq}^5 = \ln(1 + \exp[\beta_{dist\grave{a}ncia} \cdot (DIS_i - DIS_j)]) + \ln(1 + \exp[\beta_{uso_predominante_com\grave{e}rcio_servi\c{c}os} \cdot (USO_i - USO_j)]) \quad (47)$$

$$R_{iq}^6 = \ln(1 + \exp[\beta_{dist\grave{a}ncia} \cdot (DIS_i - DIS_j)]) + \ln(1 + \exp[\beta_{zona_centralidade} \cdot (ZC - ZC_j)]) \quad (48)$$

em que $\beta_{empregos_total}$, $\beta_{empregos_setor_secund\grave{a}rio}$, $\beta_{empregos_setor_terci\grave{a}rio}$, $\beta_{dist\grave{a}ncia}$, $\beta_{centro_expandido}$, $\beta_{distrito_S\acute{e}}$, $\beta_{uso_predominante_com\grave{e}rcio_servi\c{c}os}$, $\beta_{zona_centralidade}$: par\^ametros a serem estimados;

$EMPLNi$, $EMPLNj$: logaritmo do n\^umero de empregos dos destinos i e j , respectivamente;

$EMPSSi$, $EMPSSj$: logaritmo do n\^umero de empregos do setor secund\^ario dos destinos i e j , respectivamente;

$EMPSTi$, $EMPSTj$: logaritmo do n\^umero de empregos do setor terci\^ario dos destinos i e j , respectivamente;

DIS_i , DIS_j : dist\^ancia de viagem da origem para os destinos i e j , respectivamente;

$CENT_i$, $CENT_j$: *dummy* indicativa de centro expandido nos destinos i e j , respectivamente;

SE_i , SE_j : *dummy* indicativa de distrito S\acute{e} nos destinos i e j , respectivamente;

USO_i , USO_j : *dummy* indicativa de uso predominante de com\^ercio e servi\c{c}os nos destinos, respectivamente; e

ZC_i , ZC_j : *dummy* indicativa de zona de centralidade nos destinos, respectivamente.

Assume-se que os termos aleat\^orios seguem distribui\c{c}\~ao Valor Extremo do Tipo I resultando em probabilidades de escolha dos destinos pelo modelo Logit conforme a Equa\c{c}\~ao (2) e a Equa\c{c}\~ao (8) para os modelos *RUM* e *RRM*, respectivamente. Todos os modelos foram estimados pelo pacote Apollo (Hess e Palma, 2019; Hess e Palma, 2021) no software R em um computador Intel(R) Core (TM) i7-10750H e 32GB RAM.

3.5 VALIDAÇÃO

Parady, Ory e Walker (2021) definem validação como a avaliação da generalização de um modelo estatístico, ou seja, como a capacidade de um modelo de manter sua precisão preditiva em uma amostra diferente. Assim, para avaliação do desempenho dos modelos, além de medidas de ajuste recomenda-se fortemente a utilização de métodos de validação. A falta de validação de modelos aumenta o risco de *overfitting*, quando o modelo se ajusta bem aos dados da estimativa mas tem desempenho ruim em outro conjunto de dados ao ser replicado.

O processo para validação inicia-se com a estimativa do modelo na amostra para estimação, a aplicação do modelo estimado em uma amostra para validação e, por fim, avaliação de sua generalização utilizando medidas de desempenho para análise e acurácia como uma função da discrepância entre valores observados e previstos.

Os métodos de validação mais comuns na literatura são *Holdout*, *Cross-Validation* e *Bootstrapping*. A validação *Holdout* (HOV) é o método mais simples em que a base de dados é separada aleatoriamente em uma base para estimação e outra para validação. A validação cruzada (*Cross-Validation*) consiste no método *Holdout* repetido múltiplas vezes, gerando conjunto de pares de estimativa de forma aleatória. Por fim, os métodos de inicialização para validação (*Bootstrapping*) utilizam um conjunto de dados de tamanho previamente definido por amostragem aleatória (com substituição) em relação ao conjunto de dados original, sendo realizada a estimativa do modelo nesta amostra (PARADY; ORY; WALKER, 2021). Neste estudo foi utilizada a validação *holdout* devido ao esforço computacional exigido para estimativa dos parâmetros nos modelos.

Como medida de desempenho da validação, além do percentual de acertos, podem ser usadas a soma do erro quadrático (SSE), erro quadrático médio (MSE), erro quadrático da soma da raiz (RSSE), erro quadrático médio (RMSE), erro absoluto médio (MAE), erro percentual absoluto (APE) e erro percentual médio absoluto (MAPE). Não há consenso sobre a mais adequada (PARADY; ORY; WALKER, 2021).

Nesta dissertação utilizou-se 75% da base de dados para estimação dos modelos e 25% para validação frequentemente utilizados na literatura (PARADY, ORY E WALKER, 2021). Como medida de desempenho foi adotado o percentual de acertos entre estimativas de escolhas de destinos e aquelas realizadas, além da log-verossimilhança final, ρ^2 , ρ^2 ajustado, *Akaike Information Criterion* (AIC), *Bayesian Information Criterion* (BIC) e tempo computacional.

4 RESULTADOS

4.1 NÚMERO ÓTIMO DE ALTERNATIVAS E CONJUNTOS DE ESCOLHA DE DESTINOS

O procedimento para definição do número ótimo de alternativas (\tilde{J}) consistiu na aplicação do método de Guevara, Chorus e Ben-Akiva (2016) apresentado no item 3.2, em conjunto com o método de Chow *et al.* (2005), Li *et al.* (2005) e Kim e Lee (2017), apresentado no item 3.3. Nos gráficos apresentados entre a Figura 10 e a Figura 13 são ilustrados os resultados obtidos pela aplicação do método de Guevara, Chorus e Ben-Akiva (2016) pela variação do número de alternativas disponíveis ao indivíduo para determinação do número ótimo \tilde{J} .

Para cada valor de \tilde{J} foram realizadas 30 iterações com a obtenção de valores de β_{EMPLN} e β_{DIS} , cujos resultados são apresentados no Apêndice A. Ao final das 30 iterações, para cada J , foram estimados valores médios de $\bar{\beta}_{EMPLN}$ e $\bar{\beta}_{DIS}$ bem como os valores de desvio padrão, σ_{EMPLN} e σ_{DIS} .

Nota-se que para a variável empregos (EMPLN) há tendência de estabilização do valor da variável $\bar{\beta}_{EMPLN}$ a partir de $J = 20$. Para $J = 25$ há um pequeno decréscimo e posterior aumento em $J = 30$, com valor obtido para $J = 40$ muito próximo a $J = 20$. Além disso, o desvio padrão σ_{EMPLN} para $J = 20$ também é o menor observado, juntamente com o obtido para $J = 40$.

Para a variável distância (DIS), a tendência não é tão clara. Apenas é possível verificar que à medida que J aumenta, há uma tendência que $\bar{\beta}_{DIS}$ decresça, porém de maneira menos acentuada. Entretanto, ao observar o desvio padrão σ_{DIS} , nota-se que o menor valor é obtido para $J = 15$.

Após analisar os resultados foi escolhido o valor de $\tilde{J} = 20$, reforçado pela tendência clara observada para a variável empregos, e o desvio pequeno para variável distância porém estável a partir de $J = 15$. Portanto, os modelos finais consideraram o número de alternativas disponível para cada indivíduo igual a 20.

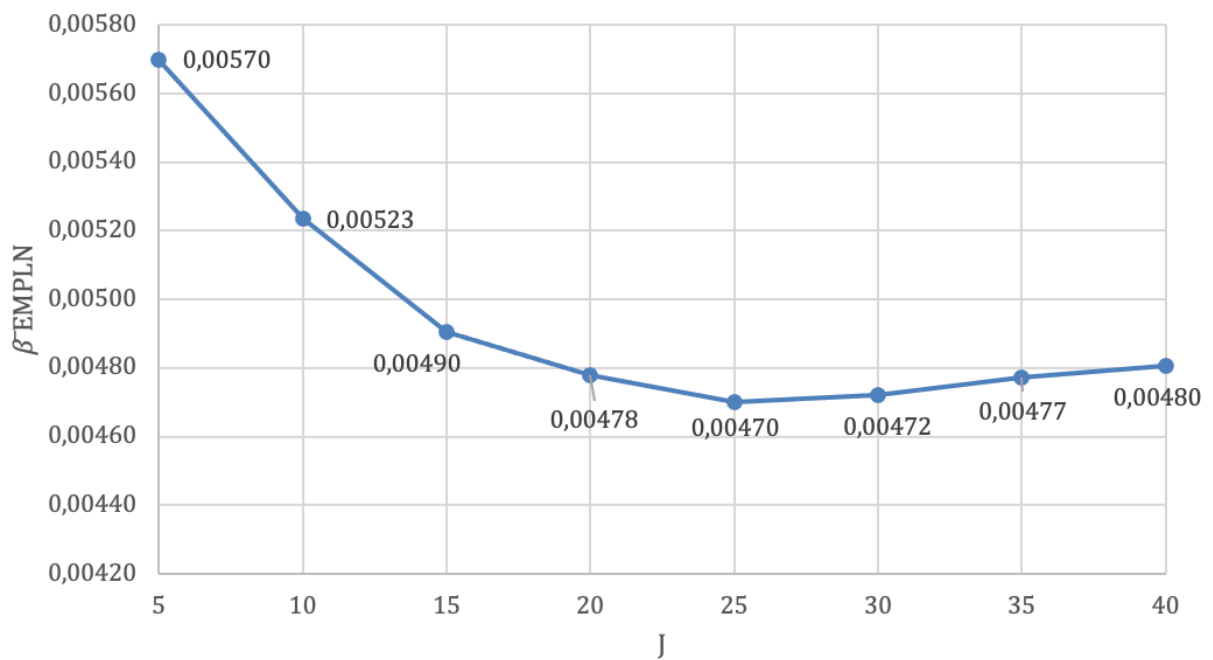


Figura 10 – Valores de $\bar{\beta}_{EMPLN}$ obtidos para cada J
 Fonte: Elaboração própria (2022).

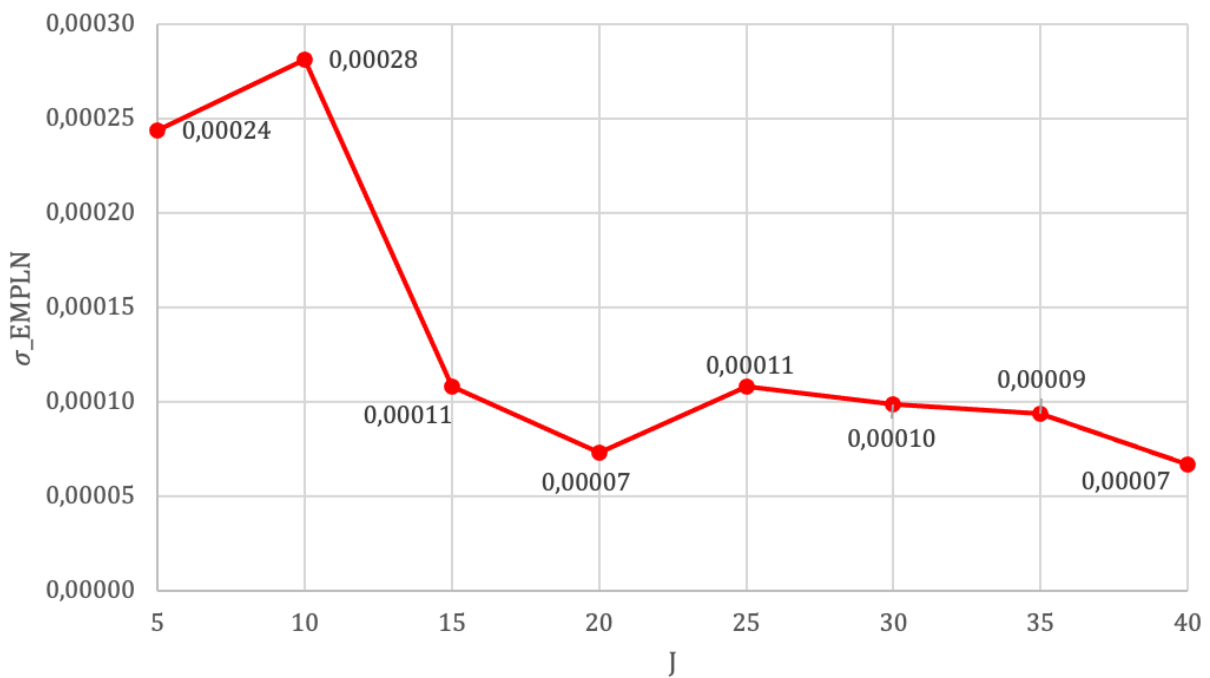


Figura 11 – Valores de σ_{EMPLN} obtidos para cada J
 Fonte: Elaboração própria (2022).

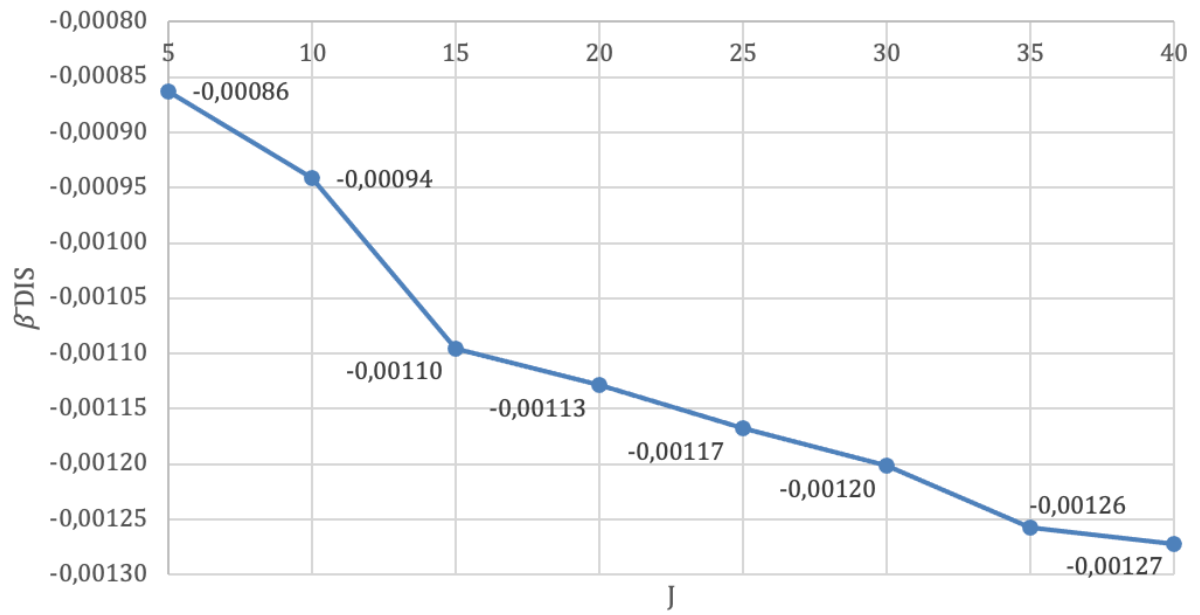


Figura 12 – Valores de $\bar{\beta}_{DIS}$ obtidos para cada J
 Fonte: Elaboração própria (2022).

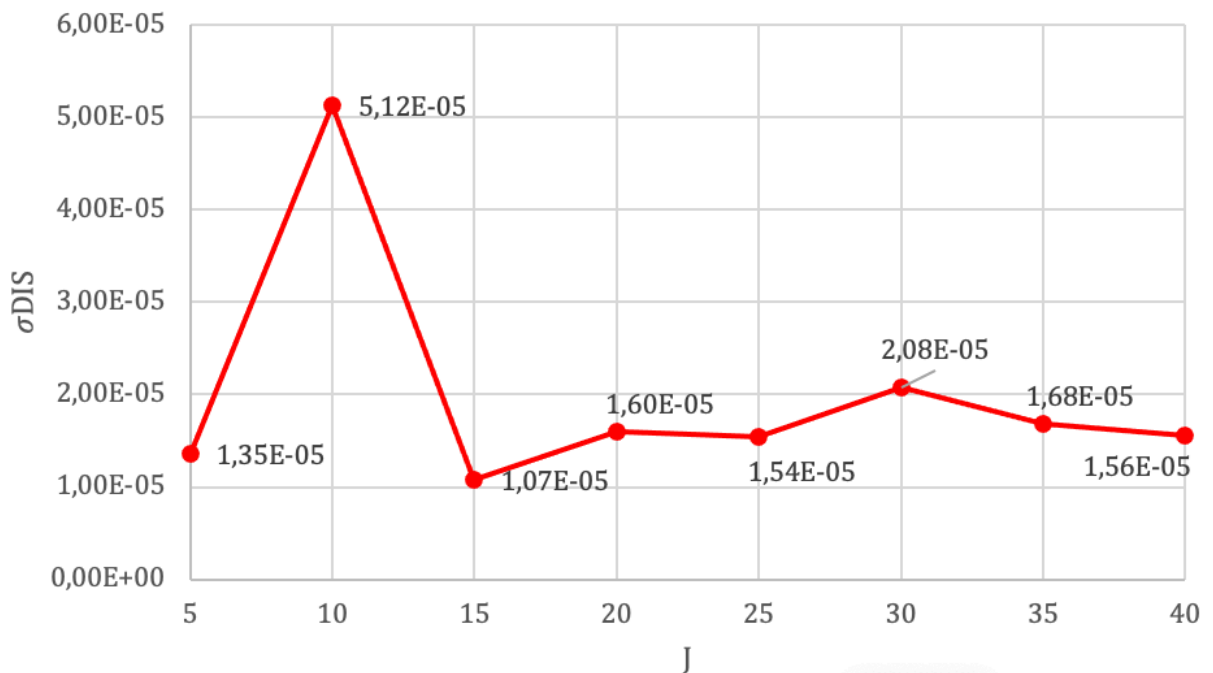


Figura 13 – Valores de σ_{DIS} obtidos para cada J
 Fonte: Elaboração própria (2022).

Após definir o número ótimo de alternativas disponíveis a cada indivíduo, foram determinadas aquelas que efetivamente compõem o conjunto de escolha de acordo com a zona de origem da viagem aplicando-se o método descrito em 3.3, incluindo a zona de destino efetivamente escolhida pelo indivíduo. Portanto, para cada zona de origem, foram gerados conjuntos de alternativas distintos com 20 alternativas disponíveis para escolha, incluindo a efetivamente escolhida. Esta base de dados foi utilizada em todos os modelos *RUM* e *RRM*.

4.2 MODELOS DE ESCOLHA DE DESTINOS

4.2.1 Medidas de desempenho

Os resultados das estimativas dos modelos em termos de medidas de desempenho são apresentados na Tabela 8. A tabela demonstra os valores de log-verossimilhança final, ρ^2 , ρ^2 ajustado, *Akaike Information Criterion (AIC)*, *Bayesian Information Criterion (BIC)*, tempo computacional e resultados de validação expressos em percentual de acertos.

Tabela 8 – Medidas de desempenho dos modelos *RUM* e *RRM*

Modelo	Abordagem	LL0	LLfinal	ρ^2	ρ^2 ajustado	AIC	BIC	Tempo computacional	Validação (% acertos)
1	<i>RRM</i>		-7281,045	0,2142	0,2138	14570,09	14594,24	01:49:00	25,90
	<i>RUM</i>		-7270,291	0,2154	0,2151	14544,58	14556,66	00:00:29	26,09
2	<i>RRM</i>		-7287,515	0,2135	0,2132	14581,03	14599,14	01:20:24	25,90
	<i>RUM</i>		-7272,344	0,2151	0,2148	14550,69	14568,8	00:00:30	25,90
3	<i>RRM</i>		-7713,106	0,1676	0,1674	15430,21	15442,29	00:33:54	20,76
	<i>RUM</i>		-7700,958	0,1689	0,1687	15405,92	15417,99	00:00:27	20,66
4	<i>RRM</i>	-9265,8	-7743,322	0,1643	0,1641	15490,64	15502,72	00:25:19	20,95
	<i>RUM</i>		-7735,201	0,1652	0,165	15474,40	15486,48	00:00:28	20,95
5	<i>RRM</i>		-7732,099	0,1655	0,1653	15468,20	15480,27	00:26:16	20,76
	<i>RUM</i>		-7723,349	0,1665	0,1663	15450,70	15462,77	00:00:40	20,76
6	<i>RRM</i>		-7730,711	0,1657	0,1655	15465,42	15477,5	00:35:01	20,85
	<i>RUM</i>		-7721,940	0,1666	0,1664	15447,88	15459,95	00:00:39	20,76

Fonte: Elaboração própria (2022).

Na comparação de desempenho entre os modelos *RUM* e *RRM* verificou-se que os primeiros apresentaram melhor desempenho de log-verossimilhança em relação ao *RRM* para todos os modelos conforme ilustrado na Figura 14. Apesar de manter as diferenças percentuais de ordem de grandeza inferior a 1%, a log-verossimilhança final da abordagem *RUM* foi 0,15% superior ao da *RRM* nos modelos 1; 0,21% nos modelos 2; 0,16 % nos modelos 3; 0,10% nos modelos 4 e de 0,11% nos modelos 5 e 6. O melhor desempenho em termos de log-verossimilhança final foi do modelo 1, que relaciona as variáveis EMPLN e DIS, na abordagem

RUM, com valor de -7270,291. Por outro lado, o pior desempenho foi do modelo 4 pela abordagem *RRM*, que relaciona as variáveis DIS e SE, com valor de -7743,322.

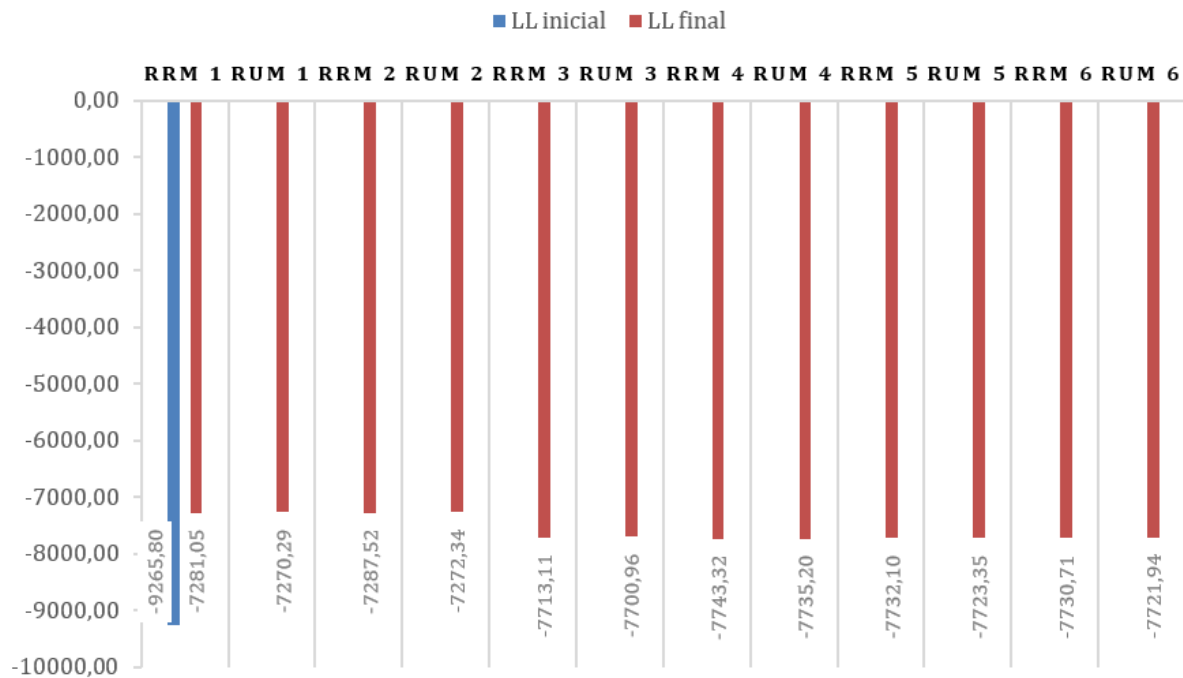


Figura 14 – Valores de log-verossimilhança iniciais e finais dos modelos

Fonte: Elaboração própria (2022).

Analisando a aderência dos modelos às respostas observadas, traduzida pela medida ρ^2 , novamente os valores ficam muito próximos. Seguindo o padrão de comparação de valores de log-verossimilhança, o valor de ρ^2 é maior para todos os modelos na abordagem *RUM* (0,56% maior para o modelo 1; 0,75% nos modelos 2; 0,78% nos modelos 3; 0,55% nos modelos 4; 0,60% nos modelos 5 e 0,54% nos modelos 6).

O modelo *RUM* 1, que tem como variáveis o logaritmo de empregos totais (EMPLN), e distância (DIS), apresentou o melhor resultado de ρ^2 com valor de 0,2154. Por outro lado, o modelo *RRM* 4, que utiliza as variáveis de distância e a *dummy* SE, obteve pior desempenho com o valor de 0,1643. Os valores de aderência ajustados são muito próximos e apresentam mesma tendência.

Da mesma forma, o *Akaike Information Criterion* (*AIC*) indica melhor desempenho para o modelo *RUM* 1, que relaciona EMPLN e DIS, com valor de 14.544,58, e pior desempenho do modelo *RRM* 4, com valor de 15.490,64. Assim como o *Bayesian Information Criterion*

(*BIC*) indica melhor desempenho do modelo *RUM* 1, com valor de 14.544,58 e pior desempenho do modelo *RRM* 4 (*BIC* = 15.502,72). As diferenças percentuais dos valores entre as abordagens mantêm-se abaixo de 1%. O desempenho dos modelos *RRM* para *AIC* e *BIC* sob abordagem *RUM* são melhores em todos os casos.

Além disso, observa-se que foi necessário maior esforço computacional para convergência dos modelos *RRM* dados os tempos requeridos para estimativa dos parâmetros. Nota-se que há expressiva diferença no tempo de estimação entre as duas abordagens (Figura 15). Os modelos *RUM* foram estimados em menos de um minuto (32 segundos em média) enquanto aqueles da abordagem *RRM* apresentaram tempos de estimação elevados (52 minutos, em média). O maior tempo de estimação de modelos *RRM* se dá principalmente pela não linearidade da função de arrependimento conforme discutido por Guevara, Chorus e Ben-Akiva (2016), que é uma limitação da abordagem.

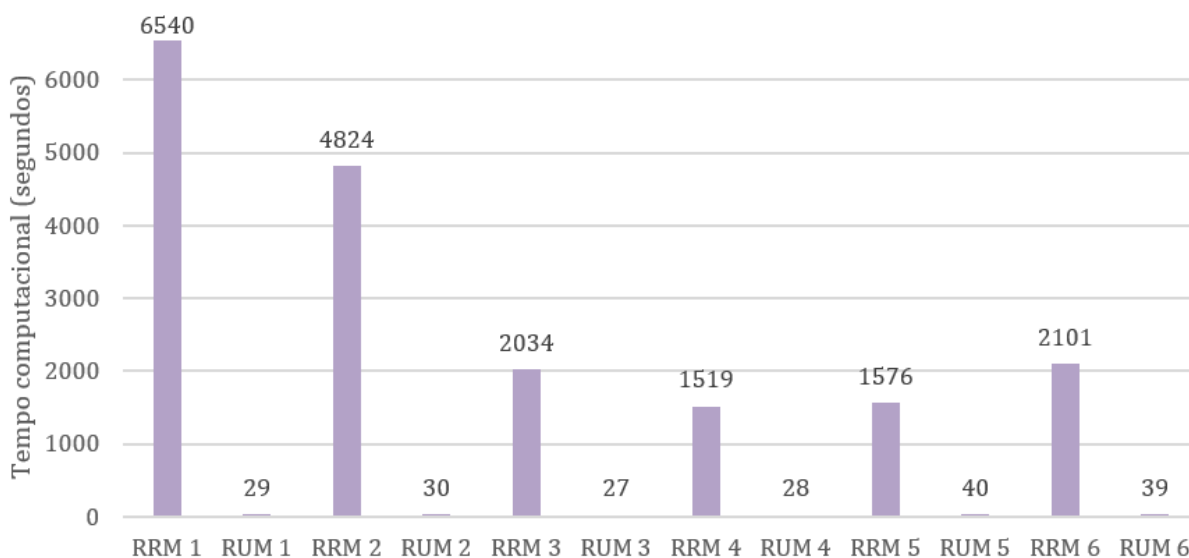


Figura 15 – Tempos de processamento computacional para estimativa dos modelos
 Fonte: Elaboração própria (2022).

Em termos de validação pelo método de *holdout*, os modelos *RUM* e *RRM* têm desempenho semelhante em porcentagem de acertos (Figura 16). Todos os modelos têm percentual de acerto superior a 20%, sendo que os modelos 1 e 2 que incluem variáveis quantitativas de empregos apresentam percentual de acertos maior que 25,90%. O percentual de acertos foi relativamente baixo possivelmente devido à variabilidade e grande quantidade de alternativas disponíveis para escolha.

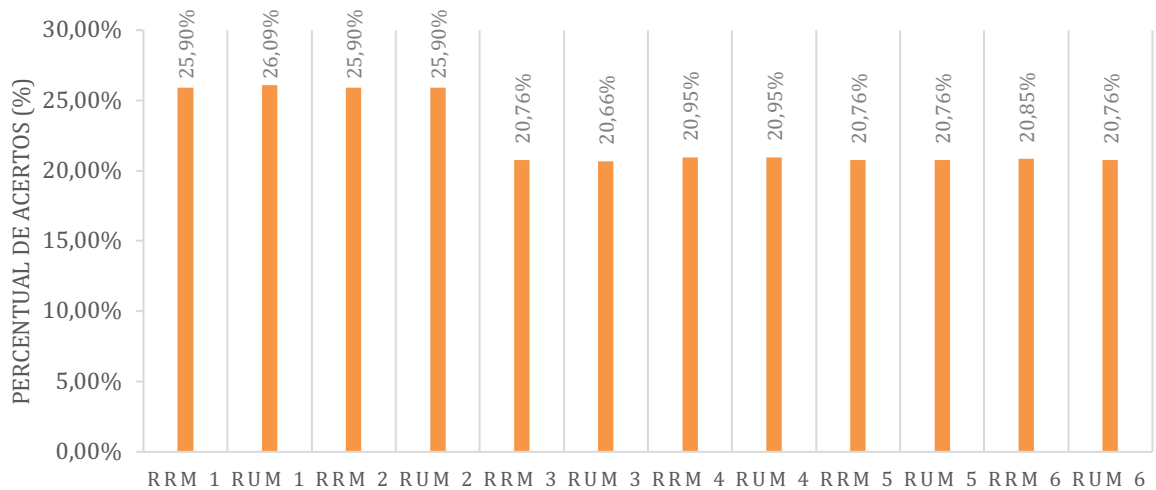


Figura 16 – Percentual de acertos dos modelos

Fonte: Elaboração própria (2022).

Portanto, apesar das diferenças percentuais pequenas em valores de desempenho entre as duas abordagens, verifica-se que a abordagem *RUM* explica mais adequadamente o fenômeno de escolha de destinos no contexto de viagens urbanas a trabalho a partir dos dados da pesquisa de preferência revelada na cidade de São Paulo, indicando que indivíduos tendem a maximizar a utilidade relativa aos destinos ao invés de tentar minimizar o arrependimento nas escolhas. O modelo com melhor desempenho em termos de log-verossimilhança foi o modelo *RUM 1*, que utilizou as variáveis *EMPLN* e *DIS*. Em geral, os modelos que relacionaram distância e variáveis *dummies* tiveram pior desempenho, demonstrando que as variáveis quantitativas de empregos são importantes na estimação.

4.2.2 Parâmetros

Os resultados de parâmetros dos modelos são apresentados Tabela 9. A tabela apresenta os valores dos parâmetros obtidos em cada modelo e os respectivos teste *t*, p-value e erro padrão obtidos, a 95% de confiança.

Tabela 9 – Resultados dos modelos *RUM* e *RRM*

Modelo	Abordagem	EMPLN*	EMPSS*	EMPST*	DIS*	CENT*	SE*	USO*	ZC*
1	<i>RRM</i>	0,0835 (28,948 / 0 / 0,0029)**			-0,0197 (-39,8343 / 0 / 0,0005)**				
	<i>RUM</i>	0,8252 (29,9694 / 0 / 0,0275)**			-0,1913 (-42,5322 / 0 / 0,0045)**				
2	<i>RRM</i>		0,0085 (4,8414 / 0 / 0,0018)**	0,0725 (23,7643 / 0 / 0,0031)**	-0,0197 (-39,8618 / 0 / 0,0005)**				
	<i>RUM</i>		0,0816 (4,6038 / 0 / 0,0177)**	0,723 (24,5792 / 0 / 0,0294)**	-0,1911 (-42,538 / 0 / 0,0045)**				
3	<i>RRM</i>				-0,0199 (-41,1122 / 0 / 0,0005)**	0,0328 (7,9468 / 0 / 0,0041)**			
	<i>RUM</i>				-0,1921 (-43,8659 / 0 / 0,0044)**	0,3499 (8,4684 / 0 / 0,0413)**			
4	<i>RRM</i>				-0,0201 (-42,1726 / 0 / 0,0005)**		0,0255 (1,9489 / 0,0513 / 0,0131)**		
	<i>RUM</i>				-0,1935 (-44,9756 / 0 / 0,0043)**		0,2698 (2,0823 / 0,0373 / 0,1296)**		
5	<i>RRM</i>				-0,0202 (-42,1073 / 0 / 0,0005)**			0,0243 (5,1765 / 0 / 0,0047)**	
	<i>RUM</i>				-0,1942 (-44,9265 / 0 / 0,0043)**			0,2505 (5,3694 / 0 / 0,0467)**	
6	<i>RRM</i>				-0,0201 (-42,0861 / 0 / 0,0005)**				0,0270 (5,4531 / 0 / 0,0049)**
	<i>RUM</i>				-0,1939 (-44,8978 / 0 / 0,0043)**				0,2776 (5,6561 / 0 / 0,0491)**

*(teste t / p-value / erro padrão) ; **significativo a 95% de confiança. Fonte: Elaboração própria (2022).

Nota-se que nos modelos *RUM*, o parâmetro relativo ao logaritmo de empregos apresentou sinal positivo (*RUM 1*), assim com o parâmetro relativo ao logaritmo de empregos por tipo no modelo *RUM 2* indicando que quanto maior o número de empregos, maior a probabilidade de escolha da zona de destino. Além disso os parâmetros são significativos a 95% de confiança para todos os modelos. O parâmetro de distância apresentou sinal negativo em todos os modelos, indicando que quanto maior a distância da viagem entre a origem e o destino, menor a probabilidade do destino ser escolhido. Os parâmetros relativos às variáveis mencionadas também são significativos a 95% de confiança para todos os modelos.

No que se refere às variáveis *dummies* testadas, todas apresentam sinal positivo (*CENT*, *SE*, *USO* e *ZC*) indicando que a localização espacial das oportunidades de emprego também influencia positivamente no aumento da probabilidade da escolha da zona. Além disso, todas as variáveis são significativas a 95% de confiança. Observa-se no entanto que os modelos 3 ao 6, que relacionam *dummies* e a variável *DIS*, conforme mencionado anteriormente, apresentam medidas de desempenho (log-verossimilhança final, aderência, *AIC*, *BIC*) inferiores aos demais modelos.

No caso dos parâmetros dos modelos sob a abordagem *RRM*, a interpretação é diferente daqueles sob a abordagem *RUM*. Chorus (2012b) destaca que sob a abordagem *RRM* o parâmetro estimado refere-se ao potencial acréscimo/decrécimo no arrependimento causado pela comparação da alternativa considerada em relação às demais alternativas devido ao acréscimo/decrécimo no nível do atributo. Por outro lado, sob a abordagem *RUM*, o parâmetro indica um acréscimo/decrécimo na utilidade de uma alternativa causado pelo acréscimo/decrécimo em uma unidade no valor do atributo.

Analisando-se os parâmetros do modelo *RRM* observa-se que têm comportamento semelhante ao observado nos modelos *RUM*. Nos modelos *RRM*, os parâmetros das variáveis de logaritmo de empregos (*EMPLN*) e empregos por tipo (*EMPSS* e *EMPST*), assim como a variável de distância (*DIS*), são significativos a 95% de confiança. Analogamente aos modelos *RUM*, os parâmetros das variáveis de empregos têm sinal positivo, enquanto aqueles de distância tem sinal negativo, indicando que a variável de empregos contribui para influenciar positivamente no aumento da função arrependimento relativa ao destino, assim como a variável distância influencia de forma negativa.

Nos modelos *RRM*, assim como nos modelos *RUM*, as variáveis *dummies* apresentam

sinal positivo e são significativas a 95% de confiança, conforme esperado, indicando que a localização espacial dos empregos influencia positivamente na escolha do destino.

5 CONCLUSÕES

Apesar de serem extensivamente utilizados para modelagem de distribuição de viagens, os modelos sintéticos, e particularmente o gravitacional, apresentam limitações quanto à inclusão de variáveis que potencialmente explicam o fenômeno de escolha de destinos, além de resultarem em estimativas agregadas da quantidade de viagens entre origens e destinos e não incluírem aspectos comportamentais nas análises. Nesse sentido, os modelos desagregados têm se mostrado eficazes para a modelagem de tal fenômeno, dentre os quais se destacam os modelos de escolha discreta.

Além disso, apesar da abordagem de maximização da utilidade aleatória (*RUM*) ser a de maior uso em aplicações teóricas e empíricas no âmbito da modelagem de escolha discreta, a abordagem de minimização do arrependimento aleatório (*RRM*) tem se apresentado como alternativa para caracterização de diferentes fenômenos comportamentais. Entretanto, no caso específico da escolha de destinos, nota-se que a aplicação de modelos sob a abordagem *RRM* ainda são escassas na literatura, inclusive inexistentes no contexto de países do Sul Global.

Nesta dissertação foi apresentada uma comparação das abordagens *RUM* e *RRM* para modelagem da escolha de destinos no contexto de viagens urbanas a trabalho pelo modo automóvel no município de São Paulo. Um método de amostragem estratificada para definição do conjunto de alternativas disponíveis para escolha pelos indivíduos foi utilizado para aplicação de ambas as abordagens, e a definição do número de alternativas foi feito com o uso da abordagem de Guevara, Chorus e Ben-Akiva (2016). Os dados da pesquisa OD 2017 executada pelo Companhia do Metropolitano de São Paulo na Região Metropolitana de São Paulo foram utilizados na modelagem, cujos resultados indicaram que a abordagem *RUM* obteve um desempenho melhor que a *RRM* em termos de valores finais de log-verossimilhança para todos os modelos estimados.

Em termos de desempenho, destacaram-se os modelos que relacionaram o logaritmo de empregos totais (*EMPLN*) e distância (*DIS*), e os modelos que relacionaram as primeiras variáveis e a de centro expandido (*CENT*), demonstrando que essas variáveis são importantes na estimação de destinos de viagens a trabalho no município de São Paulo.

Os modelos sob a abordagem *RRM* apresentaram tempos elevados de estimação em comparação àqueles sob a abordagem *RUM*, podendo ser um fator limitante em aplicações práticas e com maior número de variáveis naqueles casos. Apesar disso, buscou-se verificar o

melhor desempenho para os modelos *RRM* com a aplicação de método para determinação do número de alternativas ótimo, e apesar da diferença nos tempos computacionais, os modelos em termos de log-verossimilhança final têm diferença percentual menor que 1%.

Todos os parâmetros das variáveis de logaritmo de empregos totais, empregos por tipo, variáveis *dummies* e de distância foram significativas a 95% de confiança e apresentaram sinal esperado, indicando que as primeiras afetam positivamente a escolha de destinos em viagens motivo trabalho, e a última afeta de forma negativa essas escolhas.

A pergunta de pesquisa sobre qual é o desempenho da abordagem da Minimização do Arrependimento Aleatório (*RRM*) na modelagem de escolha de destinos comparada à Maximização da Utilidade Aleatória (*RUM*) foi respondida. A hipótese de que a abordagem *RRM* pode representar de forma satisfatória o comportamento dos indivíduos na escolha de destinos pode ser considerada verdadeira, já que os modelos *RRM* obtiveram resultados muito semelhantes aos modelos *RUM* em termos de medidas de desempenho. Entretanto, o elevado tempo computacional pode não justificar aplicações práticas desta abordagem em comparação a outra.

Portanto, conclui-se que, apesar do *RUM* apresentar desempenho superior na maioria dos modelos, a diferença menor que 1% não descarta sua utilização para modelos de escolha de destinos. Os modelos *RRM* podem ser utilizados para a estimativa, apesar do maior esforço computacional para estimativa dos modelos e para definição do número ótimo de alternativas.

Os modelos de previsão de demanda aplicados ao processo de planejamento de transportes são úteis para elaboração de diversas políticas públicas, tais como dos planos norteadores dos aspectos da mobilidade de um município. Considera-se que os modelos de escolha discreta aplicados à etapa de distribuição de viagens analisados nesta pesquisa são importantes na modelagem para o planejamento de transportes pois possibilitam a utilização de diferentes variáveis disponíveis envolvidas no processo, além daquelas de impedância tradicionalmente usadas nos modelos agregados. Em aplicações práticas, muitas vezes, não há disponibilidade de determinadas informações favorecendo a aplicação dos modelos de escolha discreta, à medida que permitem flexibilidade na seleção dos atributos não somente relacionados à impedância. Além disso, o uso dos modelos com a abordagem *RUM* exige menos tempo computacional, permitindo maior rapidez no processo de modelagem.

Como principal limitação deste trabalho, cabe mencionar a dificuldade para estimação

dos modelos *RRM* devido ao elevado tempo computacional, o que limitou maior número de testes com diferentes números de alternativas. Além disso, o recorte da dissertação limitou-se a analisar apenas viagens por motivo trabalho.

Possibilidade de pesquisas futuras incluem a modelagem de escolha discreta para outros motivos de viagens no município de São Paulo, como por exemplo de viagens a lazer e saúde, testando os métodos de amostragem aqui aplicados e comparando o desempenho sob abordagem *RUM* e *RRM* e, assim, aprofundar-se na modelagem de distribuição de viagens desagregada não só para o município como para demais cidades brasileiras. Além disso, recomenda-se a continuidade do estudo dos modelos *RRM* em aplicações práticas em modelos de escolha discreta.

6 REFERÊNCIAS BIBLIOGRÁFICAS

BADDOE, D. A.; MILLER, E. J. **Transportation–land-use interaction: empirical findings in North America, and their implications for modeling.** Transportation Research Part D: Transport and Environment, v. 5, n. 4, p. 235–263, 1 jul. 2000.

BANCO CENTRAL DO BRASIL. **Calculadora do cidadão:** correção de valores. Disponível em:

<https://www3.bcb.gov.br/CALCIDADA0/publico/exibirFormCorrecaoValores.do?method=exibirFormCorrecaoValore>. Acesso em 20 de mai. de 2020.

BELL, D. E. **Regret in Decision Making Under Uncertainty.** Operations Research, v. 30, n. 5, p. 961–981, 1982.

BEKHOR, S.; PRASHKER, J. N. **GEV-based destination choice models that account for unobserved similarities among alternatives.** Transportation Research Part B: Methodological, v. 42, n. 3, p. 243–262, 2008.

BEN-AKIVA, M. E. **Structure of passenger travel demand models.** 1973.

BEN-AKIVA, M. E.; BIERLAIRE, M. **Discrete Choice Methods and Their Applications to Short Term Travel Decisions.** N.p., 1999. 5–33.

BEN-AKIVA, M. E.; LERMAN, S. R. **Discrete Choice Analysis: Theory and Application to Travel Demand.** London: Press, The MIT, 1985.

BHAT, C., GOVINDARAJAN, A., PULUGURTA, V. **Disaggregate attraction-end choice modeling formulation and empirical analysis.** Transportation Research Record, (1645), 60–68. 1998.

BHAT, C.; ZHAO, H. **The spatial analysis of activity stop generation.** Transportation Research Part B: Methodological, v. 36, n. 6, p. 557–575, 1 jul. 2002.

BOERI, M.; MASIERO, L. **Regret minimisation and utility maximisation in a freight transport context.** Transportmetrica A: Transport Science, v. 10, n. 6, p. 548–560, 2014.

BOLFARINE, H.; BUSSAB, W. O. **Elementos de amostragem.** São Paulo: Edgard Blucher, 2005.

BOWMAN, J. L.; BEN-AKIVA, M. **Activity-Based Disaggregate Travel Demand Model**

System with Activity Schedules. Transportation Research Part A: Policy and Practice, v. 35, p. 1–28, 2001.

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil.** Brasília, DF: Senado Federal: Centro Gráfico, 1988.

CARRASCO, N. **Deciding where to shop: Disaggregate random utility destination choice modeling of grocery shopping in Canton Zurich.** Doctoral dissertation, IVT, ETH Zurich. 2008.

CALDEIRA, G. P. **Dependência espacial sob a abordagem de minimização do arrependimento em modelos de escolha de modo de transporte.** 2021. Dissertação (Mestrado em Engenharia de Transportes) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2021. doi:10.11606/D.3.2021.tde-10022022-110603. Acesso em: 02 out. 2022.

CALDEIRA, G. P.; ISLER, C. A. **Escolha de modos de transporte em viagens urbanas a trabalho: uma abordagem pela minimização do arrependimento.** 34o Congresso de Pesquisa e Ensino em Transporte da ANPET. Anais...100% online: ANPET, 2019.

CHORUS, C. G.; ARENTZE, T.; TIMMERMANS, H.J.P. **A Random Regret-Minimization model of travel choice.** Transportation Research Part B: Methodological. 42. 1-18. 2008.

CHORUS, C. G. **A New Model of Random Regret Minimization.** EJTIR, v. 10, n. 10, p. 181–196, jun. 2010.

CHORUS, C. G. **Random Regret Minimization: An Overview of Model Properties and Empirical Evidence.** Transport Reviews, v. 32, n. 1, p. 75–92, jan. 2012.

CHOW, L. F. *et al.* **Development and evaluation of aggregate destination choice models for trip distribution in Florida.** Transportation Research Record, n. 1931, p. 18–27, 2005.

COMPANHIA DO METROPOLITANO DE SÃO PAULO. **Pesquisa Origem Destino 2017 - A mobilidade urbana da região metropolitana de São Paulo em detalhes.** Disponível em: http://www.metro.sp.gov.br/pesquisa-od/arquivos/Ebook%20Pesquisa%20OD%202017_final_240719_versao_4.pdf. Acesso em: 16 ago. 2021.

CONWAY, M.; BYRD, A.; VAN DER LINDEN, M. **Evidence-based transit and land use sketch planning using interactive accessibility methods on combined schedule and**

- headway-based networks.** Transportation Research Record, v. 2653, p. 45–53, 2017.
- COSTA NETO, P. L. O. **Estatística.** 2. Ed. São Paulo: Edgard Blucher, 2002.
- DALY, A. **Estimating choice models containing attraction variables.** Transportation Research Part B, v. 16, n. 1, p. 5–15, 1982.
- FÁVERO, L. P. L.; BELFIORE, P. P. **Manual de análise de dados: Estatística e modelagem multivariada com excel, SPSS e stata.** [S.l: s.n.], 2017.
- FREUND, J. E. **Estatística Aplicada - 11.ed.: Economia, Administração e Contabilidade.** [s.l.] Bookman Editora, 2009.
- GUEVARA, C. A.; CHORUS, C. G.; BEN-AKIVA, M. E. **Sampling of Alternatives in Random Regret Minimization Models Sampling of Alternatives in Random Regret Minimization Models.** Transportation Science, v. 50, n. 1, p. 306–321, 2016.
- GOMIDE, A. D. Á.; MORATO, R. **Instrumentos de desestímulo ao uso do transporte individual motorizado: lições e recomendações** IEMA. Iema, p. 1–64, 2011.
- HAMMADOU, H. *et al.* **How to incorporate the spatial dimension in destination choice models: The case of Antwerp.** Transportation Planning and Technology, v. 31, n. 2, p. 153–181, 2008.
- HANDY, S.; CAO, X.; MOKHTARIAN, P. **Correlation or causality between the built environment and travel behavior? Evidence from Northern California.** Transportation Research Part D: Transport and Environment, v. 10, n. 6, p. 427–444, 1 nov. 2005.
- INMAN, J. J.; DYER, J. S.; JIANMIN, J. **A generalized utility model of disappointment and regret effects on post-choice valuation.** Marketing Science, v. 16, n. 2, p. 97–111, 1997.
- JING, P. *et al.* **Travel mode and travel route choice behavior based on Random Regret Minimization: A systematic review.** Sustainability (Switzerland), v. 10, n. 4, 2018.
- JONNALAGADDA, N., FREEDMAN, J., DAVIDSON, W. A., & HUNT, J. D. **Development of Microsimulation Activity-Based Model for San Francisco: Destination and Mode Choice Models.** Transportation Research Record, 2001.
- KAPLAN, S.; PRATO, C. G. **The application of the random regret minimization model to drivers' choice of crash avoidance maneuvers.** Transportation Research Part F: Traffic

Psychology and Behaviour, v. 15, n. 6, p. 699–709, nov. 2012.

KEYA, N.; ANOWAR, S.; ELURU, N. **Joint model of freight mode choice and shipment size: A copula-based random regret minimization framework.** Transportation Research Part E: Logistics and Transportation Review, v. 125, p. 97–115, maio 2019.

KIM, C. *et al.* **A comparative study of aggregate and disaggregate gravity models using Seoul metropolitan subway trip data.** Transportation Planning and Technology, v. 32, n. 1, p. 59–70, 2009.

KIM, J.; LEE, S. **Comparative analysis of traveler destination choice models by method of sampling alternatives.** Transportation Planning and Technology, v. 40, n. 4, p. 465–478, 2017a.

KOPPELMAN, F. S.; BHAT, C. R. **A Self Instructing Course in Mode Choice Modeling : Multinomial and Nested Logit Models.** Austin: [s.n.].

KORNBROT, D. **Point Biserial Correlation.** In Wiley StatsRef: Statistics Reference Online - eds N. Balakrishnan, T. Colton, B. Everitt, W. Piegorisch, F. Ruggeri and J.L. Teugels. 2014.

LI, M. T. *et al.* **Geographically stratified importance sampling for the calibration of aggregated destination choice models for trip distribution.** Transportation Research Record, n. 1935, p. 85–92, 2005.

LOOMES, GRAHAM AND SUGDEN, R. **Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty** Published by : Wiley on behalf of the Royal Economic Society Stable URL : <http://www.jstor.org/stable/2232669> Accessed : 06-04-2016 00. The Economic Journal, v. 92, n. 368, p. 805–824, 1982.

LUCE, R. D.; SUPPES, P. **Preference, utility, and subjective probability.** In R. D. Luce, R. R. Bush, & E. Galanter (Fds.), Handbook of mathematical psychology, Vol. 111. New York: Wiley, 1965. Pp. 249-410.

MCFADDEN, D. **Conditional logit analysis of qualitative choice behavior. Working paper institute of urban and regional,** n. 199, 1972.

MISHRA, S. *et al.* **Comparison between Gravity and Destination Choice Models for Trip Distribution in Maryland.** Transportation Research Board, n. January, p. 1–22, 2013.

MOLLOY, J. **Development of a Destination Choice Model for Ontario**. 2016. Disponível em: www.msm.bgu.tum.de. Acesso em: 20 out. 2020.

NEYMAN, J. **On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection**. Journal of the Royal Statistical Society, v. 97, n. 4, p. 558, 1934.

ORTÚZAR, O.; WILLUMSEN, L. G. **Modelling Transport**. Fourth Edition. 2011.

PARADY, G., ORY, D., e WALKER, J. **The overreliance on statistical goodness-of-fit and under-reliance on model validation in discrete choice models: A review of validation practices in the transportation academic literature**. Journal of Choice Modelling 38. 2021.

PEREIRA, R. H. M., SARAIVA, M., HERSZENHUT, D., BRAGA, C. K. V., e CONWAY, M. W. **r5r: Rapid Realistic Routing on Multimodal Transport Networks with R 5 in R**. Findings, 1–10, 2021.

PITOMBO, C. S.; DE SOUZA, A. D.; LINDNER, A. **Comparing decision tree algorithms to estimate intercity trip distribution**. Transportation Research Part C: Emerging Technologies, v. 77, p. 16–32, 1 abr. 2017.

PREFEITURA MUNICIPAL DE SÃO PAULO. **GeoSampa**, 2022. Disponível em: http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx. Acesso em: 18 de set. de 2022.

POZSGAY, M. A.; BHAT, C. R. **Destination Choice Modeling for Home-Based Recreational Trips Analysis and Implications for Land Use, Transportation, and Air Quality Planning**. [s.l: s.n.].

PRASHKER, J.; SHIFTAN, Y.; HERSHKOVITCH-SARUSI, P. **Residential choice location, gender and the commute trip to work in Tel Aviv**. Journal of Transport Geography, v. 16, n. 5, p. 332–341, 2008.

R CORE TEAM. **R: The R Project for Statistical Computing**. Vienna R Foundation for Statistical Computing, 2019. Disponível em: <https://www.r-project.org/>. Acesso em: out. de 2022.

SADHU, S. L. N. S.; TIWARI, G. **An activity pattern - destination land use choice model of low income households of informal settlements - Case study of Delhi**. Transportation

Research Part A: Policy and Practice, v. 85, p. 265–275, 2016.

SÃO PAULO. Lei no 16.402, de 22 de março de 2016

SHIFTAN, Y. **Practical Approach to Model Trip Chaining**. Transportation Research Record, 1645(1), 17–23, 1998.

SIMMA, A.; SCHLICH, R.; AXHAUSEN, K. W. **Destination choice modelling for different leisure activities. BRISK Binary Robust Invariant Scalable Keypoints**, n. May, p. 12–19, 2001.

SINGH, A. S.; SINGH, A. S.; MASUKU, M. B. **Sampling Techniques and the determinants of sample size in applied statistics**. 2014.

SIMONSON, I. **The Influence of Anticipating Regret and Responsibility on Purchase Decisions**. Journal of Consumer Research, v. 19, n. 1, p. 105, 1992.

SIVAKUMAR, A.; BHAT, C. R. **Comprehensive, unified framework for analyzing spatial location choice**. Transportation Research Record, 2007.

SPEAR, B. D. **Applications of New Travel Demand Forecasting Techniques to Transportation Planning**. Disponível em: <https://ntlrepository.blob.core.windows.net/docs/SICM.html>. Acesso em: 5 nov. 2021. 1977.

STEAD, D. **Relationships between Land Use, Socioeconomic Factors, and Travel Patterns in Britain**. Environment and Planning B: Planning and Design, 28(4), 499–528, 2001.

TAHERDOOST, H. **Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research**. SSRN Electronic Journal, v. 5, n. 2, p. 18–27, 2018.

THIENE, M.; BOERI, M.; CHORUS, C. G. **Random Regret Minimization: Exploration of a New Choice Model for Environmental and Resource Economics**. Environmental and Resource Economics, v. 51, n. 3, p. 413–429, set. 2012.

TRAVEL FORECASTING RESOURCE. Disponível em: http://tfresource.org/Category:Destination_Choice_Models. Acesso em: 5 de dez. de 2019.

VAN WEE, B. **Land use and transport: research and policy challenges**. Journal of Transport Geography, v. 10, n. 4, p. 259–271, 1 dez. 2002.

YANG, M. *et al.* **Modeling destination choice behavior incorporating spatial factors, individual sociodemographics, and travel mode.** *Journal of Transportation Engineering*, v. 136, n. 9, p. 800–810, 2010.

ZEELLENBERG, M. **The use of crying over spilled milk: A note on the rationality and functionality of regret.** *Philosophical Psychology*, v. 12, n. 3, p. 325–340, 1999.

APÊNDICE A – Resultados obtidos no processo para obtenção do J ótimo nos modelos *RRM*

Tabela 1 – Resultados para J = 5

Iteração	Empregos	Distância	N	LL0	LLfinal	ρ_2	ρ_2 ajustado	AIC	BIC	Tempo computacional	Iterações
1	0,0058934	-0,0008663	3134	-5043,98	-3599,49	0,29	0,29	7202,99	7215,09	01:27:46,8	15
2	0,0059633	-0,0008536	3122	-5043,98	-3604,14	0,29	0,29	7212,27	7224,37	01:28:57,67	15
3	0,0047358	-0,0008415	3122	-7188,67	-5482,55	0,24	0,24	7224,37	10981,19	01:29:39,97	13
4	0,0059603	-0,0008528	3135	-5045,59	-3619,12	0,28	0,28	7242,24	7254,34	01:26:54	14
5	0,0055620	-0,0008707	3134	-5043,98	-3614,64	0,28	0,28	7233,28	7245,38	01:15:10	15
6	0,0059727	-0,0008846	3134	-5043,98	-3565,13	0,29	0,29	7134,25	7146,35	01:21:42	15
7	0,0056403	-0,0008483	3132	-5040,76	-3642,31	0,28	0,28	7288,62	7300,72	01:25:50,25	15
8	0,0057682	-0,0008416	3134	-5043,98	-3634,36	0,28	0,28	7272,72	7284,82	01:30:45,87	15
9	0,0055617	-0,0008721	3135	-5045,59	-3625,70	0,28	0,28	7255,40	7267,50	01:35:28,19	14
10	0,0057911	-0,0008527	3130	-5037,54	-3626,23	0,28	0,28	7256,46	7268,55	01:35:12,89	15
11	0,0056449	-0,0008570	3134	-5043,98	-3627,14	0,28	0,28	7258,27	7270,37	01:36:19,21	15
12	0,0058413	-0,0008438	3135	-5045,59	-3627,61	0,28	0,28	7259,23	7271,33	01:36:25,16	15
13	0,0056864	-0,0008866	3135	-5045,59	-3590,59	0,29	0,29	7185,18	7197,28	01:33:27,67	15
14	0,0053304	-0,0008365	3133	-5042,37	-3693,51	0,27	0,27	7391,02	7403,12	01:32:21,37	14
15	0,0059364	-0,0008810	3135	-5045,59	-3583,24	0,29	0,29	7170,48	7182,58	01:38:21,57	15
16	0,0059034	-0,0008565	3130	-5037,54	-3601,72	0,29	0,28	7207,43	7219,53	00:43:8,44	15
17	0,0056524	-0,0008666	3131	-5039,15	-3610,73	0,28	0,28	7225,46	7237,56	01:41:22,26	15
18	0,0058251	-0,0008707	3131	-5039,15	-3583,78	0,29	0,29	7171,56	7183,66	01:30:17,41	15
19	0,0057207	-0,0008457	3132	-5040,76	-3620,74	0,28	0,28	7245,48	7257,58	01:28:52,6	15
20	0,0055391	-0,0008705	3135	-5045,59	-3626,25	0,28	0,28	7256,50	7268,60	01:25:49,94	15
21	0,0055391	-0,0008705	3135	-5045,59	-3626,25	0,28	0,28	7256,50	7268,60	01:25:49,06	15

Iteração	Empregos	Distância	N	LL0	LLfinal	ρ_2	ρ_2 ajustado	AIC	BIC	Tempo computacional	Iterações
22	0,0056532	-0,0008792	3135	-5045,59	-3615,05	0,28	0,28	7234,09	7246,19	01:27:34,93	15
23	0,0057989	-0,0008707	3134	-5043,98	-3609,91	0,28	0,28	7223,81	7235,91	01:56:26,81	15
24	0,0059720	-0,0008702	3132	-5040,76	-3591,62	0,29	0,29	7187,24	7199,34	01:37:22,48	15
25	0,0055648	-0,0008646	3135	-5045,59	-3651,24	0,28	0,28	7306,47	7318,58	01:44:17,59	13
26	0,0057325	-0,0008542	3135	-5045,59	-3620,31	0,28	0,28	7244,62	7256,72	01:35:4,92	13
27	0,0058409	-0,0008687	3133	-5042,37	-3609,13	0,28	0,28	7222,26	7234,36	01:34:33,95	15
28	0,0056074	-0,0008666	3135	-5045,59	-3631,26	0,28	0,28	7266,53	7278,63	01:27:11,57	13
29	0,0055532	-0,0008784	3134	-5043,98	-3620,34	0,28	0,28	7244,68	7256,78	01:36:57,41	13
30	0,0058048	-0,0008624	3131	-5039,15	-3629,70	0,28	0,28	7263,39	7275,49	01:38:17,37	13

Tabela 2 – Resultados para J = 10

Iteração	Empregos	Distância	N	LL0	LLfinal	ρ_2	ρ_2 ajustado	AIC	BIC	Tempo computacional	Iterações
1	0,0053975	-0,0009982	3126	-7197,88	-5448,66	0,24	0,24	10901,31	10913,41	01:47:21,64	13
2	0,0047405	-0,0008416	3122	-7188,67	-5481,89	0,24	0,24	10967,78	10979,87	01:10:38,26	13
3	0,0047358	-0,0008415	3122	-7188,67	-5482,55	0,24	0,24	10969,10	10981,19	01:29:39,97	13
4	0,0051836	-0,0009720	3116	-7174,86	-5480,85	0,24	0,24	10965,71	10977,80	01:31:22,2	14
5	0,0054809	-0,0010028	3119	-7181,76	-5408,76	0,25	0,25	10821,52	10833,61	01:22:42,42	14
6	0,0055238	-0,0009626	3123	-7190,97	-5452,92	0,24	0,24	10909,84	10921,93	01:35:25,2	13
7	0,0047358	-0,0008415	3122	-7188,67	-5482,55	0,24	0,24	10969,10	10981,19	01:27:49,57	13
8	0,0052838	-0,0009756	3123	-7190,97	-5484,40	0,24	0,24	10972,81	10984,90	02:00:44,4	14
9	0,0054025	-0,0009445	3123	-7190,97	-5503,97	0,23	0,23	11011,93	11024,02	01:36:51,57	14
10	0,0052339	-0,0009456	3114	-7170,25	-5519,66	0,23	0,23	11043,33	11055,41	01:54:4,86	14
11	0,0056753	-0,0009419	3120	-7184,07	-5448,19	0,24	0,24	10900,38	10912,47	01:37:28,41	13

Iteração	Empregos	Distância	N	LL0	LLfinal	ρ^2	ρ^2 ajustado	AIC	BIC	Tempo computacional	Iterações
12	0,0053271	-0,0009857	3118	-7179,46	-5448,03	0,24	0,24	10900,06	10912,15	01:35:50,39	14
13	0,0053693	-0,0009794	3119	-7181,76	-5458,99	0,24	0,24	10921,98	10934,07	01:37:0,31	14
14	0,0052202	-0,0009490	3122	-7188,67	-5519,90	0,23	0,23	11043,80	11055,89	01:33:54,33	14
15	0,0053123	-0,0009701	3120	-7184,07	-5482,94	0,24	0,24	10969,89	10981,98	01:37:4,27	14
16	0,0054275	-0,0009864	3123	-7190,97	-5433,23	0,24	0,24	10870,46	10882,56	01:49:0,94	14
17	0,0053234	-0,0009460	3123	-7190,97	-5504,35	0,23	0,23	11012,70	11024,79	01:37:13,7	14
18	0,0053958	-0,0009732	3122	-7188,67	-5449,27	0,24	0,24	10902,54	10914,63	01:39:43,19	14
19	0,0055062	-0,0009604	3116	-7174,86	-5466,01	0,24	0,24	10936,01	10948,10	01:37:46,57	14
20	0,0053603	-0,0009510	3120	-7184,07	-5481,43	0,24	0,24	10966,85	10978,94	02:21:23,55	14
21	0,0053603	-0,0009510	3120	-7184,07	-5481,43	0,24	0,24	10966,85	10978,94	01:38:8,74	14
22	0,0053576	-0,0009811	3119	-7181,76	-5455,37	0,24	0,24	10914,75	10926,84	01:39:17,3	14
23	0,0055454	-0,0009778	3119	-7181,76	-5434,73	0,24	0,24	10873,47	10885,56	01:38:31,33	14
24	0,0053219	-0,0009455	3117	-7177,16	-5498,79	0,23	0,23	11001,59	11013,68	01:49:26,79	14
25	0,0051753	-0,0009472	3123	-7190,97	-5522,26	0,23	0,23	11048,51	11060,60	02:04:31,61	14
26	0,0051898	-0,0009536	3120	-7184,07	-5514,54	0,23	0,23	11033,08	11045,18	01:41:45,02	14
27	0,0053030	-0,0009658	3118	-7179,46	-5477,09	0,24	0,24	10958,18	10970,27	01:44:56,25	14
28	0,0046631	-0,0008699	3117	-7177,16	-5456,87	0,24	0,24	10917,73	10929,82	01:33:31,38	13
29	0,0047404	-0,0008415	3122	-7188,67	-5482,09	0,24	0,24	10968,17	10980,26	01:34:43,29	13
30	0,0047358	-0,0008415	3122	-7188,67	-5482,55	0,24	0,24	10969,10	10981,19	01:06:49,19	13

Tabela 3 – Resultados para J = 15

Iteração	Empregos	Distância	N	LL0	LLfinal	ρ_2	ρ_2 ajustado	AIC	BIC	Tempo computacional	Iterações
1	0,0048590	-0,0011141	3111	-8424,74	-6506,68	0,23	0,23	13017,36	13029,45	0,03000382	13
2	0,0050343	-0,0010752	3117	-8440,99	-6547,39	0,22	0,22	13098,77	13110,86	0,02913519	14
3	0,0049114	-0,0011060	3113	-8430,16	-6522,46	0,23	0,23	13048,93	13061,02	0,03087141	14
4	0,0048185	-0,0011001	3113	-8430,16	-6534,76	0,22	0,22	13073,52	13085,61	0,03016111	14
5	0,0051164	-0,0010835	3108	-8416,62	-6505,76	0,23	0,23	13015,53	13027,61	0,03021806	14
6	0,0049849	-0,0010900	3117	-8440,99	-6534,76	0,23	0,23	13073,51	13085,60	0,03144051	14
7	0,0049519	-0,0010910	3114	-8432,87	-6524,59	0,23	0,23	13053,17	13065,26	0,03045231	14
8	0,0050535	-0,0011124	3119	-8446,41	-6500,75	0,23	0,23	13005,50	13017,59	0,03145648	13
9	0,0046853	-0,0011032	3117	-8440,99	-6545,48	0,22	0,22	13094,96	13107,05	0,02981713	14
10	0,0048106	-0,0010865	3115	-8435,58	-6561,22	0,22	0,22	13126,44	13138,53	0,02861238	14
11	0,0049700	-0,0010779	3114	-8432,87	-6553,68	0,22	0,22	13111,36	13123,44	0,02819988	14
12	0,0049752	-0,0010846	3115	-8435,58	-6541,63	0,22	0,22	13087,25	13099,34	0,02857639	14
13	0,0048367	-0,0010942	3120	-8449,12	-6552,52	0,22	0,22	13109,04	13121,13	0,0296934	14
14	0,0049938	-0,0010964	3116	-8438,28	-6522,11	0,23	0,23	13048,23	13060,32	0,02807894	14
15	0,0048154	-0,0010959	3116	-8438,28	-6542,59	0,22	0,22	13089,18	13101,26	0,02875428	14
16	0,0049358	-0,0011183	3112	-8427,45	-6485,35	0,23	0,23	12974,69	12986,78	0,02795613	13
17	0,0049649	-0,0010892	3108	-8416,62	-6513,94	0,23	0,23	13031,89	13043,97	0,02809306	14
18	0,0050430	-0,0010922	3102	-8400,37	-6481,02	0,23	0,23	12966,04	12978,12	0,03039491	14
19	0,0048337	-0,0011039	3118	-8443,70	-6531,50	0,23	0,23	13067,00	13079,09	0,02903762	13
20	0,0047105	-0,0010943	3114	-8432,87	-6561,26	0,22	0,22	13126,53	13138,61	0,02939734	14
21	0,0047105	-0,0010943	3114	-8432,87	-6561,26	0,22	0,22	13126,53	13138,61	0,02904664	14
22	0,0050053	-0,0010817	3111	-8424,74	-6521,10	0,23	0,23	13046,20	13058,29	0,02862708	14
23	0,0049018	-0,0010996	3116	-8438,28	-6531,55	0,23	0,23	13067,10	13079,19	0,02979375	13
24	0,0049083	-0,0010889	3108	-8416,62	-6519,31	0,23	0,23	13042,62	13054,71	0,02914236	14

Iteração	Empregos	Distância	N	LL0	LLfinal	ρ^2	ρ^2 ajustado	AIC	BIC	Tempo computacional	Iterações
25	0,0049692	-0,0010845	3111	-8424,74	-6532,68	0,22	0,22	13069,36	13081,44	0,03049086	14
26	0,0048562	-0,0011080	3114	-8432,87	-6511,53	0,23	0,23	13027,07	13039,15	0,03115289	14
27	0,0049678	-0,0010987	3108	-8416,62	-6496,62	0,23	0,23	12997,23	13009,32	0,02992361	14
28	0,0049030	-0,0010889	3114	-8432,87	-6531,35	0,23	0,23	13066,69	13078,78	0,02939398	14
29	0,0047408	-0,0011091	3118	-8443,70	-6536,52	0,23	0,23	13077,05	13089,14	0,02946296	13
30	0,0048731	-0,0010934	3114	-8432,87	-6534,23	0,23	0,22	13072,46	13084,54	0,03133137	14

Tabela 4 – Resultados para J = 20

Iteração	Empregos	Distância	N	LL0	LLfinal	ρ^2	ρ^2 ajustado	AIC	BIC	Tempo computacional	Iterações
1	0,0047788	-0,0011000	3102	-9292,76	-7371,14	0,21	0,21	14746,28	14758,36	00:42:45,93	14
2	0,0047534	-0,0011104	3098	-9280,78	-7340,51	0,21	0,21	14685,02	14697,10	00:40:35,41	14
3	0,0047196	-0,0011386	3093	-9265,80	-7300,16	0,21	0,21	14604,31	14616,39	00:40:47,09	14
4	0,0048546	-0,0011274	3103	-9295,76	-7315,76	0,21	0,21	14635,51	14647,59	00:40:39,73	14
5	0,0048154	-0,0011279	3093	-9265,80	-7292,48	0,21	0,21	14588,97	14601,04	00:43:13,64	14
6	0,0047826	-0,0011318	3099	-9283,77	-7311,30	0,21	0,21	14626,59	14638,67	00:40:54,78	14
7	0,0049310	-0,0011246	3095	-9271,79	-7296,01	0,21	0,21	14596,03	14608,10	00:40:18,42	14
8	0,0047360	-0,0011559	3094	-9268,80	-7259,79	0,22	0,22	14523,57	14535,65	00:39:37,17	13
9	0,0047306	-0,0011407	3099	-9283,77	-7293,34	0,21	0,21	14590,68	14602,76	00:40:20,31	14
10	0,0046688	-0,0011096	3096	-9274,79	-7350,42	0,21	0,21	14704,84	14716,92	00:40:4,56	14
11	0,0047607	-0,0011679	3105	-9301,75	-7279,32	0,22	0,22	14562,63	14574,71	00:39:54,94	13
12	0,0047657	-0,0011219	3102	-9292,76	-7325,45	0,21	0,21	14654,90	14666,98	00:40:13,49	14
13	0,0047745	-0,0011330	3096	-9274,79	-7301,70	0,21	0,21	14607,40	14619,47	00:39:59,61	14
14	0,0047179	-0,0011266	3096	-9274,79	-7330,34	0,21	0,21	14664,69	14676,76	00:39:56,85	14
15	0,0048720	-0,0011328	3106	-9304,74	-7311,66	0,21	0,21	14627,31	14639,39	00:40:37,77	14

Iteração	Empregos	Distância	N	LL0	LLfinal	ρ^2	ρ^2 ajustado	AIC	BIC	Tempo computacional	Iterações
16	0,0047449	-0,0011272	3102	-9292,76	-7330,43	0,21	0,21	14664,87	14676,95	00:39:53,84	14
17	0,0048179	-0,0011569	3090	-9256,81	-7244,24	0,22	0,22	14492,47	14504,54	00:39:59,6	14
18	0,0047146	-0,0011358	3102	-9292,76	-7330,82	0,21	0,21	14665,65	14677,73	00:40:18,13	14
19	0,0049100	-0,0011046	3100	-9286,77	-7323,07	0,21	0,21	14650,13	14662,21	00:40:56,66	14
20	0,0048260	-0,0011396	3102	-9292,76	-7295,31	0,21	0,21	14594,63	14606,71	00:41:57,89	13
21	0,0048260	-0,0011396	3102	-9292,76	-7295,31	0,21	0,21	14594,63	14606,71	00:49:26,63	13
22	0,0048094	-0,0011201	3098	-9280,78	-7315,44	0,21	0,21	14634,87	14646,95	00:48:45,15	14
23	0,0048353	-0,0011298	3102	-9292,76	-7309,05	0,21	0,21	14622,09	14634,17	00:46:44,23	14
24	0,0048137	-0,0011044	3100	-9286,77	-7346,72	0,21	0,21	14697,45	14709,53	00:46:49,11	14
25	0,0046548	-0,0011300	3098	-9280,78	-7338,23	0,21	0,21	14680,46	14692,53	00:41:1,82	14
26	0,0047356	-0,0011397	3102	-9292,76	-7293,05	0,22	0,22	14590,10	14602,18	00:41:47,56	14
27	0,0048206	-0,0011030	3095	-9271,79	-7340,04	0,21	0,21	14684,08	14696,16	00:40:28,16	14
28	0,0045888	-0,0011235	3102	-9292,76	-7359,87	0,21	0,21	14723,75	14735,83	00:42:19,35	14
29	0,0048102	-0,0011377	3102	-9292,76	-7290,76	0,22	0,22	14585,53	14597,61	00:42:33,76	14
30	0,0048204	-0,0011245	3103	-9295,76	-7323,14	0,21	0,21	14650,27	14662,35	00:47:46,34	14

Tabela 5 – Resultados para J = 25

Iteração	Empregos	Distância	N	LL0	LLfinal	ρ^2	ρ^2 ajustado	AIC	BIC	Tempo computacional	Iterações
1	0,0046760	-0,0011673	3092	-9952,76	-7930,54	0,20	0,20	15865,07	15877,14	00:41:53,21	13
2	0,0048034	-0,0011518	3081	-9917,36	-7903,16	0,20	0,20	15810,32	15822,38	00:40:54,59	13
3	0,0045008	-0,0011773	3093	-9955,98	-7947,15	0,20	0,20	15898,30	15910,38	00:42:59,58	13
4	0,0046226	-0,0011646	3091	-9949,55	-7941,34	0,20	0,20	15886,68	15898,76	00:40:39,56	14
5	0,0047096	-0,0011686	3088	-9939,89	-7910,80	0,20	0,20	15825,60	15837,68	00:40:50,18	12
6	0,0047178	-0,0011910	3085	-9930,23	-7865,23	0,21	0,21	15734,46	15746,53	00:42:37,68	12
7	0,0045027	-0,0011755	3090	-9946,33	-7949,44	0,20	0,20	15902,88	15914,95	00:44:54,14	13

Iteração	Empregos	Distância	N	LL0	LLfinal	ρ^2	ρ^2 ajustado	AIC	BIC	Tempo computacional	Iterações
8	0,0046929	-0,0011411	3087	-9936,67	-7955,31	0,20	0,20	15914,61	15926,68	00:43:53,69	14
9	0,0046256	-0,0011874	3094	-9959,20	-7918,22	0,20	0,20	15840,45	15852,52	00:39:43,58	13
10	0,0048433	-0,0011558	3099	-9975,30	-7952,71	0,20	0,20	15909,41	15921,49	00:39:53,16	14
11	0,0048476	-0,0011859	3095	-9962,42	-7890,87	0,21	0,21	15785,73	15797,81	00:39:47,6	13
12	0,0047187	-0,0011560	3090	-9946,33	-7937,91	0,20	0,20	15879,81	15891,89	00:39:39	14
13	0,0046750	-0,0012052	3086	-9933,45	-7849,03	0,21	0,21	15702,07	15714,14	00:38:50,98	13
14	0,0046600	-0,0011580	3094	-9959,20	-7940,50	0,20	0,20	15885,00	15897,08	00:40:24,39	14
15	0,0046709	-0,0011759	3095	-9962,42	-7934,46	0,20	0,20	15872,91	15884,99	00:51:0,66	13
16	0,0047745	-0,0011665	3093	-9955,98	-7917,74	0,20	0,20	15839,47	15851,54	00:51:57,85	13
17	0,0046528	-0,0011646	3087	-9936,67	-7921,83	0,20	0,20	15847,65	15859,72	00:41:51,66	14
18	0,0047429	-0,0011572	3088	-9939,89	-7919,40	0,20	0,20	15842,79	15854,86	00:42:5,6	14
19	0,0046655	-0,0011807	3090	-9946,33	-7912,14	0,20	0,20	15828,27	15840,34	00:43:22,23	13
20	0,0048617	-0,0011592	3099	-9975,30	-7947,21	0,20	0,20	15898,41	15910,49	00:44:12,62	13
21	0,0048617	-0,0011592	3099	-9975,30	-7947,21	0,20	0,20	15898,41	15910,49	00:41:56,41	13
22	0,0046986	-0,0011562	3102	-9984,95	-7963,45	0,20	0,20	15930,89	15942,97	00:42:16,95	14
23	0,0046838	-0,0011642	3090	-9946,33	-7929,61	0,20	0,20	15863,23	15875,30	00:43:6,63	13
24	0,0048260	-0,0011442	3087	-9936,67	-7928,84	0,20	0,20	15861,68	15873,75	00:43:39,35	14
25	0,0047704	-0,0011882	3091	-9949,55	-7902,59	0,21	0,21	15809,18	15821,25	00:41:56,72	13
26	0,0044666	-0,0011471	3095	-9962,42	-8018,82	0,20	0,19	16041,64	16053,71	00:42:16,06	14
27	0,0048389	-0,0011559	3091	-9949,55	-7915,12	0,20	0,20	15834,23	15846,31	00:51:14,56	13
28	0,0047420	-0,0011656	3095	-9962,42	-7945,60	0,20	0,20	15895,20	15907,27	00:50:44,38	13
29	0,0046627	-0,0011868	3087	-9936,67	-7890,04	0,21	0,21	15784,09	15796,16	00:50:54,62	13
30	0,0045272	-0,0011802	3098	-9972,08	-7957,82	0,20	0,20	15919,64	15931,71	00:45:32,83	13

Tabela 6 – Resultados para J = 30

Iteração	Empregos	Distância	N	$LL0$	$LLfinal$	$\rho2$	$\rho2$ ajustado	AIC	BIC	Tempo computacional	Iterações
1	0,0048289	-0,0012192	3093	-10519,90	-8371,12	0,20	0,20	16746,25	16758,32	00:40:43	13
2	0,0047263	-0,0012160	3091	-10513,10	-8408,19	0,20	0,20	16820,38	16832,45	00:48:39	12
3	0,0047647	-0,0011883	3088	-10502,90	-8424,45	0,20	0,20	16852,89	16864,96	00:44:30	13
4	0,0046367	-0,0011995	3080	-10475,69	-8398,81	0,20	0,20	16801,61	16813,68	00:44:20	13
5	0,0047568	-0,0011938	3090	-10509,70	-8421,47	0,20	0,20	16846,93	16859,00	00:40:51	13
6	0,0047249	-0,0011907	3083	-10485,89	-8419,23	0,20	0,20	16842,46	16854,53	00:39:27	13
7	0,0047166	-0,0011825	3089	-10506,30	-8440,71	0,20	0,20	16885,41	16897,49	00:39:33	13
8	0,0043675	-0,0011918	3081	-10479,09	-8460,23	0,19	0,19	16924,45	16936,52	00:39:44	13
9	0,0045999	-0,0011965	3088	-10502,90	-8440,02	0,20	0,20	16884,05	16896,12	00:39:37	13
10	0,0045564	-0,0012611	3090	-10509,70	-8357,92	0,20	0,20	16719,84	16731,92	00:43:55	12
11	0,0048499	-0,0012286	3081	-10479,09	-8325,83	0,21	0,21	16655,66	16667,73	00:48:18	13
12	0,0046888	-0,0011839	3084	-10489,29	-8426,94	0,20	0,20	16857,87	16869,94	00:53:07	12
13	0,0047110	-0,0012024	3089	-10506,30	-8428,21	0,20	0,20	16860,42	16872,49	00:39:45	13
14	0,0047626	-0,0011948	3082	-10482,49	-8399,37	0,20	0,20	16802,74	16814,81	00:40:43	12
15	0,0046948	-0,0012156	3093	-10519,90	-8409,87	0,20	0,20	16823,74	16835,81	00:43:00	13
16	0,0048546	-0,0012000	3082	-10482,49	-8365,65	0,20	0,20	16735,30	16747,37	00:40:40	13
17	0,0047492	-0,0011794	3089	-10506,30	-8437,53	0,20	0,20	16879,06	16891,13	00:40:40	13
18	0,0047800	-0,0012243	3079	-10472,29	-8339,61	0,20	0,20	16683,22	16695,29	00:41:00	13
19	0,0048119	-0,0012143	3084	-10489,29	-8382,91	0,20	0,20	16769,81	16781,88	00:42:35	13
20	0,0048352	-0,0011951	3086	-10496,10	-8399,14	0,20	0,20	16802,28	16814,35	00:43:56	13
21	0,0047184	-0,0011839	3082	-10482,49	-8429,51	0,20	0,20	16863,02	16875,08	00:46:8,2	13
22	0,0046967	-0,0011960	3099	-10540,31	-8467,71	0,20	0,20	16939,42	16951,50	00:42:55	12
23	0,0047526	-0,0012206	3096	-10530,11	-8409,17	0,20	0,20	16822,34	16834,41	00:41:40	13

Iteração	Empregos	Distância	N	<i>LL0</i>	<i>LLfinal</i>	$\rho2$	$\rho2$ ajustado	AIC	BIC	Tempo computacional	Iterações
24	0,0047780	-0,0012000	3085	-10492,69	-8408,33	0,20	0,20	16820,67	16832,73	00:43:55	13
25	0,0045789	-0,0011969	3078	-10468,89	-8407,82	0,20	0,20	16819,64	16831,71	00:41:28	13
26	0,0046731	-0,0012317	3084	-10489,29	-8358,40	0,20	0,20	16720,79	16732,86	00:43:05	13
27	0,0047516	-0,0011893	3087	-10499,50	-8427,06	0,20	0,20	16858,13	16870,20	00:43:49	13
28	0,0047319	-0,0012217	3084	-10489,29	-8363,67	0,20	0,20	16731,34	16743,41	00:42:27	13
29	0,0047639	-0,0011599	3082	-10482,49	-8448,30	0,19	0,19	16900,60	16912,67	00:41:58	14
30	0,0047648	-0,0011662	3090	-10509,70	-8465,40	0,19	0,19	16934,79	16946,86	00:41:25	13

Tabela 7 – Resultados para J = 35

Iteração	Empregos	Distância	N	<i>LL0</i>	<i>LLfinal</i>	$\rho2$	$\rho2$ ajustado	AIC	BIC	Tempo computacional	Iterações
1	0,0047051	-0,0012496	3095	-11003,80	-8843,50	0,20	0,20	17690,99	17703,07	00:41:30	13
2	0,0048976	-0,0012833	3098	-11014,47	-8760,23	0,20	0,20	17524,46	17536,53	00:40:09	13
3	0,0049209	-0,0012711	3109	-11053,58	-8814,75	0,20	0,20	17633,49	17645,58	00:40:25	13
4	0,0048736	-0,0012466	3093	-10996,69	-8811,74	0,20	0,20	17627,48	17639,55	00:41:50	13
5	0,0048696	-0,0012869	3108	-11050,02	-8788,22	0,20	0,20	17580,43	17592,52	00:41:54	12
6	0,0046292	-0,0012589	3092	-10993,14	-8832,05	0,20	0,20	17668,10	17680,18	00:41:14	13
7	0,0048165	-0,0012907	3098	-11014,47	-8765,04	0,20	0,20	17534,08	17546,15	00:40:46	13
8	0,0047332	-0,0012618	3082	-10957,58	-8775,74	0,20	0,20	17555,49	17567,55	00:40:05	13
9	0,0046935	-0,0012657	3097	-11010,91	-8817,54	0,20	0,20	17639,08	17651,16	00:40:03	13
10	0,0047527	-0,0012672	3094	-11000,25	-8795,87	0,20	0,20	17595,73	17607,81	00:40:12	13
11	0,0047990	-0,0012752	3099	-11018,02	-8788,08	0,20	0,20	17580,15	17592,23	00:40:09	13
12	0,0048110	-0,0012445	3098	-11014,47	-8841,12	0,20	0,20	17686,24	17698,32	00:42:18	13
13	0,0048285	-0,0012547	3098	-11014,47	-8819,55	0,20	0,20	17643,11	17655,18	00:39:55	13
14	0,0047826	-0,0012376	3098	-11014,47	-8853,53	0,20	0,20	17711,07	17723,14	00:40:41	12

Iteração	Empregos	Distância	N	<i>LL0</i>	<i>LLfinal</i>	$\rho2$	$\rho2$ ajustado	AIC	BIC	Tempo computacional	Iterações
15	0,0048203	-0,0012404	3101	-11025,13	-8853,27	0,20	0,20	17710,54	17722,62	00:42:29	13
16	0,0049034	-0,0012305	3099	-11018,02	-8853,84	0,20	0,20	17711,68	17723,76	00:39:59	13
17	0,0048791	-0,0012554	3096	-11007,36	-8806,16	0,20	0,20	17616,32	17628,39	00:40:01	13
18	0,0046618	-0,0012250	3092	-10993,14	-8873,74	0,19	0,19	17751,48	17763,56	00:41:11	13
19	0,0046164	-0,0012602	3097	-11010,91	-8844,80	0,20	0,20	17693,59	17705,67	00:42:43	13
20	0,0046868	-0,0012419	3093	-10996,69	-8855,94	0,19	0,19	17715,89	17727,96	00:45:21	13
21	0,0047237	-0,0012315	3096	-11007,36	-8880,96	0,19	0,19	17765,92	17777,99	00:46:17	13
22	0,0047045	-0,0012562	3095	-11003,80	-8830,67	0,20	0,20	17665,34	17677,41	00:46:12	13
23	0,0045765	-0,0012717	3097	-11010,91	-8829,17	0,20	0,20	17662,34	17674,42	00:41:22	13
24	0,0048542	-0,0012534	3088	-10978,91	-8781,94	0,20	0,20	17567,87	17579,94	00:43:33	13
25	0,0047817	-0,0012721	3096	-11007,36	-8800,28	0,20	0,20	17604,56	17616,63	00:41:28	13
26	0,0047860	-0,0012718	3093	-10996,69	-8794,34	0,20	0,20	17592,68	17604,76	00:42:25	12
27	0,0048261	-0,0012667	3089	-10982,47	-8775,01	0,20	0,20	17554,02	17566,09	00:41:15	13
28	0,0048713	-0,0012384	3096	-11007,36	-8825,67	0,20	0,20	17655,34	17667,41	00:43:58	12
29	0,0046367	-0,0012535	3090	-10986,03	-8843,39	0,20	0,19	17690,77	17702,84	00:42:15	12
30	0,0047385	-0,0012512	3095	-11003,80	-8820,45	0,20	0,20	17644,90	17656,98	00:42:50	13

Tabela 8 – Resultados para J = 40

Iteração	Empregos	Distância	N	<i>LL0</i>	<i>LLfinal</i>	$\rho2$	$\rho2$ ajustado	AIC	BIC	Tempo computacional	Iterações
1	0,0048309	-0,0012757	3088	-11391,26	-9162,29	0,20	0,20	18328,58	18340,65	00:42:14,25	12
2	0,0047995	-0,0012795	3093	-11409,70	-9169,93	0,20	0,20	18343,85	18355,93	00:40:38,68	12
3	0,0048120	-0,0012883	3101	-11439,22	-9178,05	0,20	0,20	18360,09	18372,17	01:26:51,52	13
4	0,0048031	-0,0012740	3095	-11417,08	-9195,53	0,19	0,19	18395,07	18407,14	00:47:15,34	13
5	0,0047595	-0,0012751	3101	-11439,22	-9214,27	0,19	0,19	18432,55	18444,63	00:47:26,78	12
6	0,0048661	-0,0012812	3108	-11465,04	-9211,51	0,20	0,20	18427,01	18439,10	00:45:26,33	13

Iteração	Empregos	Distância	N	LL0	LLfinal	$\rho2$	$\rho2$ ajustado	AIC	BIC	Tempo computacional	Iterações
ou7	0,0047402	-0,0012585	3090	-11398,64	-9205,63	0,19	0,19	18415,25	18427,32	00:39:30,15	13
8	0,0047884	-0,0012570	3090	-11398,64	-9201,55	0,19	0,19	18407,09	18419,16	00:40:4,66	13
9	0,0047665	-0,0012623	3087	-11387,57	-9202,13	0,19	0,19	18408,27	18420,34	00:40:12,68	13
10	0,0047702	-0,0012688	3094	-11413,39	-9204,51	0,19	0,19	18413,02	18425,10	00:40:10,86	13
11	0,0047606	-0,0012701	3092	-11406,02	-9194,16	0,19	0,19	18392,32	18404,39	00:39:39,07	13
12	0,0046588	-0,0012667	3077	-11350,68	-9167,39	0,19	0,19	18338,78	18350,84	00:40:1,54	13
13	0,0048965	-0,0012778	3102	-11442,90	-9180,99	0,20	0,20	18365,99	18378,07	00:43:56,14	13
14	0,0050138	-0,0013158	3090	-11398,64	-9084,34	0,20	0,20	18172,68	18184,75	00:45:44,59	13
15	0,0047365	-0,0012881	3096	-11420,77	-9174,84	0,20	0,20	18353,68	18365,75	00:47:0,11	13
16	0,0048457	-0,0012748	3093	-11409,70	-9168,55	0,20	0,20	18341,09	18353,17	00:39:49,44	13
17	0,0047696	-0,0012571	3099	-11431,84	-9246,75	0,19	0,19	18497,50	18509,58	00:39:17,13	13
18	0,0047889	-0,0012651	3096	-11420,77	-9198,63	0,19	0,19	18401,27	18413,34	00:38:35,92	13
19	0,0047201	-0,0012654	3095	-11417,08	-9221,42	0,19	0,19	18446,83	18458,91	00:39:34,2	13
20	0,0047944	-0,0012749	3096	-11420,77	-9208,92	0,19	0,19	18421,85	18433,92	00:39:32,23	13
21	0,0048171	-0,0012612	3097	-11424,46	-9204,99	0,19	0,19	18413,99	18426,06	00:39:35,06	13
22	0,0047976	-0,0012784	3091	-11402,33	-9171,32	0,20	0,20	18346,64	18358,72	00:40:16,91	13
23	0,0047644	-0,0012639	3092	-11406,02	-9203,47	0,19	0,19	18410,93	18423,00	00:40:38,37	12
24	0,0049061	-0,0012976	3090	-11398,64	-9123,60	0,20	0,20	18251,20	18263,27	00:40:30,49	13
25	0,0048648	-0,0012806	3110	-11472,42	-9211,15	0,20	0,20	18426,29	18438,38	00:39:48,06	13
26	0,0048184	-0,0012682	3090	-11398,64	-9184,83	0,19	0,19	18373,66	18385,74	00:39:26,06	13
27	0,0047808	-0,0012728	3097	-11424,46	-9196,17	0,20	0,19	18396,33	18408,41	00:41:22	13
28	0,0047641	-0,0012822	3097	-11424,46	-9195,08	0,20	0,20	18394,16	18406,24	00:42:5	13
29	0,0048978	-0,0012469	3089	-11394,95	-9192,02	0,19	0,19	18388,03	18400,11	00:46:21,55	13
30	0,0048125	-0,0012295	3090	-11398,64	-9241,73	0,19	0,19	18487,46	18499,53	00:50:18,62	12