

ANO
2020

TIAGO LUIZ MACELAY

IDENTIFICAÇÃO DE FATORES QUE INFLUENCIAM A
ESCOLHA DO MODO DE TRANSPORTE DE ESTUDANTES
EM UM POLO UNIVERSITÁRIO NA CIDADE DE JOINVILLE



UDESC

UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC
CENTRO DE CIÊNCIAS TECNOLÓGICAS – CCT
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA CIVIL

O presente trabalho tem por objetivo determinar os principais fatores que influenciam na decisão pela escolha do modo de transporte de estudantes de um polo universitário localizado no município de Joinville, utilizando as técnicas de regressão logística multinomial e árvore de decisão. A área de estudo nesta pesquisa refere-se a duas instituições de ensino superior, uma pública e outra privada, localizadas na zona norte no município de Joinville, que juntas formam um polo universitário. Os resultados apontaram que fatores como tempo de deslocamento, distância percorrida, idade, gênero, renda familiar e cidade onde reside, são significativos para a escolha do modo de transporte dos estudantes do Polo Universitário em estudo.

Orientador: Elisa Henning

Joinville, 2020

DISSERTAÇÃO DE MESTRADO

**IDENTIFICAÇÃO DE FATORES QUE
INFLUENCIAM A ESCOLHA DO
MODO DE TRANSPORTE DE
ESTUDANTES EM UM POLO
UNIVERSITÁRIO NA CIDADE DE
JOINVILLE**

TIAGO LUIZ MACELAY

JOINVILLE, 2020

TIAGO LUIZ MACELAY

**IDENTIFICAÇÃO DE FATORES QUE INFLUENCIAM A ESCOLHA DO MODO DE
TRANSPORTE DE ESTUDANTES EM UM POLO UNIVERSITÁRIO NA CIDADE
DE JOINVILLE**

Dissertação apresentada como requisito
parcial para obtenção do título de Mestre em
Engenharia Civil, na Universidade do Estado
de Santa Catarina.
Orientadora: Profa. Dra. Elisa Henning

Joinville - SC

2020

**Ficha catalográfica elaborada pelo programa de geração automática da
Biblioteca Setorial do CCT/UEDESC,
com os dados fornecidos pelo(a) autor(a)**

Macelay, Tiago Luiz

Identificação de fatores que influenciam a escolha do modo de transporte de estudantes em um Polo Universitário na cidade de Joinville / Tiago Luiz Macelay. -- 2020.
112 p.

Orientadora: Elisa Henning

Dissertação (mestrado) -- Universidade do Estado de Santa Catarina, Centro de Ciências Tecnológicas, Programa de Pós-Graduação em Engenharia Civil, Joinville, 2020.

1. Regressão logística. 2. Árvore de decisão. 3. Escolha do modo de transporte. 4. Estudantes universitários. I. Henning, Elisa. II. Universidade do Estado de Santa Catarina, Centro de Ciências Tecnológicas, Programa de Pós-Graduação em Engenharia Civil. III. Título.

O original deste documento é eletrônico e foi assinado utilizando Assinatura Digital SGP-e por ANA MIRTHES HACKENBERG e ELISA HENNING em 21/05/2020 às 15:32:52, conforme Decreto Estadual nº 39, de 21 de fevereiro de 2019. Para verificar a autenticidade desta cópia impressa, acesse o site <https://portal.sgp.ea.sc.gov.br/portal-externo> e informe o processo UDESC 00010435/2020 e o código X9B993LF.

Tiago Luiz Macelay

Dedico este trabalho ao meu filho Lorenzo,
que ilumina todos os meus dias e acalenta
o meu coração.

AGRADECIMENTOS

Primeiramente agradeço aos meus familiares que sempre acreditaram e incentivaram na realização deste trabalho, especialmente minha esposa Lisiara, que me incentivou durante todo o mestrado, e todo dia me motiva a ser uma pessoa melhor.

À professora orientadora, Elisa Henning, pela sugestão do tema desenvolvido neste trabalho, além da ajuda, interesse e dedicação com que orientou o mesmo.

À colega Thamires Ferreira Schubert, pelo fornecimento dos dados para realização deste trabalho.

E por fim à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pela concessão de bolsa de estudos durante parte do período de realização da pesquisa.

RESUMO

A busca pela mobilidade urbana sustentável é fundamental para proporcionar qualidade de vida nas cidades. As instituições de ensino superior demandam um grande número de viagens diárias concentradas no tempo e no espaço, e se forem predominantemente de carro, geram mais impactos no sistema viário, causando congestionamentos. Sendo assim, é de suma importância compreender os fatores que levam os estudantes universitários a optar por determinado meio de transporte. O incentivo ao estudante para trocar o uso do automóvel por transportes coletivos ou meios de locomoção não motorizados é fundamental nas regiões metropolitanas. Por isso, é necessário inicialmente entender o modo de deslocamento dos estudantes, para que, posteriormente, possa se propor alternativas para a utilização de modos de transporte mais sustentáveis. Diante disto, o objetivo deste trabalho é determinar os principais fatores que influenciam na decisão pela escolha do modo de transporte de estudantes de um polo universitário localizado no município de Joinville, utilizando as técnicas de regressão logística multinomial e árvore de decisão. Os objetivos específicos são estimar um modelo estatístico de regressão logística multinomial que descreva a variável de estudo; estimar um modelo de árvore de decisão da escolha do modo de transporte entre estudantes universitários; comparar os resultados encontrados pelos dois métodos entre si; e comparar os resultados obtidos pelos dois métodos adotados na presente pesquisa com a literatura. A área de estudo nesta pesquisa refere-se a duas instituições de ensino superior, uma pública e outra privada, localizadas na zona norte no município de Joinville, que juntas formam um polo universitário. O banco de dados utilizado para a estimação dos modelos trata-se de um questionário, desenvolvido como parte da dissertação de mestrado de Schubert (2019), aplicado a alunos de graduação e pós-graduação no ano de 2018. Para realização das análises estatísticas, serão escolhidos os seguintes modos de transportes: carro, modo ativo (a pé e bicicleta), transporte público e transporte público fretado. Para os testes dos modelos foram geradas as matrizes de confusão, com a acurácia e o índice Kappa; para escolha dos modelos de regressão multinomial, também foram obtidos os pseudos R^2 de McFadden, Cox & Snell e Nagelkerke, e o critério de informação Akaike (AIC). Os modelos apresentaram boas medidas de desempenho para as duas técnicas, obtendo para o Polo Universitário valores de acurácias para a amostra teste de 73,75% e 67,50% para a árvore de decisão e regressão multinomial, respectivamente. Os resultados apontaram que fatores como tempo de deslocamento, distância percorrida, idade, gênero, renda familiar e cidade onde reside, são significativos para a escolha do modo de transporte dos estudantes do Polo Universitário em estudo. A melhor compreensão dos fatores que influenciam a escolha modal pode fornecer subsídios para os órgãos públicos, auxiliando na tomada de decisões que envolvam o planejamento e investimento no sistema viário e de transportes para a área de estudo, e contribuir para a busca de políticas que incentivem a escolha por modos de transporte mais sustentáveis.

Palavras-chave: Regressão logística. Árvore de decisão. Escolha do modo de transporte. Estudantes universitários.

ABSTRACT

The search for sustainable urban mobility is essential to provide quality of life in cities. Higher education institutions demand a larger number of daily trips concentrated in time and space, and if they are predominantly by car, they generate more impacts on the road system, causing congestion. Therefore, it is extremely important to understand the factors that lead university students to choose a particular means of transport. The incentive for students to exchange the use of the car for public or non-motorized means of transportation is essential in metropolitan regions. For this reason, it is necessary to initially understand the way of transportation of students, so that later they can propose alternatives for more sustainable modes. In view of this, the objective of this research is to determine the main factors that influence the decision for choosing the mode of transportation of students from a university center, located in the city of Joinville, using the techniques of multinomial logistic regression and decision tree. The specific objectives are to estimate a statistical model of multinomial logistic regression that describes the study variable; estimate a model of decision tree for choosing the mode of transport among university students; compare the results found by the two methods with each other; and to compare the results obtained by the two methods adopted in the present work with the literature. The study area in this research refers to two institutions of higher education, one public and the other private, located in the northern area in the municipality of Joinville, which together form a university center. The database used to estimate the models is a questionnaire developed as part of Schubert's master's dissertation (2019), applied to undergraduate and graduate students in 2018. For the statistical analysis, the following modes of transport: car, active mode (on foot and bicycle), public and chartered public transport. For model tests, confusion matrices were generated, with accuracy and Kappa index; for choosing multinomial regression models, McFadden, Cox & Snell and Nagelkerke pseudo R^2 were also obtained, as well as the Akaike information criterion (AIC). The models presented good performance measures for both techniques, obtaining accuracy values for the University Pole for the test sample of 73.75% and 67.50% for the decision tree and multinomial regression, respectively. The results showed that factors such as travel time, distance traveled, age, gender, family income and city where they live, are significant for the choice of transportation mode for students at the University Pole under study. A better understanding of the factors that influence the modal choice, can provide subsidies for public agencies assisting in making decisions that involve planning and investing in the road and transport system for the study area, and contribute to the search for policies that encourage the choice for more sustainable modes of transport.

Keywords: Logistic regression. Decision tree. Choice of mode of transport. University students.

LISTA DE FIGURAS

Figura 1 - Congestionamento na Rua Tenente Antônio João, uma das principais vias de acesso ao polo universitário.	15
Figura 2 - Exemplo de uma árvore de decisão	28
Figura 3 - Gráfico de dispersão para variável dependente com apenas duas categorias.	37
Figura 4 - Curva S de probabilidade para a variável dependente binária.	37
Figura 5 - Evolução dos passageiros de ônibus em Joinville	44
Figura 6 - Evolução dos passageiros do ônibus em relação a população em Joinville	45
Figura 7 - Localização das instituições de ensino superior que contemplam a pesquisa	45
Figura 8 - Esquema representativo do procedimento de pesquisa.....	47
Figura 9 - Esquema da árvore de decisão para escolha do modo de transporte	49
Figura 10 - Esquema representativos dos métodos utilizados para a modelagem de escolha modal	50
Figura 11 - Gênero de acordo com a Universidade	54
Figura 12 - Renda familiar de acordo com gênero e Universidade.....	55
Figura 13 - Idade de acordo com gênero e Universidade.....	56
Figura 14 - Distância percorrida de acordo com a Universidade	57
Figura 15 - Árvore de decisão para o Polo Universitário	60
Figura 16 - Árvore de decisão para o Polo Universitário após a etapa de poda.....	61
Figura 17 - Árvore de decisão para a Universidade A	64
Figura 18 - Árvore de decisão para a Universidade B	66
Figura 19 - Árvore de decisão para a Universidade B após a etapa de poda	67
Figura 20 - Árvore de decisão para o Polo Universitário para outras categorias.....	83

LISTA DE TABELAS

Tabela 1 - Fatores encontrados na revisão da literatura.....	26
Tabela 2 - Categorização e frequência das variáveis presentes na pesquisa	53
Tabela 3 - Característica da amostra por modo de transporte	58
Tabela 4 - Característica da amostra por modo de transporte	63
Tabela 5 - Medidas de desempenho do modelo para o Polo Universitário	63
Tabela 6 - Matriz de confusão do modelo para a Universidade A.....	65
Tabela 7 - Medidas de desempenho do modelo para a Universidade A.....	65
Tabela 8 - Matriz de confusão do modelo para a Universidade B.....	69
Tabela 9 - Medidas de desempenho do modelo para a Universidade B.....	69
Tabela 10 - Descrição das variáveis utilizadas nos modelos	70
Tabela 11 - Comparação dos modelos para o Polo Universitário	71
Tabela 12 - Medidas de avaliação dos modelos para o Polo Universitário	73
Tabela 13 - Matriz de confusão do Modelo 2 para o Polo Universitário	73
Tabela 14 - Comparação dos modelos para a Universidade A.....	74
Tabela 15 - Medidas de avaliação do modelo para a Universidade A	74
Tabela 16 - Matriz de confusão do Modelo 3 para a Universidade A.....	75
Tabela 17 - Comparação dos modelos para a Universidade B.....	75
Tabela 18 - Medidas de avaliação dos modelos para a Universidade B.....	77
Tabela 19 - Matriz de confusão do Modelo 6 para a Universidade B.....	77
Tabela 20 - Razão de Chance para os modelos escolhidos	78
Tabela 21 - Medidas de desempenho das técnicas abordadas	79
Tabela 22 - Variáveis significativas encontradas em cada técnica abordada na pesquisa.....	80
Tabela 23 - Variáveis influentes na escolha modal encontradas na pesquisa e na literatura	81
Tabela 24 - Descrição das variáveis utilizadas nos modelos	82
Tabela 25 - Matriz de confusão da amostra teste	83
Tabela 26 - Medidas de desempenho da amostra teste	84
Tabela 27 - Modelo para o Polo Universitário	84
Tabela 28 - Medidas de avaliação do modelo para Polo Universitário	85
Tabela 29 - Razão de Chance para o modelo escolhido	85

SUMÁRIO

1	INTRODUÇÃO.....	13
1.1	JUSTIFICATIVA	14
1.2	OBJETIVOS	16
1.2.1	Geral	16
1.2.2	Específicos	16
2	REVISÃO DA LITERATURA	17
2.1	MOBILIDADE URBANA	17
2.2	ESCOLHA DO MODO DE TRANSPORTE	20
2.2.1	Estudos envolvendo a escolha do modo de transporte	22
2.2.2	Fatores influentes na escolha do modo de transporte por estudantes universitários	25
2.3	REVISÃO DOS MÉTODOS.....	26
2.3.1	Árvore de decisão	27
2.3.1.1	Algoritmos de indução de árvores de decisão	32
2.3.1.2	Avaliação do desempenho do modelo.....	34
2.3.2	Regressão logística.....	36
2.3.2.1	Regressão logística simples	37
2.3.2.2	Regressão logística múltipla.....	38
2.3.2.3	Regressão logística multinomial	38
2.3.2.4	Estimativa do modelo de regressão logística	39
2.3.2.5	Testes de significância	40
2.3.2.6	Medidas de avaliação e ajuste do modelo.....	41
3	METODOLOGIA.....	43
3.1	ÁREA DE ESTUDO	43
3.2	BANCO DE DADOS	46
3.3	MÉTODO.....	46
4	RESULTADOS E DISCUSSÕES	51
4.1	CARACTERIZAÇÃO DA AMOSTRA.....	51
4.2	RESULTADOS DAS ÁRVORES DE DECISÃO	59
4.3	RESULTADOS DOS MODELOS DE REGRESSÃO MULTINOMIAL.....	69

4.4	COMPARAÇÃO DOS RESULTADOS DA ÁRVORE DE DECISÃO E DA REGRESSÃO MULTINOMIAL.....	79
4.5	COMPARAÇÃO DOS RESULTADOS DOS MÉTODOS UTILIZADOS NA PESQUISA COM A LITERATURA	80
4.6	MODELOS OBTIDOS CONSIDERANDO APENAS AS CATEGORIAS TRANSPORTE PÚBLICO FRETADO, CARONA E MOTO	81
5	CONCLUSÕES	86
	REFERÊNCIAS.....	89
	APÊNDICE A - TESTES T E WILCOXON REALIZADOS ENTRE OS MODOS DE TRANSPORTE	98
	APÊNDICE B – ÁRVORE DE DECISÃO NO R.....	103
	APÊNDICE C – REGRESSÃO MULTINOMIAL NO R.....	104
	APÊNDICE D – REGRESSÃO MULTINOMIAL MODELO 1.....	106
	APÊNDICE E – REGRESSÃO MULTINOMIAL MODELO 2.....	107
	APÊNDICE F – REGRESSÃO MULTINOMIAL MODELO 3	108
	APÊNDICE G – REGRESSÃO MULTINOMIAL MODELO 4.....	109
	APÊNDICE H – REGRESSÃO MULTINOMIAL MODELO 5.....	110
	APÊNDICE I – REGRESSÃO MULTINOMIAL MODELO 6	111
	APÊNDICE J – REGRESSÃO MULTINOMIAL MODELO 7	112

1 INTRODUÇÃO

Diariamente, nas cidades, ocorre o deslocamento de pessoas para atenderem às diferentes necessidades da sociedade. Para que este deslocamento ocorra, é necessário que o indivíduo faça uso de algum modo de transporte, seja ele individual motorizado, como carro e moto, coletivo, como o ônibus, ou ativo, como bicicleta ou caminhada.

Com o crescimento das áreas urbanas, o número de viagens diárias aumentou, ocasionando uma maior quantidade de transportes motorizados, principalmente os carros, transitando nas vias urbanas, gerando congestionamentos e dificultando o deslocamento das pessoas. Por isto, a mobilidade urbana é um tema em constante discussão, pois com o aumento do uso do automóvel particular, verifica-se a necessidade da adoção de estratégias de incentivo ao uso de transportes mais sustentáveis.

As estratégias de gestão da mobilidade urbana inseridas em políticas públicas influenciam na escolha do usuário por determinado modo de transporte em seus deslocamentos diários. Diante disto, os órgãos públicos têm como um de seus principais interesses a busca pela mobilidade urbana sustentável, já que, em cidades de todo o planeta, os sistemas de transportes e uso do solo apresentam sinais de insustentabilidade, como poluição ambiental, acidentes de trânsito, congestionamentos, exclusão social, altos custos de viagens, consumo de recursos não renováveis, entre outros (LOPES, 2010).

Os modos de transporte atuais devem fornecer sustentabilidade para a população urbana e corresponder à necessidade de mudança das pessoas. Para isto, é necessário que os planejadores considerem a demanda e os hábitos de transporte das pessoas que vivem nas cidades, para investimentos efetivos no transporte. Por essas razões, os tomadores de decisão desenvolvem planos e modelos de transporte (ORAL; TECIM, 2013).

Neste contexto, a presente pesquisa propõe identificar quais os fatores que influenciam a escolha do modo de transporte entre os estudantes de ensino superior para o seu deslocamento, até a instituição de ensino. Para o desenvolvimento deste estudo foi escolhido o município de Joinville, localizado no norte do estado de Santa Catarina, considerado o maior polo industrial do estado. Joinville possui uma

população estimada em cerca de 590 mil habitantes, sendo assim, a cidade mais populosa de Santa Catarina (IBGE, 2019).

Atualmente, o carro é o modo de transporte mais utilizado em Joinville, sendo que, em 17 anos, a quantidade de automóveis quase triplicou e a das motocicletas quadruplicou (SEPUD, 2018). Com isto, os congestionamentos tornaram-se comuns na maioria das regiões da cidade nos horários de pico, interferindo diretamente na qualidade de vida da população.

Diante disto, esta pesquisa teve como enfoque os Campi Universitários, tentando compreender quais são os fatores determinantes na escolha do modo de transporte entre os estudantes. Assim, os resultados obtidos podem servir de auxílio aos planos de mobilidade para reduzir o uso do automóvel particular, com o intuito de diminuir os congestionamentos, na busca pelo incentivo ao uso de modos de transporte mais sustentáveis.

1.1 JUSTIFICATIVA

A mobilidade urbana é atualmente tema de estudo interdisciplinar, pois diversas áreas buscam compreender como se dá a escolha pelos diferentes modos de transporte e, assim, construir estratégias interdisciplinares que viabilizem a melhora da mobilidade (BATTISTON; DEBATIN NETO; OLEKSZECHEN, 2017).

A busca pela mobilidade urbana sustentável parte inicialmente em entender quais os fatores que influenciam na decisão pela escolha do modo de transporte. A criação de modelos objetivos e analíticos para determinação dos impactos sobre os padrões de escolha modal, apresenta-se como um fator que poderá tornar mais eficientes as decisões políticas sobre investimentos nos sistemas viários e de transporte (GONÇALVES, 2012).

O sistema viário, no qual estão inseridas instituições de ensino de superior, tende a ser impactado, de forma negativa, pelas viagens diárias geradas por essas instituições (OLIVEIRA, 2015). Sendo assim, as universidades podem ser consideradas como polos geradores de viagens, pois, diariamente, em todo início e final do turno escolar, ocorre um grande número de deslocamento de pessoas, podendo ocasionar congestionamento em suas vias de acesso. Por isso, compreender padrões de deslocamentos dos estudantes universitários é de suma importância para mitigar os efeitos negativos das viagens produzidas, com o intuito de

promover a mobilidade sustentável e eficiente de estudantes, funcionários e visitantes (FERREIRA; PITOMBO; CUNHA, 2016).

Assim, com a presente pesquisa, pretende-se contribuir para identificação e análise dos fatores relevantes que influenciam na escolha do modo de transporte de estudantes universitários, com o intuito de auxiliar nas tomadas de decisões do planejamento do sistema viário, além de contribuir para o desenvolvimento de políticas públicas que foquem no desenvolvimento sustentável dos municípios.

É importante salientar que, nesta dissertação, foram utilizados dois modelos para identificação dos fatores significativos que influenciam na escolha dos estudantes universitários do polo em estudo por determinado modo de transporte: a regressão logística multinomial e a árvore de decisão, posteriormente fez-se a comparação dos resultados e obteve-se respostas que tentem descrever a realidade atual do objeto de pesquisa.

A região de estudo do trabalho foi um polo universitário formado por duas instituições de ensino, uma universidade pública e outra privada. Este polo gera um grande deslocamento de pessoas, ocasionando congestionamentos em todas as suas vias de acesso e ao entorno, conforme pode ser observado na Figura 1. Sendo assim, é de relevante importância pesquisas que busquem avaliar e entender como ocorre o acesso ao polo, principalmente pelos estudantes, que geram a maior parte dos deslocamentos.

Figura 1 - Congestionamento na Rua Tenente Antônio João, uma das principais vias de acesso ao polo universitário.



Fonte: NSC Total (2019).

Portanto, as informações obtidas com o uso de modelos para previsão da escolha do modo de transporte de estudantes universitários podem ser utilizadas nas decisões políticas de investimentos nos sistemas de transportes, de modo que possam satisfazer o deslocamento do público alvo da presente pesquisa, de maneira sustentável. Além disto, a metodologia utilizada pode servir de base para novas pesquisas em outros polos similares.

1.2 OBJETIVOS

1.2.1 Geral

O presente trabalho tem como objetivo determinar os principais fatores que influenciam na decisão pela escolha do modo de transporte de estudantes em um polo universitário localizado no município de Joinville, utilizando as técnicas de regressão logística multinomial e árvore de decisão.

1.2.2 Específicos

- Estimar um modelo estatístico de regressão logística multinomial que descreva a variável de estudo;
- Estimar um modelo de árvore de decisão da escolha do modo de transporte entre estudantes universitários;
- Comparar os resultados encontrados pelos dois métodos entre si;
- Comparar os resultados obtidos pelos dois métodos adotados na presente pesquisa com a literatura.

2 REVISÃO DA LITERATURA

No presente capítulo será apresentada uma breve revisão sobre os conceitos e problemas que envolvem a mobilidade urbana e a escolha do modo de transporte, além de apresentar exemplos de trabalhos já realizados sobre o assunto. Também neste capítulo serão apresentados conceitos teóricos dos métodos de análise propostos na pesquisa.

2.1 MOBILIDADE URBANA

Na década de 70 houve um grande crescimento urbano nas cidades brasileiras, aumentando assim, a demanda por transportes. Devido a isto, buscou-se nos modelos técnicos-econômicos o apoio necessário para o planejamento de transportes urbanos (SILVA, 2010).

De acordo com Machado e Piccinini (2018), a mobilidade substituiu o termo transporte pelo fato de sugerir o deslocamento de pessoas englobando também os deslocamentos não motorizados, com intuito do desenvolvimento sustentável, tendência mundial desde a publicação do Relatório Brundtland, em 1987. O pensamento global e as atuações locais iniciaram um pensamento voltado para a sustentabilidade urbana.

Segundo Gonçalves (2012), na maioria das cidades de médio¹ e grande² porte ao redor do mundo ocorrem problemas de mobilidade, pois o desenvolvimento urbano provocou mudanças nestes hábitos, onde a maior parte das pessoas optou pelo uso do automóvel particular, resultando em aumento dos congestionamentos nas vias urbanas.

No Brasil, os investimentos insuficientes e inadequados e a ausência de políticas públicas que priorizem os modos de transporte coletivo e não motorizados, agravam os problemas de mobilidade urbana, comprometendo os deslocamentos diários da população, principalmente nas parcelas de menor renda, muitas vezes dependentes do transporte público e sem estrutura nas vias para a utilização segura de modos não motorizados (ABDALA; PASQUALETTO, 2013).

¹ Cidade com população entre 100 mil e 500 mil habitantes (IBGE, 2011).

² Cidade com mais de 500 mil habitantes (IBGE, 2011).

Para Teixeira *et al.* (2014), a evolução do transporte público influencia no desenvolvimento urbano, pois em diversas cidades do mundo possibilitou a expansão territorial urbana do centro para as periferias, facilitando os deslocamentos de casa para o trabalho, e vice-versa, de sua população.

Um fator que afeta a qualidade de vida de uma população é tempo de deslocamento de casa ao trabalho, na medida em que influencia na jornada cotidiana do trabalhador. O tempo de deslocamento de casa ao trabalho depende de fatores como a distância, o tipo de transporte utilizado e do congestionamento, sendo muito sensível com a dinâmica de oportunidades de emprego e de moradia. Por isto, é necessária a compreensão da situação da mobilidade urbana para formulação de políticas públicas que reduzam o tempo de viagem e seus custos, de acordo com as condições espaciais e socioeconômicas de cada região (PERO; STEFANELLI, 2015).

No Brasil, as políticas públicas destinadas à mobilidade urbana inicialmente foram voltadas ao transporte individual, sendo que a ocorrência de manifestações populares envolvendo o aumento das tarifas dos transportes urbanos, por exemplo, mostraram que questões relacionadas com a mobilidade urbana estão sendo mais notadas como um direito de cidadania e uma condição de equidade social. Isto faz com que a democratização dos modos de transportes e a busca por uma mobilidade sustentável seja uma necessidade crescente e aclamada (SILVA, 2013).

De acordo com Libardi (2014), com relação ao transporte público, existem diferenças entre os países desenvolvidos e em desenvolvimento, pois a boa qualidade do transporte público nos países desenvolvidos faz com que este seja uma escolha do usuário, enquanto nos países em desenvolvimento geralmente o transporte é de má qualidade, não sendo prioridade no planejamento urbano ou financiamento público.

Conforme Resende (2017), nos países em desenvolvimento, devido às profundas diferenças sociais, a fragilidade das instituições políticas e as contradições sociais e econômicas, faz-se necessário a realização de análises, planejamentos, políticas públicas, e que estas estejam articuladas com a mobilidade urbana de forma que contemplem todas as particularidades dos processos espaciais e sociais.

A dificuldade crescente de deslocamento de bens e pessoas, devido ao elevado número de veículos em circulação, acarretam problemas como os acidentes de trânsito, a poluição do ar e os congestionamentos, onde os transportes públicos

possuem seu espaço cada vez mais reduzido e sua operação dificultada, gerando efeitos sobre a qualidade do serviço (PASSOS *et al.*, 2012).

Para Libardi (2014), o crescimento da motorização individual é o problema mais sério da mobilidade urbana atualmente, impactando o crescimento das cidades mais do que qualquer inovação tecnológica. O uso excessivo dos modos motorizados gera congestionamentos, ocasionando o consumo excessivo de energia e de espaço, onde a poluição do ar e o estresse no trânsito provocam problemas de saúde (STEIN, 2013).

Várias propostas políticas têm sido apresentadas para atenuar as consequências provocadas pelo uso do automóvel. Para melhorar as condições do deslocamento de pessoas e cargas, surgiu o conceito de gerenciamento da mobilidade, para tratar os problemas e proporcionar aos usuários de transportes alternativos mais sustentáveis (STEIN, 2013).

O gerenciamento da mobilidade deve criar estratégias com ferramentas baseadas na organização, promoção e informação, sendo que as estratégias desenvolvidas, as políticas, sua aplicação, e os objetivos são diversos, e dependem da forma de abordagem e das necessidades das cidades em estudo (PARRA, 2006).

De acordo com Rubim e Leitão (2013), para a construção de cidades diferentes das que temos atualmente, mais inclusivas e socialmente justas, são necessárias a adoção de duas ações: a melhoria do transporte público coletivo e a revisão dos benefícios concedidos ao usuário do transporte individual.

Assim, conforme Battiston, Debatin Neto e Olekszechen (2017), o deslocamento de pessoas ou de cargas em espaços urbanos deve promover o desenvolvimento sustentável pela priorização dos modos de transporte não motorizados sobre os motorizados.

Segundo Abdala e Pasqualetto (2013), o uso do transporte público ou coletivo de qualidade para o deslocamento da população é um fator influente na qualidade de vida urbana, direcionando seu desenvolvimento econômico e social. Nesta busca pela melhoria da qualidade de vida e pela sustentabilidade é essencial compreender como ocorre a acessibilidade espacial nas cidades, que é definida pela alocação de atividades e pelo modo de deslocamento da população (HARRIS, 2015).

Para Battiston, Debatin Neto e Olekszechen (2017, p. 270), “entende-se que buscar formas alternativas de transporte é, mais do que uma opção para a mobilidade, uma necessidade, considerando a atual conjuntura do sistema de trânsito”. A busca pela mobilidade urbana sustentável ideal é algo difícil de alcançar devido a adaptação

ineficaz do meio urbano aos modos suaves de deslocamento, o que, aliado à baixa qualidade do transporte público, resulta na opção modal pelos veículos individuais (SILVEIRA, 2016).

Outra alternativa de modo de transporte sustentável é a bicicleta. O incentivo ao uso de bicicletas pode ser considerado como fundamental na construção da mobilidade sustentável, reduzindo o custo de deslocamentos das pessoas e da degradação ambiental. Sua integração aos sistemas de transporte coletivos é possível e já tem acontecido em grandes cidades (LARGURA, 2012).

Em países como Holanda, Suíça e Alemanha, além de parte de países escandinavos e países asiáticos como China e Índia, a bicicleta é utilizada como principal meio de transporte, mostrando-se favorável no deslocamento em trajetos de curta distância, para realização de diversas atividades e carregamento de produtos de pequeno porte (LARGURA, 2012).

A utilização de bicicletas reduz o nível de ruído nos sistemas viários e demanda de um espaço urbano de menor circulação em comparação aos modos motorizados, gerando menos custos para a cidade e proporcionando aos usuários uma melhor qualidade de vida, incentivando a promoção da saúde e tornando o deslocamento mais calmo, agradável e limpo (SILVA, 2012).

Segundo Lindner (2015), as variáveis essenciais para o planejamento e previsão da demanda por transportes são o número de viagens produzidas e atraídas, o fluxo de tráfego e a escolha modal que é influenciada diretamente pelos fatores humanos comportamentais.

Para Ramis e Santos (2012), não só o crescimento das cidades influenciou para o aumento no uso de automóveis, também há a ocorrência de questões comportamentais e de mudanças de hábitos enraizados na cultura. Somente por meio de políticas públicas de incentivo ao uso do transporte coletivo, esclarecendo e mostrando suas vantagens, é possível atrair novos adeptos à modalidade de transporte (RAMIS; SANTOS, 2012).

2.2 ESCOLHA DO MODO DE TRANSPORTE

A escolha modal do transporte pode ser avaliada pelas condições socioeconômicas e a acessibilidade dos habitantes a diferentes modos de transporte. Essa escolha também é influenciada por diversos fatores como a distância, o custo

da viagem e a disponibilidade ou não de determinado modo de transporte (SILVA, 2010).

Segundo Nutsugbodo, Amenumey e Mensah (2018), variáveis sociais, econômicas, culturais e ambientais possuem influência significativa na escolha do modo de transporte. A distância de viagem e outras características como segurança, conveniência e acessibilidade são consideradas fatores influentes nas preferências de escolha do modo de transporte. O tempo de viagem, custo, tempo de espera e conforto também podem influenciar nas escolhas do modo de transporte.

De acordo com Haddad e Vieira (2015), análises realizadas a partir de estudos sobre escolhas modais sugerem que o tempo economizado em deslocamentos dos trabalhadores representa um valor percentual de seu salário-hora, com diferentes valores dependendo dos fatores relacionados aos tipos e contextos das viagens analisadas e da metodologia utilizada.

A modelagem tradicional para estimativas da escolha do modal é geralmente baseada nos modelos econométricos, mas em muitos estudos, além de técnicas econométricas, vêm sendo aplicados modelos não paramétricos e exploratórios (PITOMBO; COSTA, 2015).

Segundo Lindner (2015), existem estudos no qual a variável em análise é discreta ao invés de contínua, onde o fenômeno estudado é resultado de uma seleção entre diferentes alternativas, sendo mais conveniente o uso de modelos de escolha discreta. Para Lindner (2015, p. 34), os modelos de escolha discreta podem ser de três tipos:

- Categóricos: a variável dependente não tem significado por si só, representa apenas uma codificação; ex: “1”, se o indivíduo trabalha e “0”, caso contrário.
- Com resposta ordenada: a variável dependente reflete um conceito por trás de sua ordem; ex: “1” muito satisfeito, “2” satisfeito, “3” insatisfeito.
- Contáveis: a variável dependente tem significado por si só; ex: número de patentes registradas de um país.

Na modelagem da divisão modal, a especificação do modelo a ser aplicado é somente a primeira etapa do processo, porém devem-se proceder inicialmente as etapas de calibração e testes de verificação do desempenho, pois, somente assim, verifica-se a validade do modelo para realizar análises e estimar projeções futuras (SOUZA, 2016).

Uma vez que o banco de dados tenha sido tratado e o modelo devidamente calibrado, procede-se então à última (e não menos importante) etapa da modelagem da divisão modal, a etapa de testes para verificação de seu desempenho e comportamento (SOUZA, 2016, p. 35).

Para previsão de escolhas discretas de modo de transporte, um dos modelos de escolha discreta mais utilizada na demanda por transportes é o logístico (ORTÚZAR; WILLUMSEN, 2011).

Segundo Lindner (2015), também deve ser destacada a Função Utilidade, que é uma expressão matemática para determinação do grau de satisfação que o usuário obtém através da escolha do modo de transporte e é definida pela Equação (1).

$$U = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n + \varepsilon \quad (1)$$

onde U é a utilidade derivada da escolha medida pelos atributos, x_i são as variáveis de transporte, a_i são os pesos relativos às variáveis e ε é o erro aleatório que segue a distribuição de Weibull.

De acordo com Pitombo, Costa e Salgueiro (2015), é possível estimar a escolha modal considerando variáveis socioeconômicas e medidas de qualidade do sistema de transporte. A aplicação da técnica de Árvore de Decisão pode ser uma abordagem alternativa aos modelos econométricos tradicionais, porque não têm restrições relacionadas às variáveis de entrada (categórica, numérica ou dummy) e não tem suposições como normalidade, linearidade e multicolinearidade.

2.2.1 Estudos envolvendo a escolha do modo de transporte

O desenvolvimento de pesquisas relacionadas com o comportamento de escolha do modo de transporte dos usuários é de grande relevância no planejamento urbano, focado no desenvolvimento sustentável. A seguir, são apresentados trabalhos realizados indicando os fatores influentes na escolha do modo de transporte com foco em estudantes, principalmente universitários, que são o objeto de estudo da presente pesquisa.

A pesquisa de Zhang, Yao e Liu (2017) explora o comportamento de escolha do modo de viagem escolar de estudantes de 7 a 18 anos na cidade de Pequim, na China. Para examinar a influência das variáveis na escolha do modo de viagem

escolar foram feitas análises utilizando os modelos de regressão logística e árvore de decisão. Os resultados mostraram que possuir carro, o ambiente de caminhada/ciclismo ruim e a conveniência dos adultos em levar os estudantes estimula significativamente o uso do carro. A escolha pelo uso do carro é ainda mais favorável quando o horário de partida é no horário de congestionamento em comparação com outros horários. Além disso, a maior distância incentiva o uso de transporte motorizado, estando os pais mais dispostos a levar seus filhos para a escola. Também, verificou-se que residir no bairro onde a escola está localizada pode reduzir a probabilidade do uso o carro. Para os estudantes que vivem num raio de 6 km da escola, a variável gênero é altamente significativa para escolha por bicicleta, transporte público e caminhada, indicando que modos ativos e transporte público são mais utilizados por rapazes. Já para os estudantes que moram acima de 6 km da escola, a variável gênero não apresentou correlação com a escolha de viagem escolar dos alunos.

No trabalho desenvolvido por Nguyen-Phuoc *et al.* (2018), com os dados coletados de mais de 500 estudantes em seis universidades em Danang, no Vietnã, um modelo de regressão logística foi criado para explorar variáveis específicas individuais e alternativas que influenciam a escolha de modo para viagens à escola. As principais descobertas mostraram que características dos alunos como idade, gênero e renda têm um impacto na decisão de escolha de modo de transporte. Tempo de viagem de casa para a escola é encontrado com um efeito fortemente negativo sobre a escolha de caminhar como meio de locomoção para a escola. Os estudantes que viajam menos de 2 km estão mais propensos a utilizar modos de transportes ativos. Verificou-se também que, alunos que estão usando motocicletas para as escolas estão dispostos a mudar para o transporte público, se um sistema de transporte público eficiente e confiável estiver disponível.

No trabalho de Ferreira, Pitombo e Cunha (2016), foi utilizado o modelo de regressão logística binomial para previsão da escolha modal de um *campus* universitário em São Carlos, no estado de São Paulo. Os dados analisados fazem parte de um banco de dados de um questionário aplicado a estudantes de graduação, pós-graduação, docentes e demais funcionários. A análise indicou que usuários com idade superior a 30 anos estão menos propensos a irem caminhando para a universidade. Com relação ao uso do carro, notou-se que usuários mais jovens, com idade até 30 anos, possuíam menor tendência à utilização deste modo. Além disso,

percebeu-se que o acesso a meios motorizados interferem na escolha do modo de transporte para o deslocamento até a universidade.

Na pesquisa realizada por Pitombo e Gomes (2014), foi observado que, quanto maior a renda familiar ou o número de carros no domicílio, maior a probabilidade de escolher o carro como modo de deslocamento.

No estudo sobre o serviço de transporte público que alimenta o campus universitário da Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUÍ), localizado na cidade de Ijuí, desenvolvido por Specht, Brandli e Silva (2009), a partir de uma coleta de dados com aplicação do método de preferência declarada, foram gerados modelos do tipo função utilidade para identificação da percepção dos usuários com relação ao transporte público, obtendo-se como atributo mais importante a acessibilidade, seguido da frequência, segurança e custo.

Na Universidade Federal do Ceará, Maia *et al.* (2012), procuraram identificar as características mais relevantes nos deslocamentos internos, e os atributos significativos foram o atendimento e o conforto, seguidos pelo tempo de viagem.

Sanches e Ferreira (2013) verificaram que o modo de transporte mais utilizado para acesso ao campus da Universidade Federal de São Carlos é o automóvel, sendo que o conforto e a rapidez são os fatores que mais influenciam na escolha deste. Os incentivos mais citados para o uso do ônibus foram relacionados à frequência de serviço e à existência de linha direta. Dentre os fatores que fariam o usuário de automóvel passar a usar a bicicleta, destaca-se a existência de ciclovias e uma menor distância a ser percorrida.

Em Daca, capital de Bangladesh, Nasrin (2017) concluiu que custo, conforto, tempo e disponibilidade do modo de transporte, são os principais fatores para escolha do modo de transporte entre os estudantes na Universidade *Asia Pacific*. Estes resultados foram revelados através de uma análise exploratória e criação de um modelo logístico multinomial, com base em dados coletados com os estudantes da universidade em estudo.

No trabalho de Kotoula *et al.* (2018), analisou-se características dos padrões de viagem de estudantes da Universidade Democrática da Trácia, localizada na cidade de Xanthi, Grécia, e desenvolveram-se modelos logísticos multinomiais, encontrando no modo de escolha, tempo e distância como fatores mais importantes. Com relação ao transporte público, fatores como economia e segurança influenciaram na escolha por este modo.

No estudo desenvolvido por Whalen, Páez e Carrasco (2013), utilizou-se um modelo logístico multinomial para analisar a escolha do modo de transporte de estudantes da Universidade McMaster, em Hamilton, Canadá. Os resultados indicaram que custo, atitudes individuais (variável formada por várias condições estabelecidas no estudo, como conforto no uso do modo de transporte, gosto por viajar sozinho, condições de infraestrutura, entre outros), densidade de ruas e calçadas influenciam na escolha do modo. O tempo de viagem afetou os usuários de carro e bicicleta, reduzindo o uso destes com o aumento no tempo.

Romanowska, Okraszewska e Jamroz (2019) investigaram os padrões de deslocamento diário e determinaram os fatores que afetam a escolha do modo de transporte de duas universidades localizadas em Gdansk, Polônia. A pesquisa abrangeu funcionários e estudantes, apresentando notórias diferenças no modo de escolha dos mesmos. Os resultados indicaram que 57% dos funcionários escolhem o carro como modo de transporte, e 59% dos estudantes optam pelo transporte público. A escolha do modo de viagem foi determinada principalmente pela disponibilidade do carro, localização da origem da viagem (tempo e distância) e acessibilidade ao transporte público. Outros fatores também apareceram, como custo e conforto. Com relação ao gênero, as mulheres optaram pelo transporte público com mais frequência do que os homens, enquanto que os homens escolheram mais o carro e a bicicleta. A pesquisa ainda indicou que andar a pé está relacionado com residir próximo à universidade.

Portanto, de acordo com as pesquisas apresentadas, pode-se concluir que diversos fatores podem influenciar na escolha do modo de transporte pelos estudantes, e que estes fatores podem ser derivados tanto do perfil ou características socioeconômicas do estudante, quanto das características do ambiente urbano.

2.2.2 Fatores influentes na escolha do modo de transporte por estudantes universitários

Conforme mencionado anteriormente, a escolha do modo de transporte está associada a fatores socioeconômicos e também da disponibilidade dos meios. Na Tabela 1 são apresentados os fatores relevantes encontrados na revisão da literatura, para a escolha do modo de transporte dos estudantes universitários.

Tabela 1 - Fatores encontrados na revisão da literatura

Fatores	Estudos em que aparecem
Tempo de viagem	Nguyen-Phuoc <i>et al.</i> (2018); Maia <i>et al.</i> (2012); Sanches e Ferreira (2013); Nasrin (2017); Kotoula <i>et al.</i> (2018); Whalen, Páez e Carrasco (2013); Romanowska, Okraszewska e Jamroz (2019)
Distância percorrida	Nguyen-Phuoc <i>et al.</i> (2018); Sanches e Ferreira (2013); Kotoula <i>et al.</i> (2018); Romanowska, Okraszewska e Jamroz (2019)
Idade	Nguyen-Phuoc <i>et al.</i> (2018); Ferreira, Pitombo e Cunha (2016)
Renda familiar	Nguyen-Phuoc <i>et al.</i> (2018); Pitombo e Gomes (2014)
Gênero	Nguyen-Phuoc <i>et al.</i> (2018); Romanowska, Okraszewska e Jamroz (2019)
Acessibilidade ao modo	Nguyen-Phuoc <i>et al.</i> (2018); Ferreira, Pitombo e Cunha (2016); Pitombo e Gomes (2014); Specht, Brandli e Silva (2009); Nasrin (2017); Romanowska, Okraszewska e Jamroz (2019)
Custo	Specht, Brandli e Silva (2009); Nasrin (2017); Kotoula <i>et al.</i> (2018); Whalen, Páez e Carrasco (2013); Romanowska, Okraszewska e Jamroz (2019)
Conforto	Maia <i>et al.</i> (2012); Sanches e Ferreira (2013); Nasrin (2017); Whalen, Páez e Carrasco (2013); Romanowska, Okraszewska e Jamroz (2019)
Infraestrutura	Sanches e Ferreira (2013); Whalen, Páez e Carrasco (2013)
Segurança	Specht, Brandli e Silva (2009); Kotoula <i>et al.</i> (2018)

Fatores como o tempo de viagem, a distância percorrida, a acessibilidade ao modo, o custo e o conforto, apareceram em vários estudos envolvendo a escolha do modo de transporte de estudantes universitários. Outros fatores como idade, sexo e renda familiar apareceram como variáveis socioeconômicas relevantes.

2.3 REVISÃO DOS MÉTODOS

Existem vários métodos para análise da escolha do modo de transporte de usuários. Nesta pesquisa serão utilizados dois métodos: a árvore de decisão e a regressão logística multinomial. Nesta seção apresenta-se o embasamento teórico destes métodos.

2.3.1 Árvore de decisão

As árvores de decisão representam de maneira simples e eficaz um determinado conjunto de dados, e consistem na divisão sucessiva do conjunto em vários subconjuntos, até os mesmos pertencerem a uma mesma classe, ou até uma das classes serem majoritárias, não sendo necessárias novas divisões (GARCIA, 2003).

Uma árvore de decisão pode ser definida como um gráfico acíclico e direto, onde a hierarquia é denominada árvore e cada segmento é denominado nó. Existe um nó, chamado raiz, que contém todo o banco de dados, e a partir deste nó os dados podem ser subdivididos dentro de sub-nós (nó filho). Entre os nós existe sempre um único caminho e, quando os dados não podem ser mais subdivididos dentro de outro subconjunto, ele é considerado um nó terminal ou folha (FERNANDES *et al.*, 2012).

A construção de uma árvore de decisão é realizada a partir de três elementos principais: um conjunto de atributos que delimita a divisão dos dados, um critério para estabelecer a melhor divisão dos nós e uma regra de parada para as subdivisões (ROMA *et al.*, 2018).

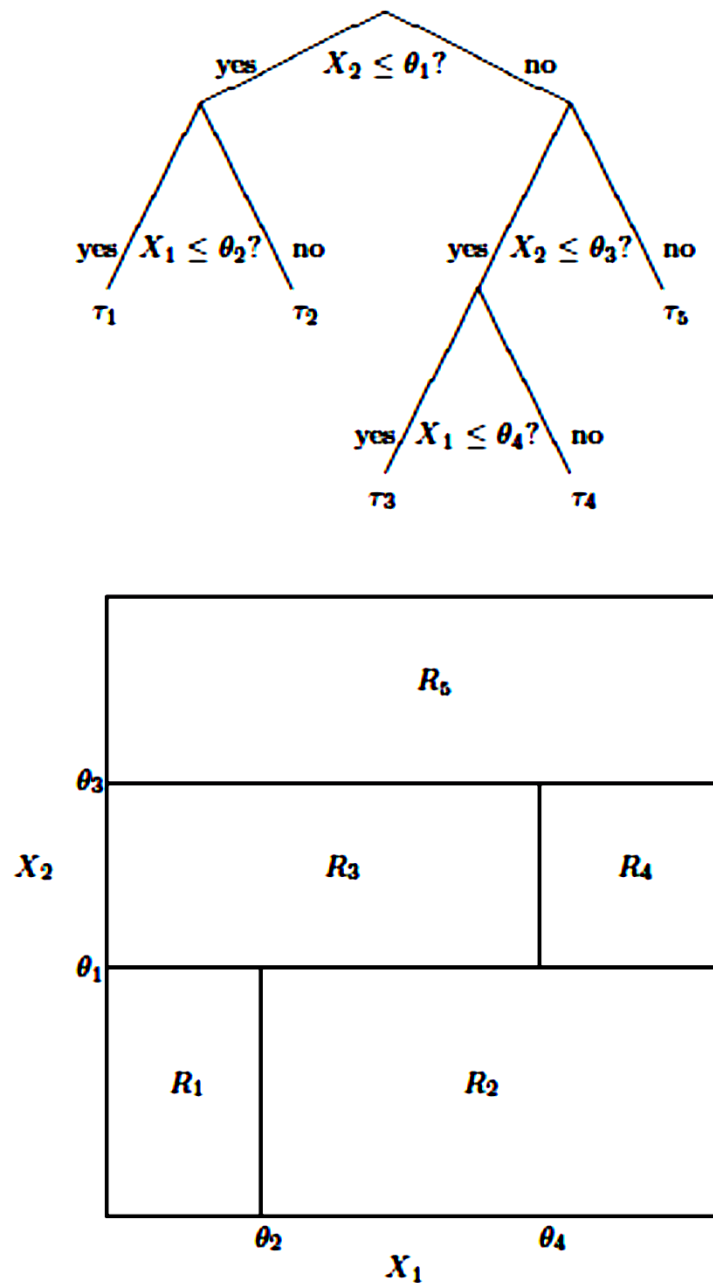
A Figura 2 apresenta um exemplo de árvore de decisão (figura superior) com duas variáveis de entrada X_1 e X_2 , quatro critérios de divisão dos nós $\theta_1, \theta_2, \theta_3$ e θ_4 , e cinco nós terminais T_1, T_2, T_3, T_4 e T_5 , correspondendo respectivamente as regiões R_1, R_2, R_3, R_4 e R_5 , identificadas na parte inferior da figura. A interpretação gráfica do exemplo de árvore de decisão é que, quando a condição do nó é satisfeita (por exemplo, $X_2 \leq \theta_1$), os dados seguem para o nó esquerdo (*yes*), e caso contrário, os dados seguem para o nó direito (*no*). Essa condição persiste até ser atingido um dos cinco nós terminais (T_1, T_2, T_3, T_4 ou T_5).

Caso a variável dependente seja qualitativa, as árvores de decisão são denominadas Árvores de Classificação e são utilizadas em problemas de classificação, que é o caso desta pesquisa. Se a variável dependente for quantitativa, as árvores de decisão são chamadas de Árvores de Regressão (PIANUCCI; PITOMBO, 2019).

Em um nó de uma árvore de decisão, os testes a serem realizados dependem das características do atributo vinculado ao nó. Em cada teste é apenas designado um atributo por nó, fazendo com que a estrutura da árvore seja de fácil compreensão. Durante os testes, é possível utilizar atributos do tipo quantitativo ou categórico.

Definidos os testes e o conjunto de dados, é decidido qual o caminho a ser percorrido na árvore durante o processo de classificação (GARCIA, 2003).

Figura 2 - Exemplo de uma árvore de decisão



Fonte: Izenman (2008, p.283).

As árvores de decisão são construídas utilizando um algoritmo de particionamento sequencial. Este algoritmo constrói a árvore dividindo sucessivamente os dados para criar subconjuntos homogêneos, resultando em

árvores de decisão úteis para previsão ou classificação de dados (PITOMBO; SOUZA; LINDNER, 2017).

De acordo com algoritmo utilizado para criação da árvore de decisão, é adotado um critério para seleção dos atributos mais representativos para a classificação da árvore. Desta forma, em uma árvore binária, para cada nó será selecionado um atributo que melhor represente a divisão.

Os critérios para a construção da árvore de decisão selecionam quais atributos representam melhor cada nó em análise. A partir do critério utilizado, é possível identificar a melhor divisão da árvore conforme a capacidade informativa do atributo.

Um dos critérios conhecidos para seleção de atributos é o ganho de informação, obtido através de uma medida conhecida como entropia. De acordo com Garcia (2003), o ganho é conceituado como a redução esperada da entropia, quando ocorre um particionamento de um conjunto de dados. Desta forma, é possível escolher o atributo que na divisão do nó permite obter um maior ganho de informação.

Num conjunto de dados, a entropia caracteriza a impureza deste, identificando assim, a falta de homogeneidade dos dados de entrada em relação a sua classificação. A interpretação da entropia na teoria da informação é que ela especifica o número mínimo de bits de informação necessários para codificar a classificação de um membro arbitrário de conjunto de dados, sendo a entropia definida pela Equação (2) (MITCHELL, 1997).

$$Entropia(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (2)$$

onde S é o conjunto de dados; c é o número de classes; e p_i é a proporção de S pertencente a classe i .

Definida a entropia como uma medida da impureza em um conjunto de dados, pode-se definir o ganho de informação como simplesmente a redução esperada na entropia, causada pelo particionamento dos dados de acordo com um atributo. Mais precisamente, o ganho de informação de um atributo relativo a um conjunto de dados, é definido pela Equação (3) (MITCHELL, 1997).

$$Ganho(S, A) = Entropia(S) - \sum_{i=1}^c \frac{|S_i|}{|S|} Entropia(S_i) \quad (3)$$

onde $Ganho(S, A)$ é o ganho do atributo A sobre conjunto S ; $|S_i|$ é o subconjunto de S no qual o atributo A tem valor i .

Assim, o atributo a ser escolhido para divisão da árvore, é o que apresentar o maior valor de ganho de informação. O cálculo do ganho de informação deve ser feito para todas as variáveis envolvidas na pesquisa, realizando as divisões na árvore até que os valores de ganho não apresentem diferenças significativas, isto é, os valores sejam homogêneos.

O ganho de informação tende a ser muito grande quando os atributos apresentam muitos valores, isto acaba resultando em árvores muito largas, que não são as mais desejáveis. Diante disto, a razão do ganho de informação é uma alternativa para contornar esse problema, pois se trata de um ganho relativo, definido pela Equação (4) (GARCIA, 2003).

$$RGanho(S, A) = \frac{Ganho(S, A)}{-\sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}} \quad (4)$$

onde $RGanho(S, A)$ é a razão de ganho do atributo A sobre conjunto S ; $|S_i|$ é o subconjunto resultante do particionamento de S pelos n valores do atributo A .

De acordo com Quinlan (1988), é sugerido que a razão de ganho seja calculada em duas etapas. Na primeira, calcula-se o ganho de informação de todos os atributos, e na segunda encontra-se a razão de ganho apenas para os atributos que apresentaram um ganho de informação acima da média. Por conseguinte, deve ser escolhido o atributo que apresentar a melhor razão de ganho.

O índice *Gini* também é utilizado para seleção de atributos. De acordo com Garcia (2003), para um conjunto de dados S contendo exemplo de n variáveis, $Gini(S)$ é definido através da Equação (5).

$$Gini(S) = 1 - \sum_{i=1}^n p_i^2 \quad (5)$$

onde p_i é a frequência relativa da classe i em S .

Particionando S em dois subconjuntos, S_1 e S_2 , com n_1 e n_2 exemplos respectivamente, o novo índice de divisão dos dados, $Gini_{split}(S)$, é dado pela Equação

(6), e o algoritmo seleciona atributos que valores maiores para *Gini* (GARCIA, 2003).

$$Gini_{split}(S) = \frac{n_1}{n} Gini(S_1) + \frac{n_2}{n} Gini(S_2) \quad (6)$$

Em um problema de duas classes, por exemplo, as probabilidades são p_1 e p_2 , assim, se um nó qualquer possui 100 dados, onde para uma variável explicativa os dados forem distribuídos em 40 e 60 para cada classe respectivamente, as probabilidades seriam:

$$p_1 = \frac{40}{100} = 0,4 \text{ e } p_2 = \frac{60}{100} = 0,6$$

Desta forma pode-se calcular o índice *Gini*, por meio da Equação (5), da seguinte maneira:

$$Gini = 1 - (p_1^2 + p_2^2) = 1 - (0,4^2 + 0,6^2) = 1 - (0,16 + 0,36) = 1 - 0,52 = 0,48$$

Caso o nó com 100 dados seja dividido em dois novos nós com 50 dados cada, pode-se calcular o novo índice *Gini* da divisão, por meio da Equação (6), supondo que as probabilidades dos dois nós sejam respectivamente, 0,3 e 0,7; e 0,4 e 0,6. Logo temos que:

$$\begin{aligned} Gini_{split} &= \frac{50}{100} [1 - (0,3^2 + 0,7^2)] + \frac{50}{100} [1 - (0,4^2 + 0,6^2)] \\ Gini_{split} &= \frac{50}{100} [1 - (0,09 + 0,49)] + \frac{50}{100} [1 - (0,16 + 0,36)] \\ Gini_{split} &= \frac{50}{100} [1 - 0,58] + \frac{50}{100} [1 - 0,52] \\ Gini_{split} &= \frac{50}{100} \cdot 0,42 + \frac{50}{100} \cdot 0,48 \\ Gini_{split} &= 0,21 + 0,24 = 0,45 \end{aligned}$$

O índice *Gini* é utilizado para medir a heterogeneidade dos dados de um nó, e seu valor varia de zero a um. Conforme o índice se aproxima de seu valor máximo, aumenta o número de classe uniformemente distribuídas no nó. Assim, é possível identificar o melhor atributo para divisão de cada nó da árvore de decisão, optando pelos atributos com os maiores valores. Esse procedimento deve ser realizado para

todas as variáveis envolvidas na pesquisa, realizando as divisões na árvore até que não exista mais ganho em dividi-la.

De acordo com Raimundo *et al.* (2008), no critério de *Gini* é medida a heterogeneidade dos dados e pode ser entendido como a probabilidade condicional de erro, dado um conjunto de treinamento selecionado de forma aleatória, testando todas as variáveis envolvidas na análise, escolhendo para a divisão do nó a que obteve a maior índice de *Gini*. O procedimento de divisão da árvore é feito até que não haja mais ganho na divisão, fazendo a associação de uma classe ao nó folha.

2.3.1.1 Algoritmos de indução de árvores de decisão

Existem diversas formas de uma árvore de decisão ser estruturada a partir de um conjunto de atributos, e o número de árvores possíveis cresce conforme o número de atributos aumenta. Assim, torna-se impraticável encontrar a melhor estrutura da árvore para um determinado estudo. Para solucionar este problema, são utilizados algoritmos para indução de árvores de decisão (BASGALUPP, 2010).

O algoritmo utilizado para divisão dos dados de acordo com a variável dependente identifica as variáveis independentes que maximizam a homogeneidade do subgrupo de dados de cada nó filho (PIANUCCI; PITOMBO, 2019). Um algoritmo bem conhecido e utilizado para indução de árvores de decisão é o *Top-Down Induction of Decision Tree* (TDIDT), sendo os mais usuais o ID3 (Quinlan, 1986), C4.5 (Quinlan, 1993) e CART (Breiman *et al.*, 1984). No TDIDT, as regras de decisão são produzidas de forma implícita por sucessivas divisões, de acordo com os valores dos atributos preditivos (BRAMER, 2007).

Desenvolvido por Quinlan (1986), o algoritmo ID3 constrói as árvores de decisão de cima para baixo, definindo inicialmente qual atributo estará na raiz da árvore. Para determinar este atributo, é utilizado como critério o ganho de informação; o melhor atributo é selecionado e usado como teste no nó raiz da árvore. Um descendente do nó raiz é então criado para cada valor possível deste atributo, e os exemplos de treinamento são classificados para o nó descendente apropriado (isto é, no ramo correspondente ao valor do exemplo para este atributo). Todo o processo é então repetido usando os exemplos de treinamento associados à cada nó descendente, a fim de selecionar o melhor atributo para testar nesse ponto na árvore (MITCHELL, 1997).

De acordo com Oral e Tecim (2013), o algoritmo ID.3 é usado para classificar valores categóricos com regras de divisão. Este algoritmo necessita que os valores dos atributos sejam tratados previamente, e a estrutura da árvore gerada é bastante simples (GARCIA, 2003).

Assim como o ID3, o algoritmo C4.5 também foi desenvolvido por Quinlan (1993), porém difere porque pode ser usado tanto para valores categóricos, como quantitativos. Para seleção dos atributos, é utilizado testes com o ganho de informação e a razão do ganho de informação.

O algoritmo C4.5 constrói árvores de decisão em quatro etapas: primeiro, seleciona-se o atributo do nó raiz; em segundo lugar, cria-se um ramo para cada valor; em terceiro lugar, divide-se a árvore em nós, até que as divisões não sejam mais significativas. A construção do modelo de árvore de decisão segue de cima para baixo, e um grupo de dados de treinamento é distribuído em cada estágio de subdivisão (nó) de acordo com a variável descritiva, selecionada com base no critério de divisão. A divisão dura até que todas as observações em um subgrupo tenham a mesmo modo de escolha na parte inferior (SAFITRI; SURJANDARI, 2017).

Um dos algoritmos mais utilizados para indução de árvores de decisão é o *Classification Algorithms and Regression Trees (CART)*, desenvolvido por Breiman et al. (1984). O CART estabelece um relacionamento dentro das variáveis independentes e dependentes, e o algoritmo é ajustado por sucessivas divisões binárias no conjunto de dados, fazendo com que os subconjuntos resultantes sejam cada vez mais homogêneos em relação a variável dependente. Essas divisões são representadas por uma estrutura de árvore binária e cada nó corresponde a uma divisão (PITOMBO; GOMES, 2014).

Os resultados do algoritmo CART são sempre árvores binárias que podem ser percorridas da sua raiz até as folhas respondendo apenas perguntas do tipo “sim” ou “não” (BASGALUPP, 2010). Em uma árvore de decisão binária cada nó possui uma pergunta e caso a resposta seja positiva os dados seguem pelo nó esquerdo, caso contrário, seguem para o nó direito.

Para divisão dos atributos, o CART utiliza a técnica de pesquisa exaustiva para definição das entradas dos nós. O algoritmo também permite o uso de combinações lineares entre atributos (BASGALUPP, 2010). De acordo com Raimundo et al. (2008), em testes realizados pelos autores do CART, o índice *Gini* apresenta os melhores

resultados para a seleção dos atribuídos de divisão dos nós, sendo recomendado o uso deste critério.

Após o término de construção da árvore de decisão, é adotada uma técnica chamada poda, que visa limitar o tamanho da árvore, removendo partes que não contribuem para a previsão da classificação, gerando uma árvore menos complexa, com melhor desempenho e mais fácil compreensão (GARCIA, 2003).

Existem diversas maneiras de realizar a poda em uma árvore de decisão, podendo ser classificadas em pré-poda e pós-poda. A pré-poda é feita durante o processo de construção da árvore, adotando um critério de parada transformando um sub-nó em um nó terminal. Já a pós-poda é realizada após a construção da árvore, onde são removidos os sub-nós não significativos, reduzindo o número de nós de decisão (BASGALUPP, 2010).

Segundo Raimundo *et al.* (2008), o algoritmo CART utiliza a técnica da poda por minimização do custo complexidade, que tem por objetivo a remoção ou corte dos sub-nós que não são relevantes para a classificação, consistindo na criação de uma árvore inicial e a podando para obter um conjunto de árvores menores. A ideia principal da técnica de poda é definir uma relação entre o custo de erro da árvore e sua complexidade. Para isto, deve ser encontrado pelo algoritmo o valor do parâmetro de complexidade, que resulta em uma sub-árvore que minimiza o erro no sub-nó da árvore (LAURETTO, 2010).

2.3.1.2 Avaliação do desempenho do modelo

Após a etapa de poda deve-se avaliar o desempenho do modelo de classificação para mensurar o poder de predição da árvore de decisão. Para que a avaliação seja eficaz é preciso ser cuidadoso na escolha das amostras utilizadas para a indução (treinamento) e avaliação (teste) da árvore. Geralmente a amostra é separada em dois terços para treinamento (aproximadamente 70%) e um terço para testes (aproximadamente 30%), quando há uma quantidade relevante de dados (BASGALUPP, 2010).

As medidas de desempenho utilizadas para avaliar o modelo podem ser calculadas a partir da Matriz de Confusão, descrita como uma tabela que identifica o desempenho de um modelo a partir de um conjunto de dados, onde os valores verdadeiros são conhecidos comparando os mesmos como os valores preditos. Em

uma matriz de confusão existem os verdadeiros positivos (VP), que foram classificados na classe referência corretamente; os falsos positivos (FP), classificados na classe referência, mas pertencem à outra classe; os verdadeiros negativos (VN), classificados nas outras classes corretamente; e os falsos negativos (FN), que pertencem à classe referência, mas foram classificados em outra classe (BASGALUPP, 2010).

Após a construção da matriz de confusão podem ser calculadas medidas de desempenho, sendo estas utilizadas para avaliar a classificação de uma árvore de decisão. Uma delas é a acurácia, Equação (7), que estima a probabilidade de acerto das previsões, podendo também ser expresso em termo de taxa de erro, que é o complemento da acurácia, isto é, a probabilidade de erro das previsões. Estas medidas são as mais usadas em problema de classificação de dados (BASGALUPP, 2010).

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (7)$$

Motta (2016) realizou uma revisão sistemática da literatura e selecionou 52 artigos que utilizavam técnicas de classificação, dos quais 75% destes artigos fizeram uso da árvore de decisão. Esses trabalhos apresentaram resultados de acurácia acima de 50%, com média das acurácias obtidas de 86,36%. Considerando somente as árvores de decisão, o valor médio da acurácia de classificação foi de 86,75%.

O uso da acurácia para avaliação de desempenho de um modelo de classificação negligencia aspectos de não homogeneidade do conjunto de dados, e o diferente custo em se classificar erroneamente cada classe existente. Neste caso, é interessante discriminar os acertos e erros que são cometidos por um classificador para cada classe (ALVARENGA JUNIOR, 2018). Assim, também podem ser obtidas outras medidas de desempenho como a especificidade, Equação (8), que classifica os acertos relacionados com as classes negativas; e a sensibilidade, Equação (9), que classifica os acertos relativos às classes positivas (BASGALUPP, 2010).

$$Especificidade = \frac{VN}{VN + FP} \quad (8)$$

$$Sensibilidade = \frac{VP}{VP + FN} \quad (9)$$

O índice Kappa, Equação (10), também pode ser estimado a partir da matriz de confusão, e mede a concordância entre as classes com variação de zero a um, sendo que valores entre 0 a 0,20 indicam concordância baixa; entre 0,21 a 0,40 concordância leve; entre 0,41 a 0,60 concordância moderada; entre 0,61 a 0,80 concordância substancial; e a partir de 0,81 a concordância é quase perfeita (LANDIS; KOCH, 1977).

$$Kappa = \frac{P(O) - P(E)}{1 + P(E)} \quad (10)$$

onde $P(O)$ é a soma da diagonal principal da matriz de confusão dividido pelo total de amostras; e $P(E)$ é a somatório do produto da probabilidade de acerto de cada classe pela probabilidade de erro de cada classe.

2.3.2 Regressão logística

Segundo Montgomery e Runger (2012), em uma análise na qual a variável resposta é quantitativa, a regressão linear frequentemente funciona bem, mas em uma situação onde a variável resposta deriva de uma observação de uma resposta qualitativa, e são atribuídos valores a essas respostas, a regressão mais indicada é a logística.

Conforme Mesquita (2014), a técnica de regressão logística foi desenvolvida no século XIX, mas obteve maior visibilidade após 1950. A regressão logística descreve a relação entre uma variável dependente qualitativa binária, associada a um conjunto de variáveis independentes qualitativas ou quantitativas.

A regressão logística é considerada uma técnica de fácil interpretação pelos estatísticos, avaliando a probabilidade de ocorrência de determinado evento e como cada variável exerce influência no resultado do evento estudado (MESQUITA, 2014).

De acordo com Hair Junior *et al.* (2009, p. 225), “regressão logística é uma forma especializada de regressão que é formulada para prever e explicar uma variável categórica binária (dois grupos), e não uma medida dependente métrica”.

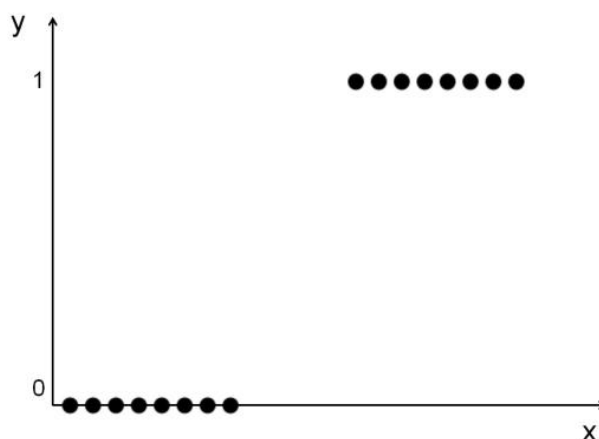
Se a discretização se restringe apenas a duas opções, a regressão logística é dita binomial, caso haja mais opções, a regressão é generalizada e nomeada multinomial. Vale salientar que essas regressões também são denominadas por Modelos Logit Binomial e Logit Multinomial (LINDNER, 2015, p.35).

A regressão logística binomial é normalmente expressada por duas categorias conforme pode ser observado na Figura 3, onde é representada a relação entre

variável dependente binária (y) e variável independente contínua (x), segundo Lindner (2015).

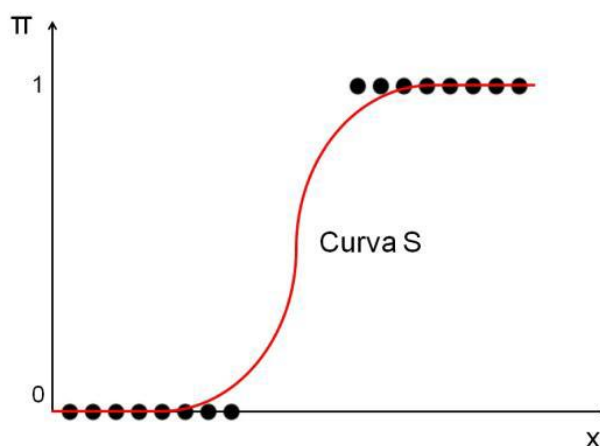
A regressão logística possibilita o uso de um modelo com curva em S para prever a probabilidade (π) de um evento, conforme representado na Figura 4, de acordo com Lindner (2015).

Figura 3 - Gráfico de dispersão para variável dependente com apenas duas categorias.



Fonte: Lindner (2015, p.35).

Figura 4 - Curva S de probabilidade para a variável dependente binária.



Fonte: Lindner (2015, p.36).

2.3.2.1 Regressão logística simples

Conforme Lindner e Pitombo (2016), a regressão logística simples permite o uso de um modelo com curva em S para prever a probabilidade π de um evento categórico. A variável dependente possui duas categorias ("0" e "1") e apenas uma

variável independente envolvida. A modelagem da curva S é dada por uma transformação logística da probabilidade π , conforme as Equações (11) a (13).

$$g(x) = \ln\left(\frac{\pi}{1-\pi}\right) \quad (11)$$

$$g(x) = \beta_0 + \beta_1 x_1 \quad (12)$$

$$\pi = \frac{1}{1 + e^{-g(x)}} \quad (13)$$

sendo $g(x)$ a função logística da variável dependente; x é a variável independente; $x_0 = 1$; e β_0, β_1 os parâmetros da equação do modelo.

2.3.2.2 Regressão logística múltipla

A regressão logística múltipla consiste em um modelo onde a variável dependente é binária e apresenta mais de uma variável independente, alterando a função logística, que agora considera as múltiplas variáveis independentes, conforme Equação (14) (LINDNER, 2015).

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (14)$$

sendo $g(x)$ a função logística da variável dependente; x é a variável independente; $x_0=1, \beta_0, \beta_1, \dots, \beta_n$ os parâmetros da equação; e n o número de variáveis independentes.

As mesmas relações de avaliação identificadas anteriormente na regressão logística simples podem ser utilizadas na regressão logística múltipla (LINDNER, 2015). A regressão logística múltipla pode ser utilizada para prever escolhas discretas modais (LINDNER; PITOMBO, 2016).

2.3.2.3 Regressão logística multinomial

A regressão logística multinomial trata-se de um modelo logístico em que é permitido que a variável categórica dependente apresente mais de duas categorias, sejam elas nominais ou ordinais. Neste modelo de regressão, deve-se escolher uma das categorias da variável dependente como referência (FAVERO *et al.*, 2009).

De acordo com Lindner (2015), desta forma, são necessárias k funções logísticas, como mostrado na Equação (15).

$$\begin{aligned} g_1(x) &= \beta_{1,0} + \beta_{1,1}x_1 + \beta_{1,2}x_2 + \cdots + \beta_{1,n}x_n \\ g_2(x) &= \beta_{2,0} + \beta_{2,1}x_1 + \beta_{2,2}x_2 + \cdots + \beta_{2,n}x_n \\ g_k(x) &= \beta_{k,0} + \beta_{k,1}x_1 + \beta_{k,2}x_2 + \cdots + \beta_{k,n}x_n \\ g_k(x) &= \sum_{i=0}^n \beta_{k,i}x_i \end{aligned} \quad (15)$$

sendo $g(x)$ a função logística da variável dependente; x é a variável independente; $x_0=1$, $\beta_{k,0}$, $\beta_{k,1}$, ..., $\beta_{k,n}$ os parâmetros da equação; e n o número de variáveis independentes.

Com isso, Lindner (2015) afirma que as probabilidades π_k para cada uma das k classes podem ser descrita na Equação (16).

$$\pi_k = \frac{1}{1 + e^{-g_k(x)}} \quad (16)$$

As mesmas relações de avaliação identificadas anteriormente na regressão logística simples e múltipla podem ser utilizadas na regressão logística multinomial (LINDNER, 2015).

2.3.2.4 Estimativa do modelo de regressão logística

A estimação dos parâmetros da equação de um modelo de regressão logística é realizada a partir dos dados da amostra. De acordo com Silva (2010), quanto maior for o tamanho da amostra, mais eficiente será o processo de estimação dos parâmetros.

Para o processo de estimação do modelo estatístico, sugere-se que a amostra inicial seja separada em duas partes, uma destinada à calibração e outra a validação do modelo. De acordo com Lindner (2015), não existe uma proporção padrão para a calibração e validação, entretanto a maior proporção deve ser destinada a calibração.

2.3.2.5 Testes de significância

Depois da estimação dos parâmetros da regressão, é necessário avaliar a significância do modelo, ou seja, determinar se as variáveis independentes do modelo estão associadas significativamente com a variável dependente.

Um teste de significância importante na regressão logística é o teste da razão de verossimilhança, onde a hipótese de que pelo menos um dos parâmetros β_{ij} é diferente de zero (exceto os interceptos – parâmetros β_{i0}) é testada. Esse teste faz uma comparação entre o valor da função de verossimilhança para o modelo, contendo apenas os interceptos e a verossimilhança do modelo final com todos os parâmetros estimados (FIGUEIRA, 2006).

De acordo com Bittencourt (2003), para testar se o modelo logístico ajustado é adequado, deve-se obter a estatística de teste D , chamada de *deviance*, havendo uma distribuição qui-quadrado, calculada pela Equação (17). A deviance é uma medida estatística que avalia a significância dos parâmetros estimados e tem por base o teste da razão de verossimilhança (ALVARENGA, 2015).

$$D = -2 \ln \frac{l(\beta_0)}{l(x, \beta)} \sim \chi^2_{(k-1)p} \quad (17)$$

sendo que, $l(\beta_0)$ é o valor da função de verossimilhança apenas com os interceptos; $l(x, \beta)$ é o valor da função de verossimilhança para o modelo final; k é o número de categorias da variável-resposta Y ; e p é o número de variáveis independentes (x) incluídas no modelo.

O teste *deviance* tem como hipótese nula que os parâmetros da equação não são significativos e a hipótese alternativa é que os parâmetros são significativos. Então devemos rejeitar a hipótese nula quando para um determinado nível de significância, a estatística χ^2 é maior que um ponto crítico χ_{α}^2 , usando a tabela da distribuição qui-quadrado a partir de uma probabilidade α e de certo número de graus de liberdade. Caso seja rejeitada a hipótese nula, temos que o parâmetro testado é significativo para o modelo.

Para a realização de testes de significância individuais para os parâmetros β_{ij} , é bastante comum a utilização da estatística de *Wald*, onde a hipótese nula é a de que

o particular coeficiente β_{ij} é igual a zero. A estatística W de *Wald* é definida como o quadrado da razão entre a estimativa de máxima verossimilhança para o coeficiente e seu respectivo erro-padrão (EP), conforme apresentado na Equação (18) (BITTENCOURT, 2003). Caso o coeficiente logístico seja estatisticamente significativo, o mesmo pode ser interpretado em termo de como impacta o modelo estimado e consequentemente seu desempenho preditivo (HAIR JUNIOR *et al*, 2009).

$$W = \left(\frac{\beta_{ij}}{EP(\beta_{ij})} \right)^2 \sim \chi^2 \quad (18)$$

2.3.2.6 Medidas de avaliação e ajuste do modelo

De acordo com Alvarenga (2015), uma medida para avaliação de um modelo logístico é o critério de informação de Akaike (AIC), sendo que esta medida não é um teste de hipóteses, mas sim uma estatística com base no logaritmo de verossimilhança obtida a partir da Equação (19). A AIC é uma medida relativa à perda de informação devido ao ajuste de um determinado modelo, então quanto menor for o seu valor, menos informação foi perdida e, portanto, melhor será o ajuste do modelo.

$$AIC = -2[\log(L) - k] \quad (19)$$

sendo que, k é o número de parâmetros do modelo, e L é o valor de verossimilhança para o modelo estimado.

Para avaliação da qualidade do ajuste de modelos logísticos existem os pseudos R^2 , como os de McFadden, McFadden Ajustado, Cox & Snell e Nagelkerke. Os modelos logísticos não tem uma medida como o R^2 nos modelos lineares, assim várias medidas estatísticas são propostas para esses modelos. Os pseudo R^2 são calculados com base na função de verossimilhança e correspondem a uma fração da variância que é compartilhada entre as variáveis. As Equações (20), (21) e (22) representam respectivamente os pseudo R^2 de McFadden, Cox & Snell e de Nagelkerke (SILVA, 2010).

$$R^2 = 1 - \left(\frac{-2LL_{null}}{-2LL_k} \right) \quad (20)$$

$$R^2 = 1 - \left(\frac{-2LL_{null}}{-2LL_k} \right)^{\frac{2}{n}} \quad (21)$$

$$R^2 = \frac{1 - \left(\frac{-2LL_{null}}{-2LL_k} \right)^{\frac{2}{n}}}{1 - (-2LL_{null})^{\frac{2}{n}}} \quad (22)$$

sendo que, $-2LL_{null}$ é a probabilidade de interceptar apenas o modelo; $-2LL_k$ é a probabilidade do modelo especificado; e n é o número de observações.

De acordo com Lindner (2015), os pseudo R^2 possuem um valor ótimo quando tendem a “1”. Isto é, quanto mais próximos do valor “1”, melhor é o ajuste do modelo. No entanto, Ortúzar e Willumsen (2011) afirmam que os pseudo R^2 não possuem uma interpretação intuitiva para valores intermediários, e obter valores próximos a 1,0 são difíceis de serem alcançados; assim, os autores declaram que valores em torno de 0,4 são geralmente considerados excelentes ajustes.

Para avaliação dos modelos também podem ser utilizadas medidas de desempenho obtidas através da matriz de confusão, como acurácia, especificidade, sensibilidade e índice Kappa; medidas estas já descritas anteriormente na seção 2.3.1.2.

3 METODOLOGIA

Neste capítulo são descritos os procedimentos metodológicos utilizados na pesquisa. Inicialmente, apresenta-se a área de estudo da pesquisa e a caracterização dos dados utilizados, descrevendo posteriormente o método usado para alcançar os objetivos propostos.

3.1 ÁREA DE ESTUDO

O Polo Universitário, que é objeto de estudo desta pesquisa, está localizado na cidade de Joinville, tendo este uma área territorial de 1.127,946 km² e população estimada de 590.466 habitantes (IBGE, 2019). Joinville é a maior cidade catarinense e polo industrial da região Sul, responsável por cerca de 20% das exportações do estado, concentrando grande parte de sua atividade econômica na indústria, com destaque para os setores metalmeccânico, têxtil, plástico, metalúrgico, químico e farmacêutico (SEPUD, 2018).

Em conformidade com a Política Nacional de Mobilidade Urbana expressa pela Lei Federal nº 12.587 (BRASIL, 2012), Joinville aprovou em 2015 o Decreto nº 24.181 que estabeleceu o Plano de Mobilidade Sustentável de Joinville (PlanMOB), que tem por finalidade orientar as ações relativas a mobilidade urbana, priorizando os modos de transporte sustentáveis, visando aos deslocamentos das pessoas e bens. Além disto, busca promover a qualidade de vida e acesso a bens, serviços e lazer.

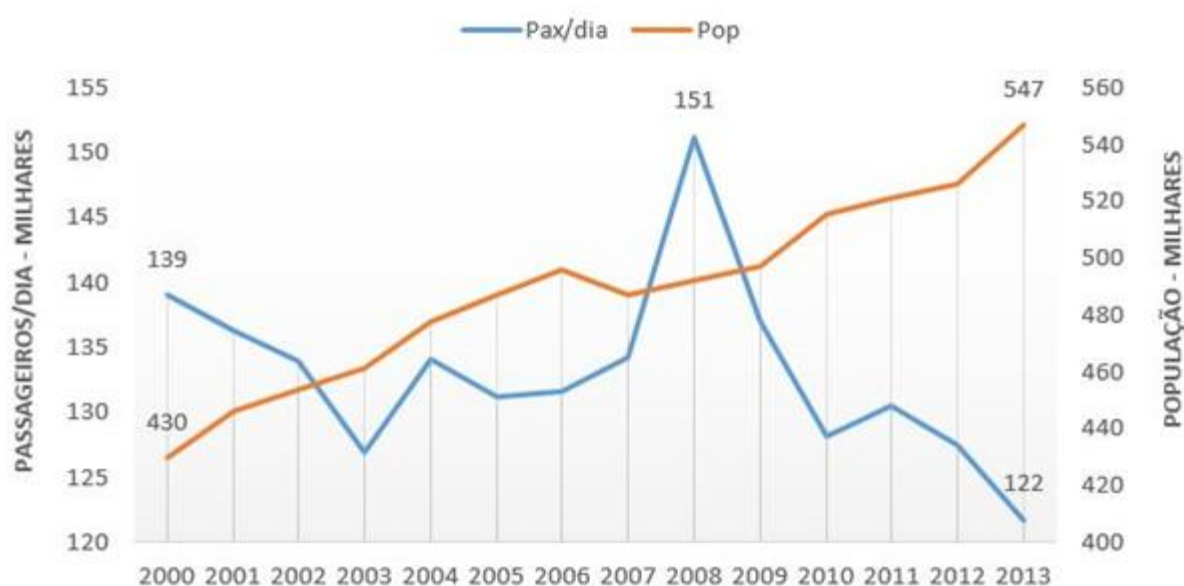
De acordo com SEPUD (2018), uma pesquisa realizada em 2010 indicou a seguinte divisão do modo de transporte em Joinville em ordem decrescente: 35% dos usuários utilizam o automóvel, 24% transporte público, 23% se deslocam a pé, 11% bicicleta, 6% motocicleta e 1% fazem uso de outro modo de transporte. Em comparação ao ano 2000, em 2017 a frota de veículos automotores aumentou significativamente, sendo que a quantidade de motocicletas quadruplicou e a de automóveis quase triplicou (SEPUD, 2018). Conforme dados do Departamento Estadual de Trânsito do Estado de Santa Catarina (DETRAN/SC, 2019) o município de Joinville em setembro de 2019 registrou 266 mil automóveis licenciados.

Com relação ao transporte coletivo (ônibus), Joinville hoje possui um sistema integrado com pagamento de passagem única. Segundo dados apresentados por SEPUD (2018), o número de usuários de ônibus vem caindo com o passar dos anos,

conforme pode ser observado nas Figuras 5 e 6. Além disso, com o crescimento da frota de veículos particulares, ocorreu um aumento no tempo de deslocamento do ônibus; em 1999 a velocidade média do ônibus era 25 km/h, reduzindo para 18 km/h em 2018.

A Figura 5 apresenta duas curvas, uma representando a população de Joinville no período de 2000 a 2013, e a outra representando a evolução dos usuários do transporte por ônibus no mesmo período. Assim, observa-se que ao longo do período ocorreu um crescimento da população do município, mas, em contrapartida, os usuários de ônibus diminuíram. Esta queda no número de passageiros de ônibus também pode ser constatada na Figura 6, onde é apresentada a porcentagem de passageiros de ônibus com relação a população de Joinville, indicando, de maneira geral, a mesma queda observada na Figura 5.

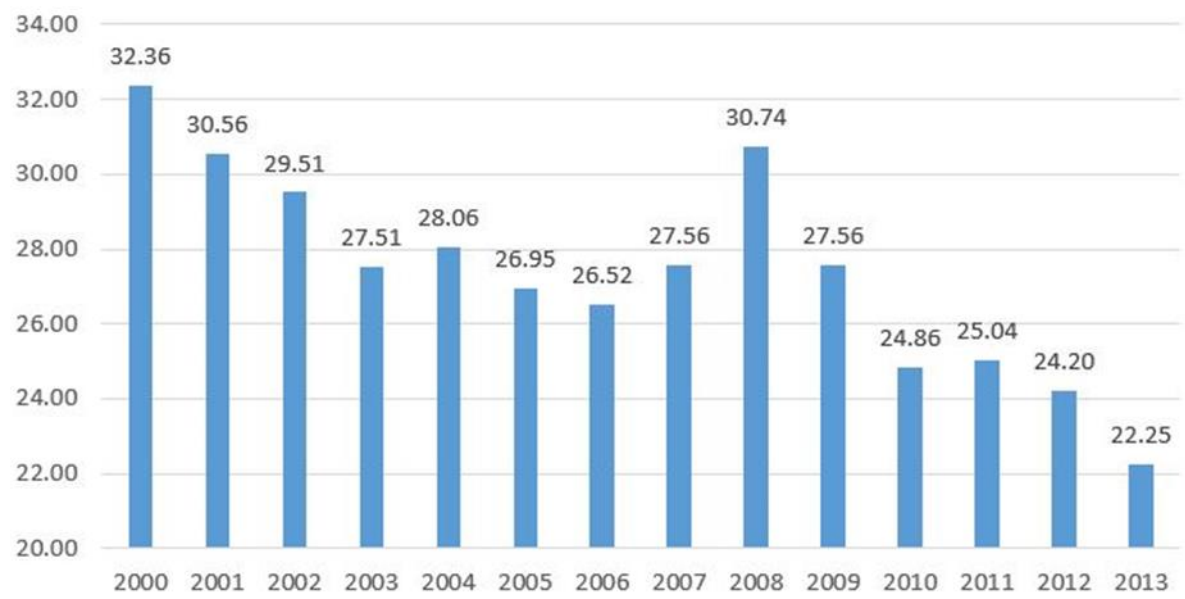
Figura 5 - Evolução dos passageiros de ônibus em Joinville



Fonte: SEPUD (2018).

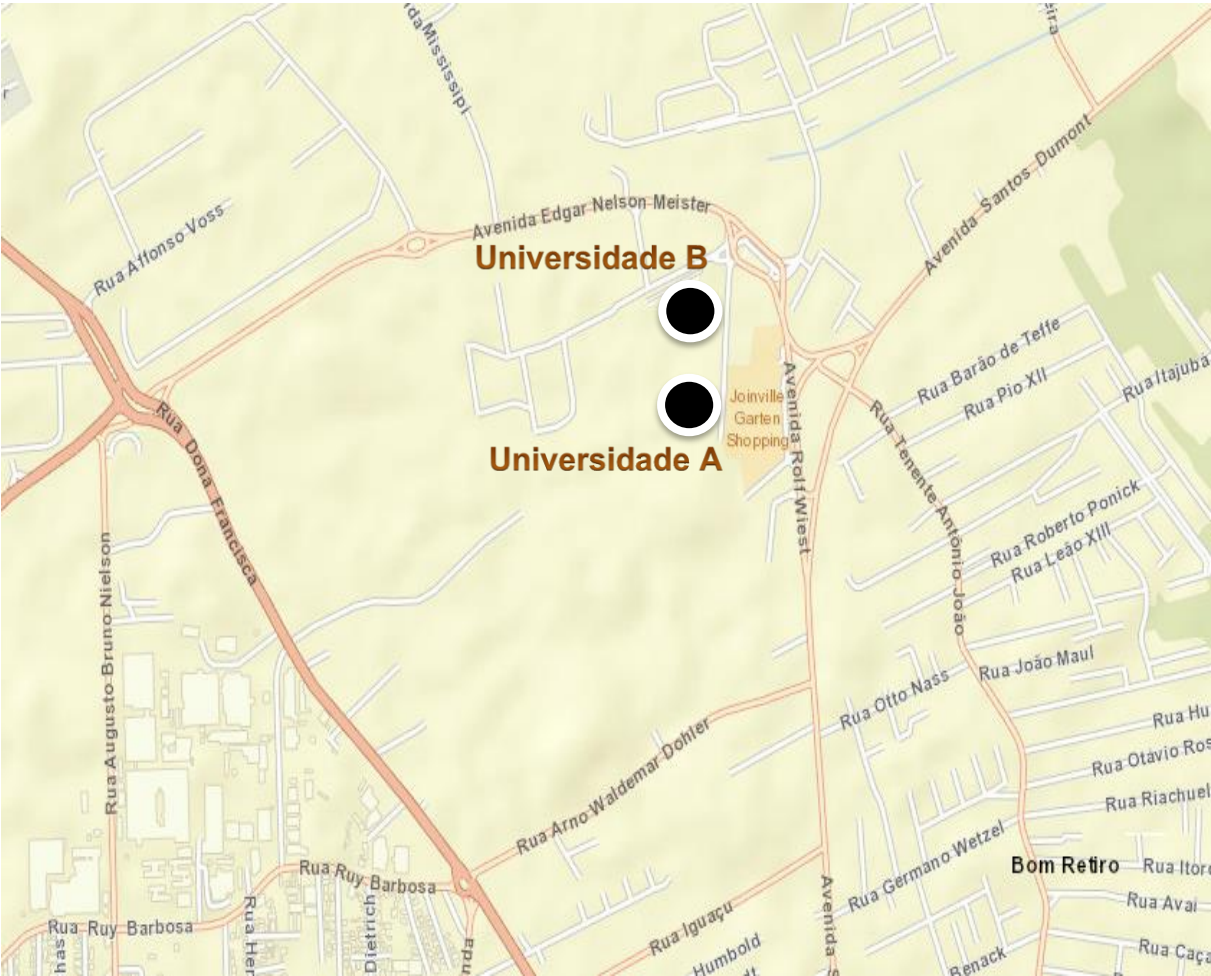
A área de estudo nesta pesquisa refere-se a duas instituições de ensino superior localizadas na zona norte de Joinville. Estas duas instituições formam um polo universitário, conhecido como o *campus* universitário do município. A Figura 7 apresenta a localização das instituições de ensino estudadas neste trabalho.

Figura 6 - Evolução dos passageiros do ônibus em relação a população em Joinville



Fonte: SEPUD (2018).

Figura 7 - Localização das instituições de ensino superior que contemplam a pesquisa



Fonte: Simgeo (2019).

Na Universidade A (pública) são ofertados cursos de graduação e pós-graduação nas áreas de engenharia, informática e licenciatura, possuindo algo em torno de dois mil e novecentos alunos (UDESC, 2019). Já a Universidade B (privada) com cursos de graduação e pós-graduação nas mais diversas áreas, conta com mais de seis mil e quinhentos alunos (INEP, 2019).

3.2 BANCO DE DADOS

O banco de dados a ser estudado trata-se de um questionário desenvolvido como parte da dissertação de mestrado de Schubert (2019), aplicado a alunos de graduação e pós-graduação no ano de 2018. O questionário conta com 514 respondentes, com informações referentes a dados pessoais, caracterização dos deslocamentos, período de curso, modo de transporte utilizado, e situações para possíveis mudanças do modo de transporte. O questionário contempla alunos de seis instituições de ensino superior localizadas em regiões distintas no município de Joinville e teve como objetivo estudar o deslocamento dos estudantes universitários.

A base de dados de Schubert (2019) contém informações referentes às pesquisas de preferência declarada e revelada, que são as mais utilizadas para a modelagem de escolha modal. Nesta pesquisa foram usados os dados de preferência revelada, ou seja, os modos que os alunos utilizam para o seu deslocamento diário no cenário existente, de acordo com a oferta atual de transportes.

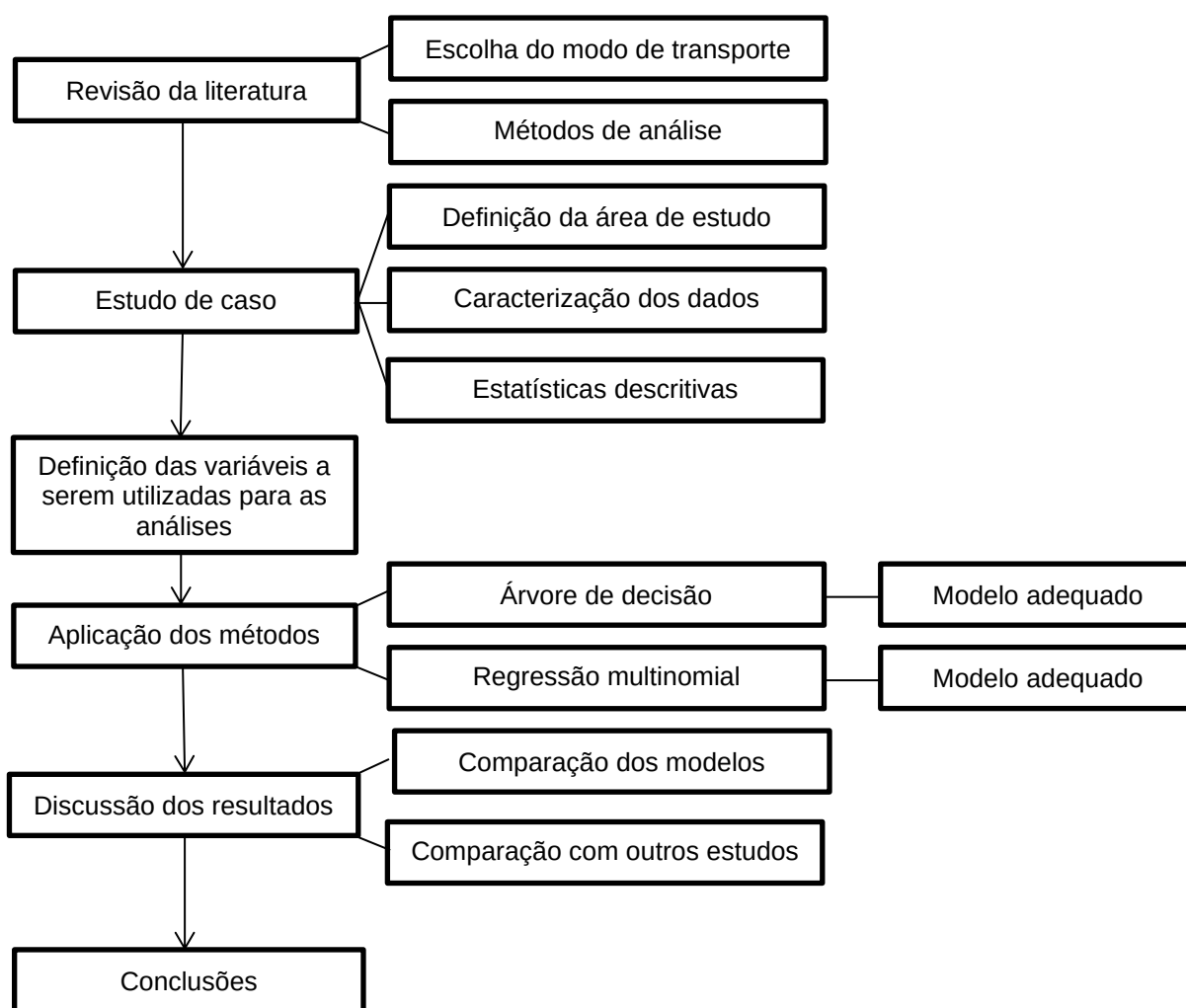
Para as modelagens foram utilizadas somente as informações relativas a duas instituições de ensino, que são o objeto de estudo como descrito anteriormente. Sendo assim, a amostra utilizada na pesquisa possui as informações de 308 respondentes, dos quais 109 são da instituição pública e 199 da instituição privada.

3.3 MÉTODO

O método proposto para este trabalho é dividido em etapas, conforme os objetivos específicos propostos. Na Figura 8 é apresentado o esquema representativo do procedimento utilizado na pesquisa. Inicialmente, buscou-se suporte na literatura para o embasamento teórico da presente pesquisa, para posterior definição do estudo de caso a ser desenvolvimento e quais os métodos seriam utilizados para análise dos dados. Após a aplicação dos métodos, definiu-se quais os melhores modelos que

representam o objeto de estudo, e comparou-se os modelos entre si e com outros estudos similares, com o objetivo de analisar os resultados e obter conclusões sobre a pesquisa.

Figura 8 - Esquema representativo do procedimento de pesquisa



Fonte: O autor (2019).

Para realização das análises estatísticas, foram escolhidos os seguintes modos de transportes: carro, modo ativo (a pé e bicicleta), transporte público (ônibus) e transporte público fretado, os demais foram considerados como categoria outros, aparecendo os modos motocicleta e carona (modo caracterizado na pesquisa como o indivíduo que utiliza o carro como modo de transporte, mas não é o condutor). As análises serão realizadas adotando um nível de significância de 5%.

Inicialmente, fez-se a criação da árvore de decisão classificando eficientemente o públicos-alvo em subgrupos, cujos membros possuem características homogêneas

que influenciam a escolha do modo de transporte. Para indução da árvore de decisão utilizou-se o algoritmo *Classification and Regression Trees (CART)*, desenvolvido por Breiman et al. (1984), juntamente com o pacote *rpart* (THERNEAU; ATKINSON, 2018) do programa computacional R versão 3.5.3 (R Core Team, 2019) e a interface RStudio. Utilizando o pacote *caret* (KUHN, 2019), selecionou-se aleatoriamente 70% da amostra para treinamento e 30% para teste e foram geradas as duas árvores. Utilizou-se também uma ferramenta para a poda das árvores com o intuito de diminuir o número de atributos significativos e obter uma melhor classificação. Esta poda foi realizada através do parâmetro de complexidade da árvore (CP), onde é possível encontrar o CP ótimo para podar uma árvore com a função *printcp*, encontrada no pacote *rpart* da linguagem R. Após a poda para avaliação da classificação da árvore de decisão, utilizou-se a função *confusionMatrix*, encontrada no pacote *caret*. Para a avaliação dos modelos foram geradas as matrizes de confusão, obtendo a partir destas as seguintes medidas de desempenho do modelo: acurácia, índice Kappa, especificidade e sensibilidade.

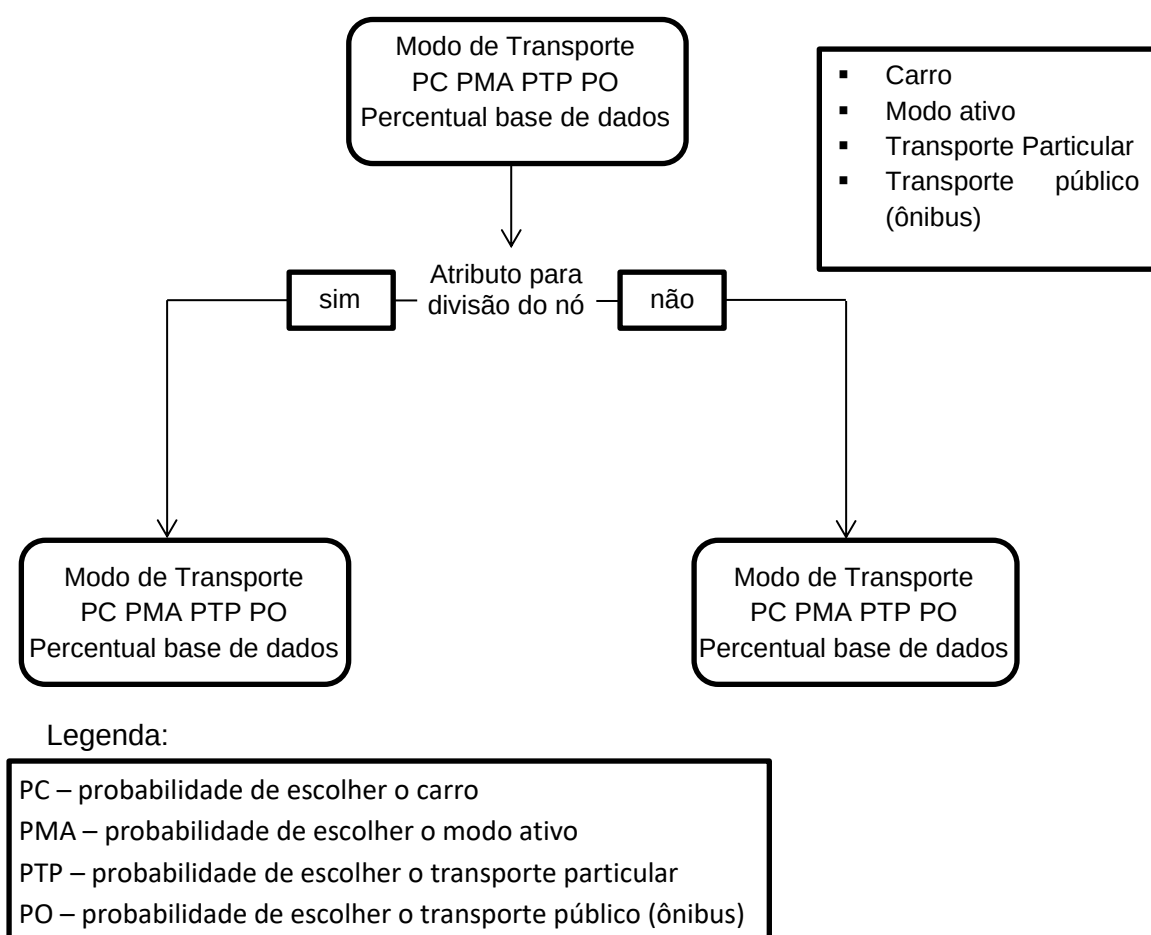
A árvore de decisão é representada de forma gráfica, sendo que em cada nó da árvore apresenta três linhas de informações, na primeira aparece o modo de transporte com a maior probabilidade de escolha, na segunda as probabilidades de escolha de cada modo de transporte, seguindo a ordem da legenda, e na terceira o percentual da base de dados para aquele nó. Abaixo de cada nó contém o atributo responsável pela divisão daquele nó específico, e qual o critério utilizado para divisão dos dados. Os dados que são afirmativos seguem sempre para o lado esquerdo e os demais para o lado direito. Quando não existem mais atributos significativos para a divisão da árvore, a mesma é finalizada com o nó terminal, podendo apresentar vários nós terminais.

A Figura 9 apresenta o esquema da árvore de decisão, onde a partir do primeiro atributo, a árvore é dividida em dois ramos (divisão do banco de dados), e cada um dos ramos apresenta uma nova divisão modal em função do atributo analisado. O modo de transporte, que aparece em destaque em cada nó, é somente uma indicação do modo predominante, sendo as análises realizadas com base em cada nova divisão modal da árvore.

Posteriormente, novamente com o auxílio do programa computacional R, juntamente com pacote *nnet* versão 7.3-12 (VENABLES; RIPLEY, 2002) e a função *multinom*, foram gerados vários modelos de regressão logística multinomial, sendo

estes testados, até que fosse encontrado o melhor modelo para representar quais fatores influenciam os estudantes universitários na escolha do modo de transporte no município de Joinville. Para os testes dos modelos foram geradas as matrizes de confusão, com a acurácia e o índice Kappa. Também foram obtidos os pseudo R^2 de McFadden, Cox & Snell e Nagelkerke, e o critério de informação Akaike (AIC).

Figura 9 - Esquema da árvore de decisão para escolha do modo de transporte



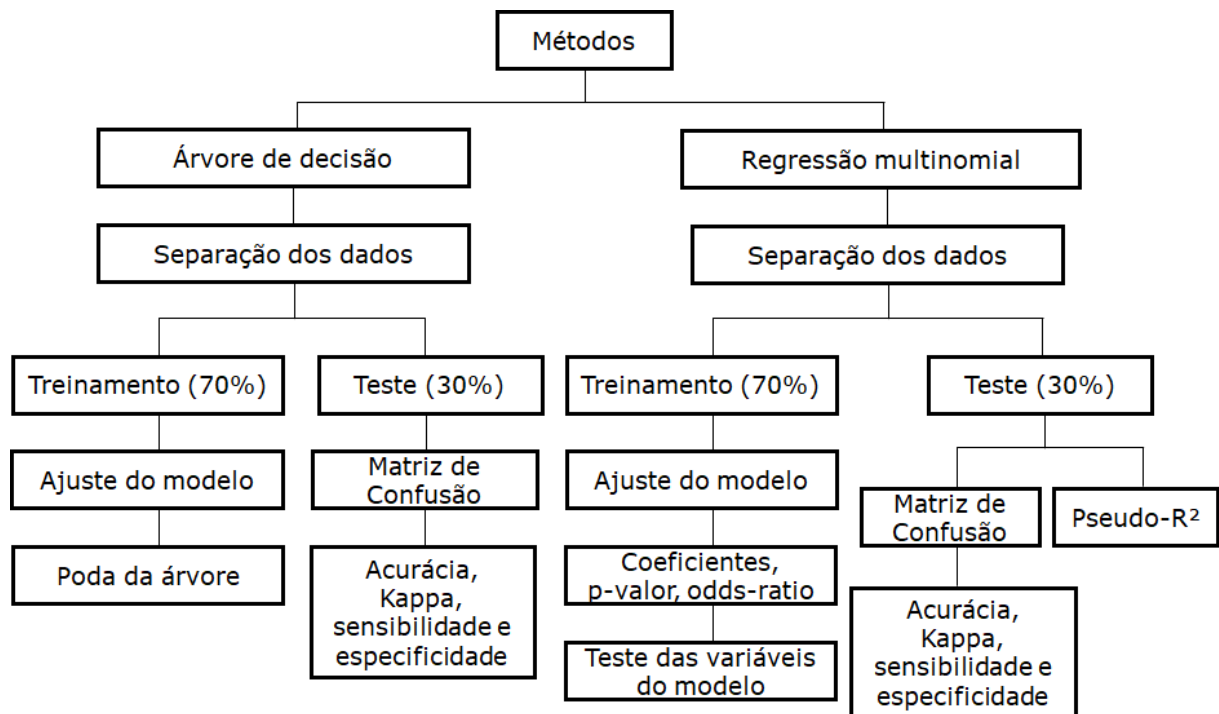
Fonte: O autor (2019).

As análises na presente pesquisa foram realizadas para três situações: a primeira considerando o Polo Universitário, que se refere aos dados das duas instituições de ensino; a segunda considerando apenas os dados da Universidade A; e a terceira considerando apenas os dados da Universidade B.

A Figura 10 apresenta o esquema representativo dos métodos utilizados para a modelagem de escolha modal, indicando as etapas anteriormente descritas, dos dois métodos utilizados, árvore de decisão e regressão multinomial.

Após a criação do modelo de regressão, comparou-se o mesmo com o modelo obtido da árvore de decisão, a fim de identificar semelhanças e diferenças nos fatores influentes na escolha do modo de transporte dos estudantes universitários, de acordo com cada método. Para a comparação do desempenho dos modelos utilizou-se duas medidas obtidas com a matriz de confusão, a acurácia e o índice Kappa. Estas medidas foram comparadas entre os modelos para a amostra de treinamento e para a teste.

Figura 10 - Esquema representativos dos métodos utilizados para a modelagem de escolha modal



Fonte: O autor (2019).

Também se realizou a comparação dos modelos obtidos por cada método com a literatura, com relação aos fatores relevantes identificados na pesquisa e os fatores encontrados na literatura envolvendo a escolha modal de estudantes universitários.

4 RESULTADOS E DISCUSSÕES

Neste capítulo é apresentada a caracterização da amostra, e os resultados obtidos para a técnica da árvore de decisão e para a regressão multinomial. As análises foram realizadas para três situações: a primeira considerando o Polo Universitário, que se refere aos dados das duas instituições de ensino; a segunda considerando apenas os dados da Universidade A; e a terceira considerando apenas os dados da Universidade B. Para as árvores de decisão e os modelos multinomiais considerou-se como variável dependente o modo de transporte, sendo que esta categoria é definida pelos modos carro, ativo (a pé e bicicleta), transporte público (ônibus) e transporte público fretado. As demais categorias, como carona e moto, não aparecem nos resultados, pois foram retiradas das análises pelo fato de que, nos modelos iniciais, os parâmetros não foram significativos.

4.1 CARACTERIZAÇÃO DA AMOSTRA

As variáveis utilizadas na pesquisa e suas respectivas categorizações aparecem na Tabela 2. Nesta tabela aparecem variáveis quantitativas (idade, tempo e distância) e variáveis qualitativas (instituição de ensino, gênero, renda familiar, estado civil, cidade onde mora, tem filhos, endereço de origem, trabalho e rotina diária). Nas variáveis qualitativas foram mantidas as categorizações definidas no questionário elaborado por Schubert (2019). Já com relação as variáveis quantitativas, que na base de dados apresentam-se de forma contínua, realizou-se a categorização, transformando essas variáveis em qualitativas ordinais.

O procedimento para categorização das variáveis idade e tempo (tempo gasto no deslocamento até a instituição de ensino), que na base de dados apresentam-se de forma contínua, foi realizado primeiramente definindo de forma aleatória pelo autor, vários intervalos para as categorias, e posteriormente sendo gerados os histogramas individuais para cada intervalo, sendo assim, com base nos histogramas, selecionados para categorização das variáveis, o intervalo em que os dados foram distribuídos de maneira mais uniforme. Este procedimento também foi feito para a variável distância (distância percorrida até a universidade), informação que não aparece na base de dados, mas foi obtida por meio do endereço de origem informado pelo respondente do questionário. Para auxiliar na obtenção das distâncias, foi

utilizada a ferramenta Rotas, presente no *Google Maps* (GOOGLE, 2019). Esta ferramenta disponibiliza vários trajetos, podendo o usuário selecionar o modo de transporte, carro, transporte público, bicicleta ou a pé. Sendo assim, foi indicado os pontos de partida (endereço de origem do estudante) e chegada (Polo Universitário), e o modo de transporte utilizado para cada estudante da base de dados, e definido como valor da variável distância, o menor valor, em quilômetros, indicado nos trajetos sugeridos. Para os modos transporte público fretado, carona e moto, selecionou-se para obtenção do valor da distância, o trajeto indicado utilizando-se o carro.

A Tabela 2 apresenta um resumo do perfil da amostra, e conta com informações relativas às características dos estudantes universitários como gênero, idade, renda familiar, estado civil, cidade onde mora, se tem filhos, se trabalha, qual sua rotina diária, e características envolvendo o seu modo de deslocamento até a universidade como modo de transporte utilizado, tempo gasto e distância percorrida.

A maioria dos respondentes da amostra são da instituição privada (64,61%). De 308 respondentes, 195 são mulheres (63,31%) e 113 são homens (36,69%), e a maioria são jovens com faixa etária de 18 a 23 anos (71,75%), e solteiros (84,42%). Relativo à renda familiar, a maior frequência é de 3 a 5 salários mínimos (46,75%), com a renda de até 2 salários mínimos aparecendo em segundo (25,32%). A maior parte dos respondentes reside em Joinville (94,48%), não tem filhos (94,16%) e trabalha (56,82%).

O modo de transporte utilizado com maior frequência pelos estudantes universitários para o deslocamento até a universidade é o transporte público (ônibus), representando 38,64% (119 respondentes), seguido pelo carro com 25,32% (78 respondentes) e o modo ativo 21,43% (66 respondentes). O maior uso do ônibus como modo de transporte está relacionado ao perfil da amostra descrito na parágrafo anterior, indicando que os respondentes são na maioria jovens solteiros, do sexo feminino e com renda familiar menor que 5 salários mínimos, representando assim, um cenário mais favorável ao uso do transporte público disponível; no caso da cidade de Joinville, o ônibus. Também é importante destacar que o maior incentivo ao uso dos modos ativos pode ser uma alternativa para redução dos congestionamentos e busca de uma mobilidade mais sustentável.

Referente ao local de origem dos deslocamentos dos estudantes, dos 308 respondentes, 222 tem como origem do deslocamento sua casa (72,08%), e os demais, o local de trabalho (27,92%). Esse fato chama a atenção pois, como

comentado anteriormente, a maior parcela dos estudantes trabalha, o que era de se presumir que o percentual de deslocamentos tendo como origem o local de trabalho fosse maior do que o apresentado na pesquisa. Essa situação pode ser explicada pela rotina diária escolhida pelos respondentes, em que parte dos estudantes que trabalha, estuda no período matutino (51 do total de 175), logo, seu deslocamento até a universidade começa em casa. Também observou-se que outros estudantes que trabalham (38 do total de 175), depois do trabalho, primeiramente se dirigem até sua residência antes de irem para a universidade.

Tabela 2 - Categorização e frequência das variáveis presentes na pesquisa

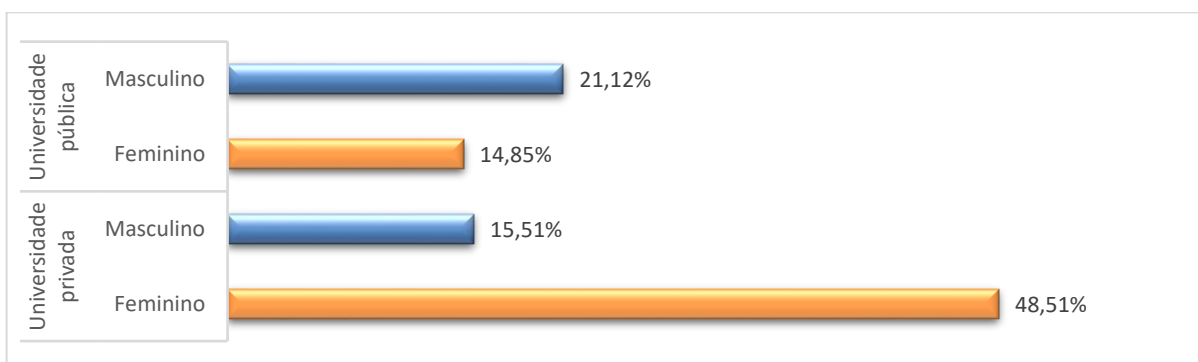
Variáveis	Variáveis Categóricas	Quantidade na Amostra	Porcentagem (%)
Instituição de ensino (Inst_ensino)	A	109	35,39
	B	199	64,61
Gênero (Sexo)	Masculino (0)	113	36,69
	Feminino (1)	195	63,31
Idade	De 18 a 20 anos (0)	117	37,98
	De 21 a 23 anos (1)	104	33,77
	Acima de 23 anos (2)	87	28,25
Renda familiar (Renda_fam)	Até 2 salários mínimos (0)	78	25,32
	De 3 a 5 salários mínimos (1)	144	46,75
	De 6 a 9 salários mínimos (2)	54	17,53
	Mais de 10 salários mínimos (3)	32	10,39
Estado Civil (Est_civil)	Solteiro (0)	260	84,42
	Casado (1)	43	13,96
	Outro (2)	5	1,62
Cidade onde mora (Cidade)	Joinville (0)	291	94,48
	Outro (1)	17	5,52
Tem filhos (Filho)	Sim (0)	18	5,84
	Não (1)	290	94,16
Endereço de origem (End_origem)	Casa (0)	222	72,08
	Trabalho (1)	86	27,92
Trabalho	Sim (0)	175	56,82
	Não (1)	133	43,18
Escolher a rotina diária (Rotina)	Casa-Aula-Casa (0)	133	43,18
	Casa-Aula-Trabalho-Casa (1)	51	16,56
	Casa-Trabalho-Aula-Casa (2)	86	27,92
	Casa-Trabalho-Casa-Aula-Casa (3)	38	12,34
Tempo (tempo_ida)	Até 15 min (0)	74	24,42
	De 16 min a 30 min (1)	105	34,65
	De 31 min a 60 min (2)	90	29,70
	Mais de 60 min (3)	34	11,22
Distancia (distancia_ida)	Até 3 km (0)	47	15,51
	De 4 a 6 km (1)	73	24,09
	De 7 a 8 km (2)	56	18,48
	De 9 a 12 km (3)	65	21,45
	Mais de 12 km (4)	62	20,46
Modo de transporte diário (Modo_transp)	Carro	78	25,32
	Modo ativo	66	21,43
	Transporte público (ônibus)	119	38,64
	Transporte público fretado	11	3,57
	Outro	34	11,04

De maneira geral, aproximadamente 1/3 dos estudantes leva para se deslocar de 16 a 30 minutos (34,65%), enquanto os demais representam respectivamente 29,70% (de 31 a 60 minutos); 24,42% (até 15 minutos); e 11,22% (mais de 60 minutos). Diante disto, pode-se observar que o tempo de deslocamento da maioria dos respondentes se concentra no intervalo de 16 a 60 minutos (64,35%), o que pode estar associado aos modos de transporte mais utilizados pelos respondentes o ônibus e o carro, pois o tempo de deslocamento sofre influência do modo de transporte escolhido.

Já com relação a distância, a maior parte dos estudantes percorre até a universidade de 4 a 6 quilômetros (24,09%), seguido pelas distâncias de 9 a 12 quilômetros (21,45%), mais de 12 quilômetros (20,46%), de 7 a 8 quilômetros (18,48%), e até 3 quilômetros (15,51%). Pode-se observar que não há grandes diferenças entre faixas de categorias relativas a distância percorrida, indicando que os pontos de partida estão distribuídos em várias regiões da cidade, e até em municípios vizinhos.

As Figuras 11, 12, 13 e 14 apresentam respectivamente, as variáveis gênero, renda familiar, idade e distância percorrida. Estas variáveis foram caracterizadas em formas de gráficos de acordo com cada uma das duas universidades estudadas. Na Figura 11 observa-se que a amostra da universidade pública possui mais respondentes do gênero masculino, mas não apresenta uma grande diferença para o feminino. Já na universidade privada, predomina o gênero feminino representando quase a metade da amostra total.

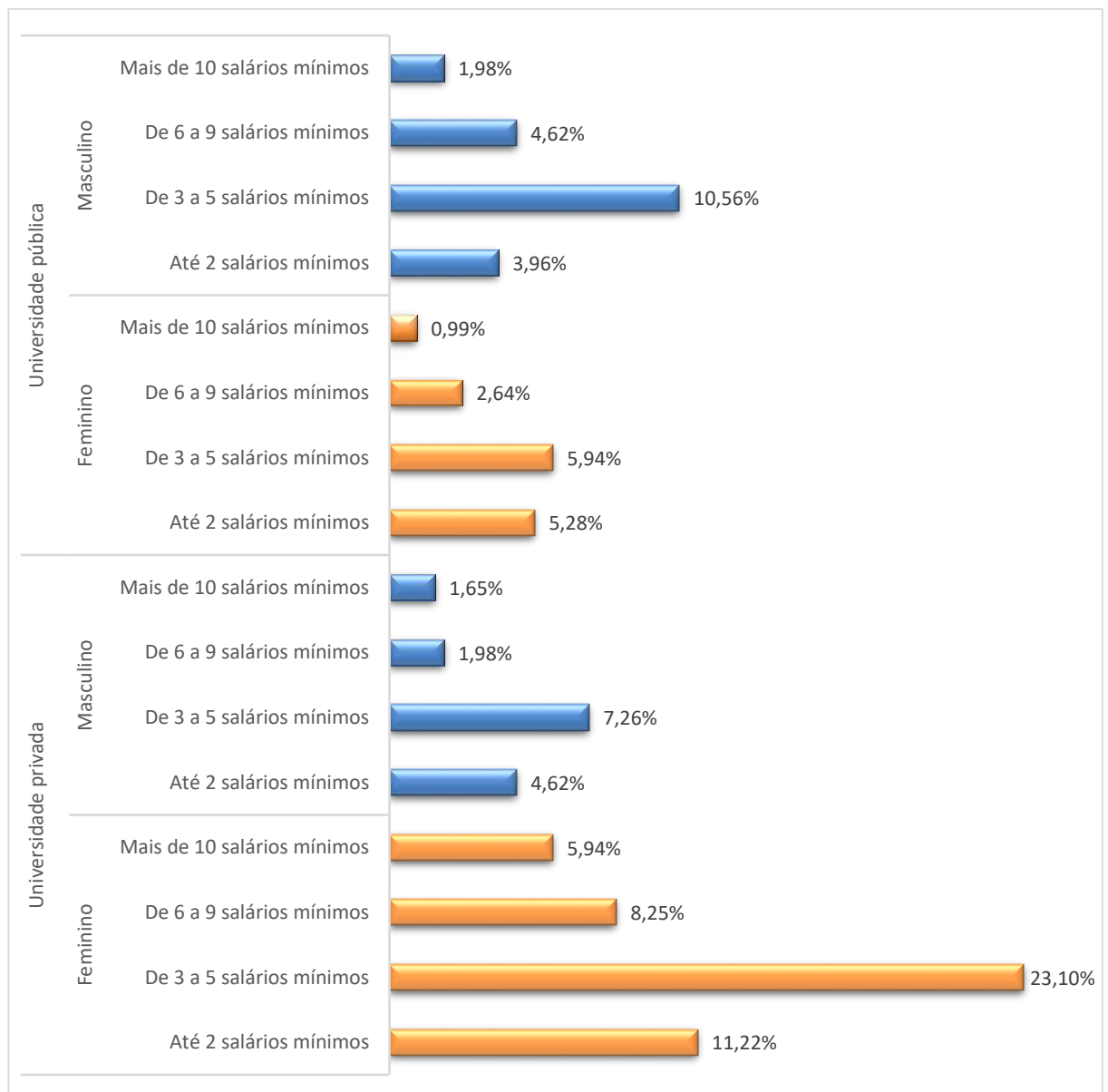
Figura 11 - Gênero de acordo com a Universidade



A Figura 12 indica a distribuição na amostra da renda familiar de acordo com o gênero e a universidade. Assim, pode-se constatar que na universidade pública há

uma predominância no gênero masculino da renda de três a cinco salários mínimos, e no feminino a maior parte também possui renda de três a cinco salários mínimos, mas a renda de até dois salários também é representativa. Na universidade privada, a maior parte dos respondentes tanto do gênero masculino quanto feminino tem renda familiar de três a cinco salários mínimos, com a renda de até dois salários mínimos aparecendo em seguida.

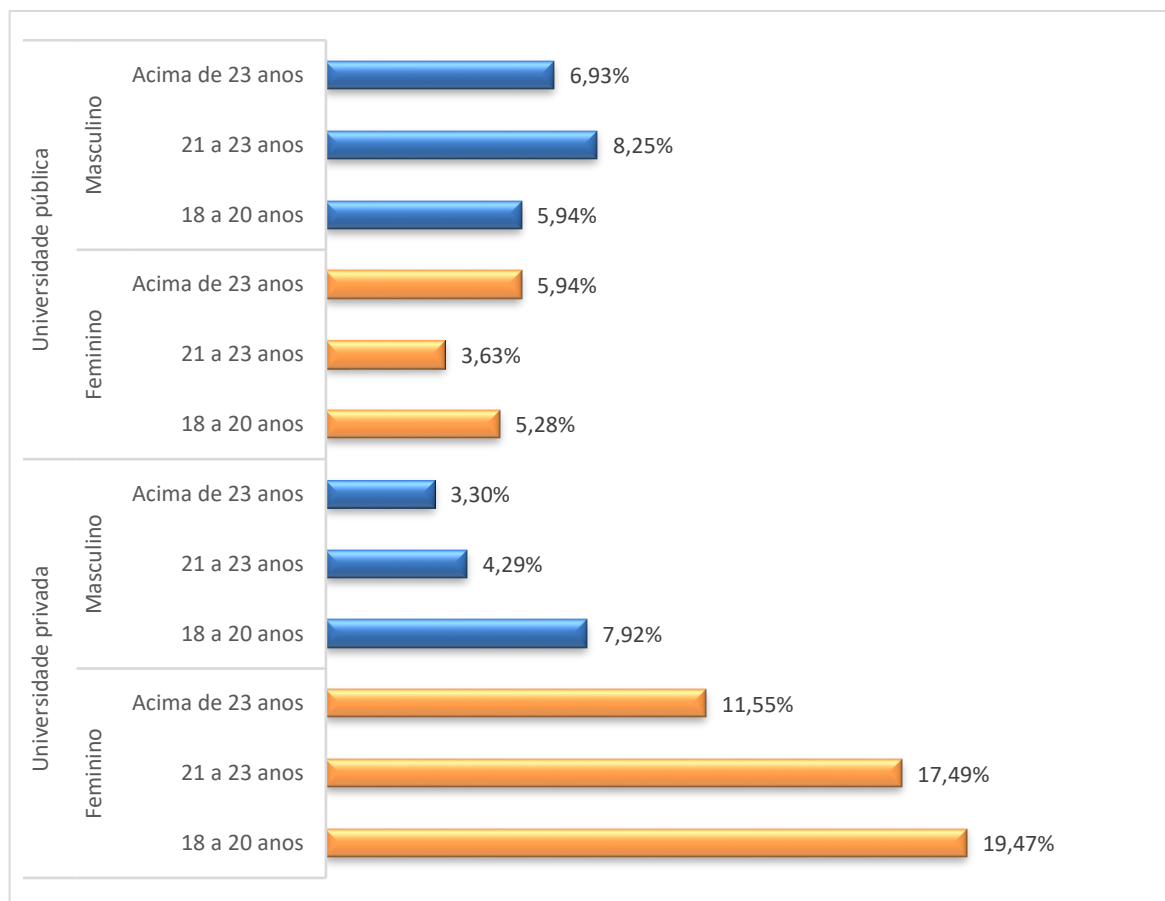
Figura 12 - Renda familiar de acordo com gênero e Universidade



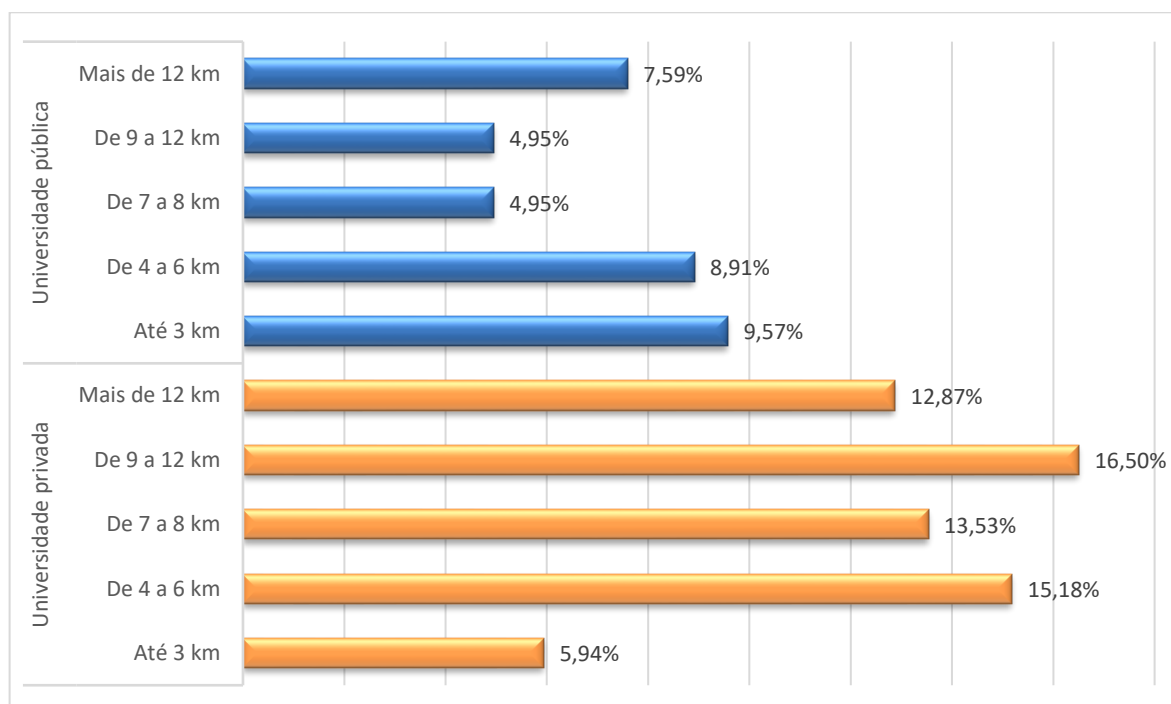
Na Figura 13 é apresentada a distribuição da idade dos respondentes de acordo com o gênero e a universidade, indicando que a maior parte da amostra é

representada por estudantes jovens com até 23 anos independente do gênero do respondente.

Figura 13 - Idade de acordo com gênero e Universidade



A Figura 14 representa a distribuição da variável distância percorrida pelos estudantes até a universidade, considerando a instituição de ensino em que estudam. Na universidade pública pode-se observar que muitos respondentes percorrem distâncias de até seis quilômetros, sendo isto explicado pelo fato desta universidade atrair muitos estudantes oriundos de outras cidades para Joinville, e estes estudantes acabam procurando residências mais próximas da instituição para morarem durante a sua graduação. Na universidade privada observa-se que os respondentes percorrem, em sua maior parte, trajetos acima de seis quilômetros, isto deve-se ao fato de que os estudantes desta instituição de ensino residem em Joinville e em municípios próximos.

Figura 14 - Distância percorrida de acordo com a Universidade

Na Tabela 3 são apresentadas as características da amostra separadas pelo modo de transporte utilizado pelos estudantes em seu deslocamento até a universidade. Com relação à instituição de ensino, os estudantes da instituição A (pública) fazem maior uso do modo ativo e o ônibus como modo de transporte, ambos com 33,03%, e o carro aparece em seguida com 26,60%. O maior uso do modo ativo na instituição pública pode ser explicado devido ao fato de muitos estudantes desta universidade, como já comentado anteriormente, serem de outros municípios, e acabam procurando residências mais próximas da instituição para morarem durante o período de graduação, assim, facilitando a escolha por modos ativos. Já na instituição B (privada), 42,27% dos estudantes utilizam o transporte público como modo de transporte, seguido pelo carro com 24,74% e o modo ativo 14,43%. O uso mais frequente do ônibus pelos alunos da instituição privada pode estar relacionado com o fato dos respondentes desta universidade serem mais jovens e do sexo feminino como já indicado anteriormente, assim, influenciando na maior escolha desse modo de transporte.

Os estudantes que moram em Joinville utilizam com maior frequência o transporte público 41,05%, com o carro e o modo ativo com percentuais de 24,91% e 22,46% respectivamente. Os alunos que residem em outra cidade utilizam o

transporte público fretado (ônibus fretado, vans) como modo de transporte 61,11%, seguido pelo carro com 33,33%.

Tabela 3 - Característica da amostra por modo de transporte

Variáveis	Modo de Transporte (%)				
	Carro	Modo ativo	Ônibus	Transporte público fretado	Outro
Instituição de ensino					
A	26,60	33,03	33,03	0	7,34
B	24,74	14,43	42,27	6,70	11,86
Gênero					
Feminino	25,52	17,71	42,71	3,65	10,42
Masculino	25,22	27,03	32,43	5,41	9,91
Idade (anos)					
Média	25,82	22,52	21,20	20,92	21,32
Desvio padrão	6,60	4,32	3,51	2,63	4,01
Renda familiar					
Até 2 salários mínimos	13,16	31,58	46,05	1,32	7,89
De 3 a 5 salários mínimos	24,65	16,20	40,84	5,63	12,68
De 6 a 9 salários mínimos	35,85	20,75	32,08	3,77	7,55
Mais de 10 salários mínimos	40,62	18,75	25,00	6,25	9,38
Estado civil					
Solteiro	20,78	23,14	40,78	4,71	10,59
Casado	48,84	11,63	30,23	2,32	6,98
Outro	60,00	20,00	20,00	0	0
Cidade onde mora					
Joinville	24,91	22,46	41,05	0,07	10,88
Outra	33,33	0	5,56	61,11	0
Tem filhos					
Sim	44,44	16,67	27,78	0	11,11
Não	24,21	21,40	39,65	4,56	10,18
Trabalho					
Sim	25,00	17,61	40,91	6,25	10,23
Não	25,98	25,98	36,22	1,58	10,24
Tempo de ida (min)					
Média	26,81	19,84	52,84	78,46	22,26
Desvio padrão	32,00	12,91	25,12	31,11	8,15
Distância de ida (km)					
Média	12,48	4,00	9,64	51,77	7,94
Desvio padrão	20,80	3,12	3,69	22,56	3,18

Os alunos do sexo masculino (32,43%) utilizam mais o transporte público, seguido pelo modo ativo (27,03%) e carro (25,22%). As alunas (42,27%) também fazem maior uso do transporte público, aparecendo na sequência o carro 25,52% e o modo ativo 17,71%. Com isto, observa-se que as mulheres utilizam mais o ônibus do que os homens, e estes utilizam mais o modo ativo em relação as mulheres. Com relação a renda familiar, estudantes com maiores rendas utilizam mais o carro, enquanto que os de rendas menores fazem maior uso do ônibus. Os estudantes

solteiros utilizam mais o ônibus, enquanto que os casados usam mais o carro. Já se o estudante tem filhos, faz maior uso do carro, caso contrário, utiliza mais o ônibus.

Aplicando-se os testes t e Wilcoxon para verificar se há diferença entre as médias das idades entre o uso dos diferentes modos de transporte, constatou-se que os estudante que utilizam o carro possuem a maior média com relação à idade, 25,82 anos (desvio padrão de 6,60 anos). Os testes indicam que a média da idade no uso do carro difere dos demais modos. Já entre os outros modos, os testes indicam que as suas médias não diferem entre si, ou seja, estatisticamente, são iguais. Os resultados dos testes t e Wilcoxon realizados entre os modos de transporte podem ser encontrados no Apêndice A.

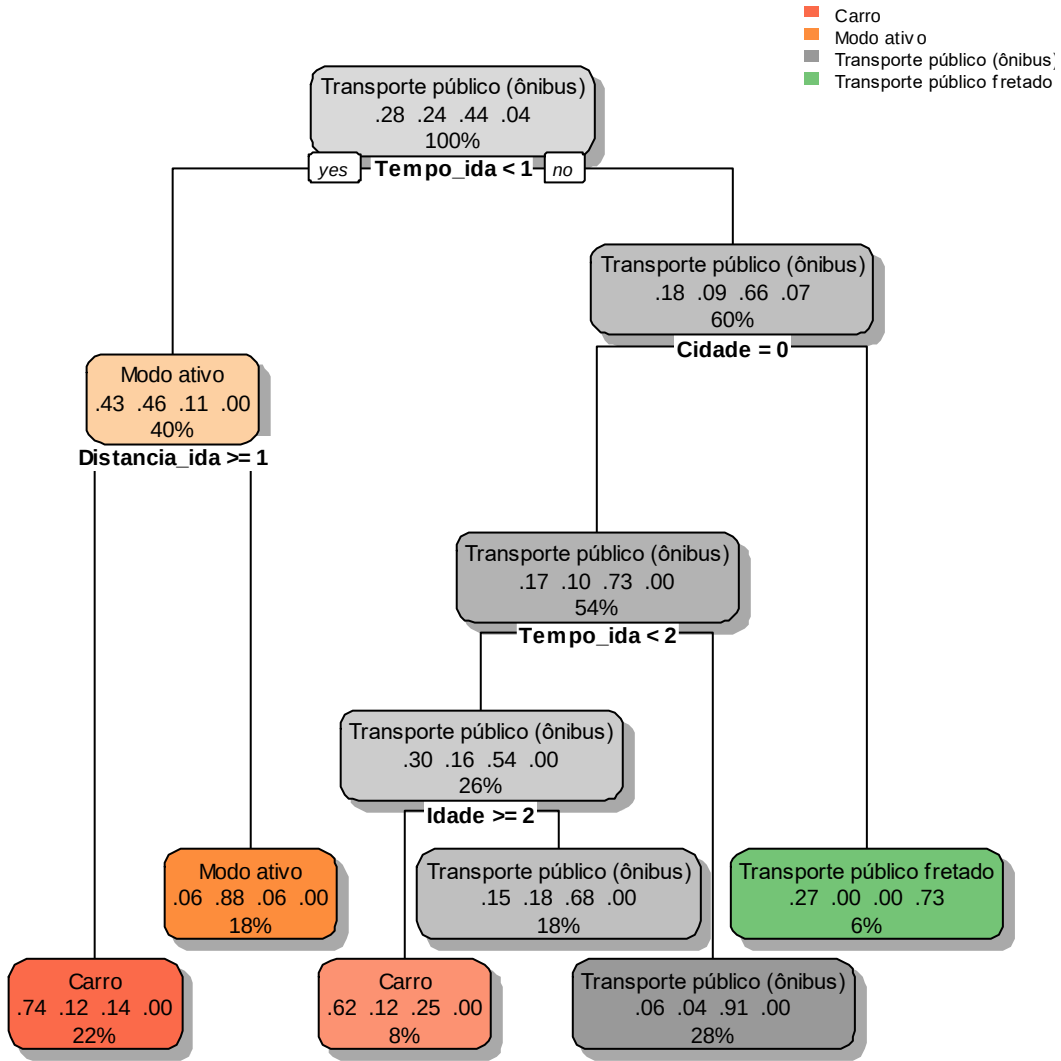
O tempo médio do ônibus é 52,84 minutos com desvio padrão de 25,12 minutos, e a distância média percorrida é 9,64 quilômetros e desvio padrão de 3,69 quilômetros. O tempo do carro apresentou média de 26,81 minutos e desvio padrão de 32,00 minutos, e a distância média foi de 12,48 quilômetros e desvio padrão de 20,80 quilômetros. Já no modo ativo o tempo médio foi 19,84 minutos e desvio padrão de 12,91 minutos, e a distância apresentou média de 4,00 quilômetros e desvio padrão 3,12 quilômetros. Comparando os resultados entre o uso do carro e do ônibus, pode-se observar que mesmo com os congestionamentos o carro ainda leva menos tempo de percurso.

Com relação ao modo ativo, pode-se constatar que o uso deste modo de transporte está associado à menores distâncias quando comparadas com os outros modos de transporte. De acordo com IEMA (2010), os deslocamentos de até cinco quilômetros são favoráveis ao uso da bicicleta, o que se aproxima ao valor médio encontrado nesta pesquisa, conforme Tabela 3.

4.2 RESULTADOS DAS ÁRVORES DE DECISÃO

Nessa seção são apresentadas as árvores de decisão obtidas para o Polo Universitário, a Universidade A e a Universidade B. A Figura 15 representa a árvore de decisão para o Polo Universitário, sendo possível observar que as variáveis representativas do modelo foram o tempo de deslocamento, a distância percorrida, a cidade onde mora e a idade.

Figura 15 - Árvore de decisão para o Polo Universitário

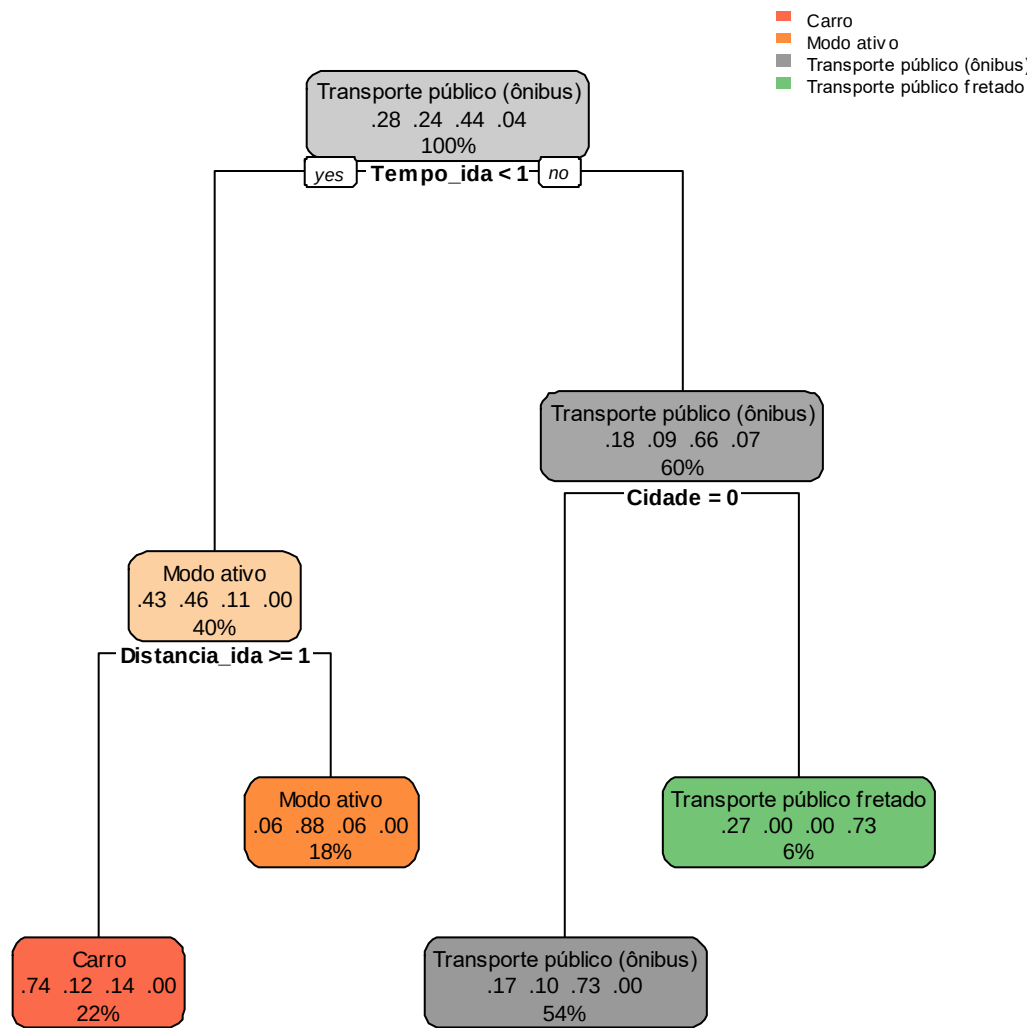


Já a Figura 16 trata-se da árvore de decisão obtida para o Polo Universitário após a realização da poda para obtenção de uma melhor classificação. Assim, o número de variáveis representativas foi reduzido, permanecendo o tempo, distância e cidade onde mora. O nó raiz que contém toda a base de dados, tem como modo de transporte predominante o transporte público (44% de probabilidade de escolha), seguido pelo carro (28%), modo ativo (24%) e transporte público fretado (4%).

A primeira variável de divisão da árvore podada foi o tempo (1 significa de 16 a 30 minutos de deslocamento); com isso, para o estudante que leva até 15 minutos de deslocamento até sua instituição de ensino, o nó apresentou uma nova divisão modal com o modo ativo como predominante (46%), seguido de perto pelo carro (43%), o transporte público (11%) e o transporte público fretado (0%). Neste nó, pode-se verificar que ocorre um aumento na escolha pelo modo ativo e o carro, e uma redução

pela escolha do transporte público e transporte público fretado. O outro nó gerado pela divisão do nó raiz, para o estudante que leva mais de 15 minutos até a universidade, apresentou a divisão modal com o transporte público como predominante para escolha (66%), seguido pelo carro (18%), o modo ativo (9%) e o transporte público fretado (7%). Assim, observa-se que ocorreu o aumento na escolha pelo transporte público e o transporte público fretado, e a redução na escolha pelo carro e o modo ativo. Diante de que foi exposto, pode-se concluir que o tempo sofre influência do modo utilizado, logo tempos maiores no deslocamento até o Polo Universitário indicam um aumento no uso do transporte público, e tempos menores sugerem um maior percentual no uso do carro e do modo ativo.

Figura 16 - Árvore de decisão para o Polo Universitário após a etapa de poda



Para os estudantes que levam até 15 minutos de deslocamento aparece outra divisão na árvore com a variável distância (1 significa de 4 a 6 quilômetros percorridos), sendo que para os estudantes que percorrem até 3 quilômetros foi gerado um novo nó, tendo como modo predominante para escolha o modo ativo (88%), seguido carro e transporte público ambos com 6%, e o transporte público fretado (0%). Neste nó, pode-se notar um aumento na escolha pelo modo ativo e uma redução na escolha pelo carro e o transporte público. O outro nó gerado pela divisão com a variável distância apresentou como modo predominante para escolha o carro (74%), seguido pelo transporte público (14%) e o modo ativo (12%). Com isso, observa-se um aumento na escolha pelo carro e o transporte público, e uma redução na escolha pelo modo ativo. Assim, pode-se notar que a variável distância influencia na escolha do modo de transporte para o Polo Universitário, indicando que o uso do modo ativo está associado a menores distâncias percorridas, e maiores distâncias favorecem o uso do carro e do transporte público.

Outra divisão presente na árvore é com a variável cidade (0 significa morar em Joinville), em que, para o estudante que possua um deslocamento superior a 15 minutos e more em Joinville, foi gerado um nó tendo como modo predominante para escolha o transporte público (73%), seguido pelo carro (17%), o modo ativo (10%) e o transporte público fretado (0%). Assim, observa-se um aumento na escolha do transporte público, e uma redução na escolha pelo transporte público fretado, sendo que a escolha pelo carro e o modo ativo tiveram uma pequena alteração, praticamente desprezível. O outro nó gerado, para os estudantes que não moram em Joinville, apresentou como modo predominante para escolha o transporte público fretado (73%), seguido pelo carro (27%), sendo observado um aumento na escolha por ambos os modos. O transporte público e o modo ativo não tiveram probabilidade de escolha, sugerindo que esses modos não são opções para o deslocamento dos estudantes que moram em outras cidades. Portanto, pode-se constatar que a variável cidade possui influência na escolha do modo transporte, indicando que estudantes que não residem em Joinville estão mais dispostos a utilizarem o transporte público fretado para seu deslocamento até o Polo Universitário.

Nas Tabelas 4 e 5 são apresentados, respectivamente, a matriz de confusão e os valores obtidos para avaliação do desempenho da amostra teste do modelo de árvore de decisão para o Polo Universitário. Na matriz de confusão observa-se que, no modelo, o modo de transporte com maior acerto é o ônibus, seguido pelo carro e o

modo ativo. O modelo apresentou uma boa acurácia de 0,7375 (73,75%) e um índice Kappa de 0,5989, indicando que este modelo tem uma concordância moderada. O resultado da sensibilidade indicou que o modelo acerta bem os verdadeiros positivos para o modo ônibus (0,8286) e carro (0,8261), e moderadamente para o modo ativo (0,5789). Para a especificidade, o modelo acerta bem os verdadeiros negativos para todos os modos, carro (0,8246), modo ativo (0,9344), ônibus (0,8444) e transporte público fretado (1,0000).

Tabela 4 - Característica da amostra por modo de transporte

Previsão / Real	Carro	Modo ativo	Ônibus	Transporte público fretado
Carro	19	4	3	3
Modo ativo	1	11	3	0
Ônibus	3	4	29	0
Transporte público fretado	0	0	0	0

Tabela 5 - Medidas de desempenho do modelo para o Polo Universitário

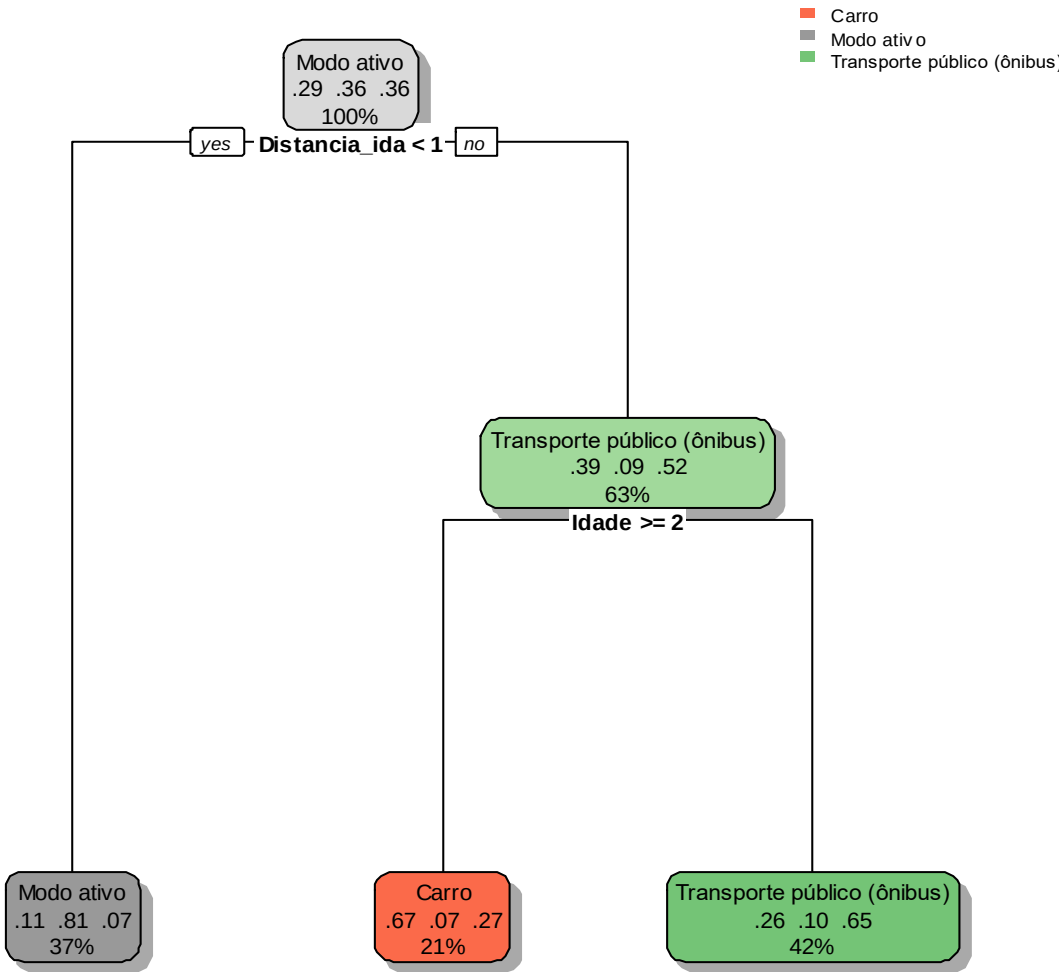
Medidas de desempenho	Modo de transporte			
	Carro	Modo ativo	Ônibus	Transporte público fretado
Sensibilidade	0,8261	0,5789	0,8286	0,0000
Especificidade	0,8246	0,9344	0,8444	1,0000
Acurácia	0,7375			
Índice Kappa	0,5989			

A Figura 17 representa a árvore de decisão para a Universidade A, sendo as variáveis representativas a distância e a idade. Neste modelo a realização da etapa da poda não alterou o esquema da árvore. Assim, o nó raiz, que contém toda a base de dados, tem como modo de transporte predominante o modo ativo e o carro, ambos com 36% de probabilidade de escolha, seguido pelo transporte público (29%). Neste modelo não apareceu o transporte público fretado, porque este modo de transporte não é característico nesta instituição de ensino, e sim na Universidade B

A primeira variável de divisão da árvore foi a distância (1 significa de 4 a 6 quilômetros percorridos), com isso, para o estudante que percorre até 3 quilômetros, o nó apresentou uma nova divisão modal com o modo ativo como predominante (81%), seguido pelo carro (11%), e o transporte público (7%). Neste nó, pode-se verificar que ocorre um aumento na escolha pelo modo ativo e uma redução pela escolha do carro e transporte público. O outro nó gerado pela divisão do nó raiz, para o estudante que percorre mais de 3 quilômetros até a universidade, apresentou a

divisão modal com o transporte público como predominante (52%), seguido pelo carro (39%), e o modo ativo (9%). Assim, observa-se que ocorreu um aumento na escolha pelo transporte público e o carro, e a redução na escolha pelo modo ativo. Dessa maneira, pode-se notar que a variável distância obteve influência na escolha do modo de transporte para a Universidade A, indicando que o uso do modo ativo está associado a menores distâncias percorridas, e maiores distâncias favorecem o uso do carro e do transporte público. Nesta instituição de ensino, é comum o uso do modo ativo no deslocamento diário, seja a pé ou de bicicleta, pois como já citado anteriormente, muitos estudantes residem próximos a universidade, facilitando a escolha deste meio de locomoção.

Figura 17 - Árvore de decisão para a Universidade A



Outra divisão que aparece na árvore é a variável idade (2 significa acima de 23 anos). Para os estudantes que percorrem mais de 3 quilômetros e possuam uma idade superior a 23 anos, foi gerado um nó tendo como modo predominante para escolha o carro (67%), seguido pelo transporte público (27%), e o modo ativo (7%). Assim, observa-se um aumento na escolha do carro, e uma redução na escolha do transporte público, sendo que a escolha do modo ativo teve uma pequena alteração. O outro nó gerado, para os estudantes que possuem até 23 anos, apresentou como modo predominante para escolha o transporte público (65%), seguido pelo carro (26%) e o modo ativo (10%), sendo observado um aumento na escolha do transporte público e uma redução na escolha do carro. Com relação ao modo ativo, a alteração ocorrida na escolha por este modo foi praticamente desprezível. Dessa maneira, pode-se concluir que a variável idade possui relevância na escolha do modo de transporte na Universidade A, sugerindo que estudantes mais velhos possuem certa tendência ao uso do carro como modo de transporte.

Nas Tabelas 6 e 7 são apresentados, respectivamente, a matriz de confusão e os valores obtidos para avaliação do desempenho da amostra teste do modelo de árvore de decisão para a Universidade A. Na matriz de confusão, observa-se que, no modelo, o modo de transporte com maior acerto é o modo ativo, seguido pelo ônibus e o carro. O modelo apresentou uma acurácia de 0,6429 (64,29%) e um índice Kappa de 0,4553, indicando que este modelo tem uma concordância moderada. O resultado da sensibilidade indicou que o modelo tem um bom acerto dos verdadeiros positivos para o modo ativo (0,7000) e para o modo ônibus (0,8000) e baixo para o carro (0,3750). Para a especificidade o modelo acerta bem os verdadeiros negativos para os modos carro (0,9000), modo ativo (0,6667) e ônibus (0,8889).

Tabela 6 - Matriz de confusão do modelo para a Universidade A

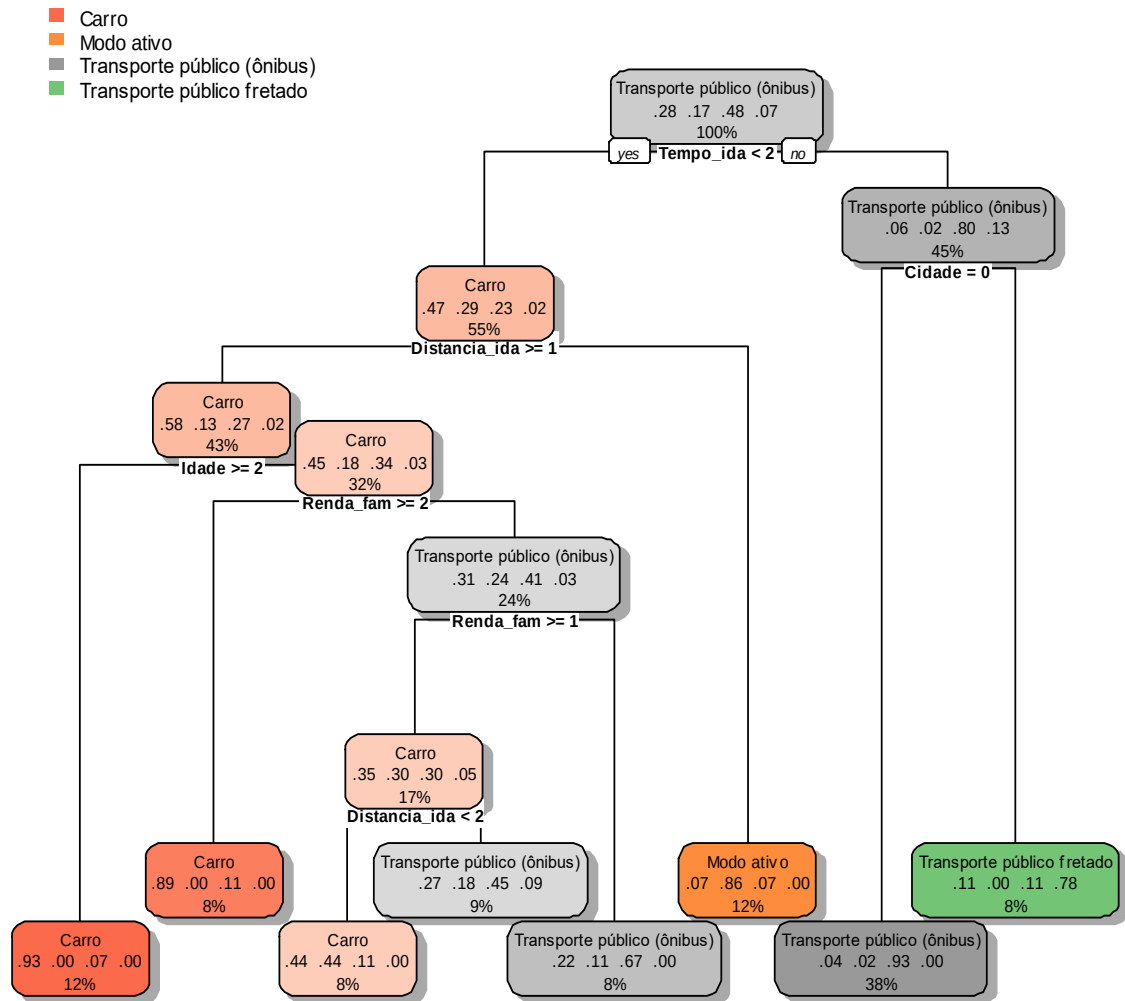
Previsão / Real	Carro	Modo ativo	Ônibus
Carro	3	2	0
Modo ativo	4	7	2
Ônibus	1	1	8

Tabela 7 - Medidas de desempenho do modelo para a Universidade A

Medidas de desempenho	Modo de transporte		
	Carro	Modo ativo	Ônibus
Sensibilidade	0,3750	0,7000	0,8000
Especificidade	0,9000	0,6667	0,8889
Acurácia	0,6429		
Índice Kappa	0,4553		

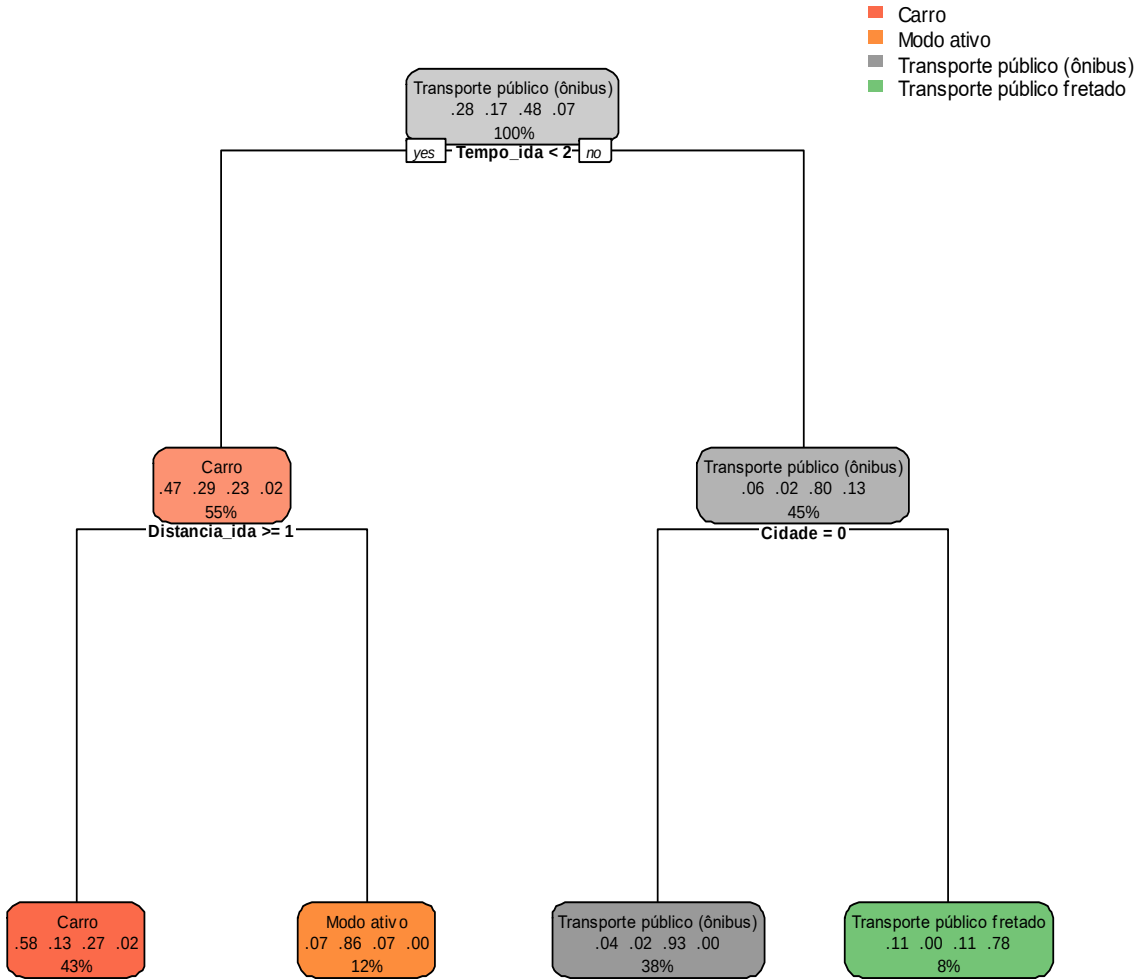
A Figura 18 representa a árvore de decisão para a Universidade B, possibilitando observar que as variáveis representativas do modelo foram o tempo, a distância, a cidade onde mora, a idade e a renda familiar.

Figura 18 - Árvore de decisão para a Universidade B



Já a Figura 19 trata-se da árvore de decisão obtida para a Universidade B após a realização da poda para obtenção de uma melhor classificação; assim, o número de variáveis representativas foi reduzido, permanecendo tempo, distância e cidade onde mora. Com isto, o nó raiz, que contém toda a base de dados, tem como modo de transporte predominante o transporte público (48% de probabilidade de escolha), seguido pelo carro (28%), modo ativo (17%) e transporte público fretado (7%).

Figura 19 - Árvore de decisão para a Universidade B após a etapa de poda



A primeira variável de divisão da árvore podada foi o tempo (2 significa de 31 a 60 minutos de deslocamento), com isso, para o estudante que leva até 30 minutos de deslocamento até sua instituição de ensino, o nó apresentou uma nova divisão modal com o carro como predominante (47%), seguido de perto pelo modo ativo (29%), o transporte público (23%) e o transporte público fretado (2%). Neste nó, pode-se verificar que ocorre um aumento na escolha pelo carro e o modo ativo, e uma redução pela escolha do transporte público e transporte público fretado. O outro nó gerado pela divisão do nó raiz, para o estudante que leva mais de 30 minutos até a universidade, apresentou a divisão modal com o transporte público como predominante para escolha (80%), seguido pelo transporte público fretado (13%), o carro (6%) e o modo ativo (2%). Dessa maneira, observa-se que ocorreu o aumento na escolha pelo transporte público e o transporte público fretado, e a redução na escolha pelo carro e o modo ativo. Assim, como comentado na árvore do Polo Universitário, tempos

maiores no deslocamento até a Universidade B indicam um aumento no uso do transporte público, e tempos menores sugerem um maior percentual no uso do carro e do modo ativo.

Para os estudantes que levam até 30 minutos de deslocamento, aparece outra divisão com a variável distância (1 significa de 4 a 6 quilômetros percorridos), sendo que, para os estudantes que percorrem até 3 quilômetros foi gerado um novo nó, tendo como modo predominante para escolha o modo ativo (86%), seguido do carro e transporte público, ambos com 7%, e o transporte público fretado (0%). Neste nó pode-se notar um aumento na escolha pelo modo ativo e uma redução na escolha pelo carro e o transporte público. O outro nó gerado pela divisão com a variável distância apresentou como modo predominante para escolha o carro (58%), seguido pelo transporte público (27%), o modo ativo (13%), e o transporte público fretado (2%). Com isso, observa-se um aumento na escolha pelo carro e o transporte público, e uma redução na escolha pelo modo ativo. Assim, pode-se notar que a variável distância influencia na escolha do modo de transporte para a Universidade, assim como foi constatado na árvore do Polo Universitário, sugerindo que menores distâncias percorridas, favorecem o uso do modo ativo, enquanto que maiores distâncias propiciam maior escolha do carro e do transporte público.

Outra divisão na árvore é com a variável cidade (0 significa morar em Joinville), em que para o estudante que possua um deslocamento superior a 30 minutos e more em Joinville, foi gerado um nó tendo como modo predominante para escolha o transporte público (93%), seguido pelo carro (4%), o modo ativo (2%) e o transporte público fretado (0%). Assim, observa-se um aumento na escolha do transporte público, e uma redução na escolha pelo transporte público fretado, sendo que a escolha pelo carro e o modo ativo tiveram uma pequena alteração, praticamente desprezível. O outro nó gerado, para os estudantes que não moram em Joinville, apresentou como modo predominante para escolha o transporte público fretado (78%), seguido pelo carro e o transporte público, ambos com 11%. Com isso, verificou-se um aumento na escolha pelo transporte público fretado e o carro. O modo ativo não teve probabilidade de escolha, sugerindo que esse modo não é opção para o deslocamento dos estudantes que moram em outras cidades. O transporte público apresentou uma redução na escolha, mas não deveria apresentar probabilidade de escolha, assim como na árvore do Polo Universitário. Portanto, pode-se constatar que a variável cidade possui influência na escolha do modo transporte, assim com indicado

anteriormente na árvore do Polo Universitário, sugerindo que estudantes que não residem em Joinville, estão mais dispostos a utilizarem o transporte público fretado para seu deslocamento até a Universidade B.

Nas Tabelas 8 e 9 são apresentados, respectivamente, a matriz de confusão e os valores obtidos para avaliação do desempenho da amostra teste do modelo de árvore de decisão para a Universidade B. Na matriz de confusão observa-se que no modelo o modo de transporte com maior acerto é o ônibus, seguido pelo carro e o transporte público fretado. O modelo apresentou uma boa acurácia de 0,7143 (71,43%) e um índice Kappa de 0,5418, indicando que este modelo tem uma concordância moderada. O resultado da sensibilidade indicou que o modelo acerta bem os verdadeiros positivos para o modo ônibus (0,8462) e transporte público fretado (1,0000), moderadamente para o carro (0,6429), e um baixo acerto para o modo ativo (0,3750). Para a especificidade, o modelo acerta bem os verdadeiros negativos para os modos carro (0,8571), modo ativo (1,000) e transporte público fretado (0,9783), e moderadamente para o modo ônibus (0,6800).

Tabela 8 - Matriz de confusão do modelo para a Universidade B

Previsão / Real	Carro	Modo ativo	Ônibus	Transporte público fretado
Carro	9	1	4	0
Modo ativo	0	3	0	0
Ônibus	4	4	20	0
Transporte público fretado	1	0	0	3

Tabela 9 - Medidas de desempenho do modelo para a Universidade B

Medidas de desempenho	Modo de transporte			
	Carro	Modo ativo	Ônibus	Transporte público fretado
Sensibilidade	0,6429	0,3750	0,8333	1,0000
Especificidade	0,8571	1,0000	0,6800	0,9783
Acurácia	0,7143			
Índice Kappa	0,5418			

4.3 RESULTADOS DOS MODELOS DE REGRESSÃO MULTINOMIAL

Nessa seção são apresentados os modelos de regressão multinomial obtidos para o Polo Universitário, a Universidade A e a Universidade B. Na Tabela 10 são indicadas as descrições das variáveis utilizadas para criação dos modelos de regressão logística multinomial, com suas respectivas categorias de referência.

Tabela 10 - Descrição das variáveis utilizadas nos modelos

Variáveis	Categorias
Dependente:	
Modo de transporte diário (Modo_transp)	Carro – referência Modo ativo Transporte público (ônibus) Transporte público fretado
Independentes:	
Gênero (Sexo)	Feminino – referência Masculino
Idade	De 18 a 20 anos – referência De 21 a 23 anos Acima de 23 anos
Renda familiar (Renda_fam)	Até 2 salários mínimos – referência De 3 a 5 salários mínimos De 6 a 9 salários mínimos Mais de 10 salários mínimos
Cidade onde mora (Cidade)	Joinville – referência Outro
Tempo (tempo_ida)	Até 15 min – referência De 16 min a 30 min De 31 min a 60 min Mais de 60 min
Distancia (distancia_ida)	Até 3 km – referência De 4 a 6 km De 7 a 8 km De 9 a 12 km Mais de 12 km

Os modelos de regressão logística multinomial foram treinados para a escolha do modo de transporte, tendo esta variável as categorias carro (referência), modo ativo, ônibus e transporte público fretado. Foram ajustados diversos modelos, servindo como testes para a seleção das variáveis independentes a ser utilizados para criação dos modelos finais. Ao final optou-se pela escolha de seis modelos, sendo dois para o Polo Universitário, dois para a Universidade A (pública) e dois para a Universidade B (privada). Com relação a Universidade A, optou-se pela retirada do modo de transporte público fretado, modo não presente no modelo de árvore de decisão para esta universidade, assim elaborando apenas um modelo, pelo fato do transporte público fretado não ter sido indicado por nenhum estudante desta universidade na amostra.

Nas Tabelas 11, 14 e 17 são apresentados respectivamente, os modelos selecionados para o Polo Universitário (Modelos 1 e 2), Universidade A (Modelos 3 e 4) e Universidade B (Modelos 5 e 6), com os coeficientes e significância das variáveis escolhidas. A diferença entre os dois modelos, para cada caso analisado, é por em

um estar presente a variável distância e no outro, a variável tempo. Optou-se em selecionar para cada modelo apenas uma destas variáveis, pois, segundo Ton *et al.* (2019), a distância e o tempo de viagem são características altamente correlacionadas e às vezes até consideradas equivalentes.

Para a seleção do melhor modelo, foram analisados os valores da acurácia e índice Kappa gerados pela matriz de confusão, assim como os valores dos pseudo- r^2 de McFadden, Cox e Snell, e Nagelkerke. Para a elaboração da matriz de confusão, foi utilizada a amostra de treinamento, sendo utilizado para os demais testes de avaliação, a amostra de teste.

Na Tabela 11 analisando os dois modelos identifica-se que no nível de $\alpha = 5\%$, a variável “Tempo” é significativa para o ônibus e transporte público fretado, e o fator “Distancia” para o modo ativo. No modelo escolhido (Modelo 2), a idade foi significativa para o ônibus e transporte público fretado no nível de $\alpha = 5\%$, o gênero significativo para o modo ativo no nível de $\alpha = 10\%$, a renda familiar significativa para o modo ativo e ônibus no nível de $\alpha = 10\%$, e a cidade onde mora significativa para o modo ativo, ônibus e transporte público fretado no nível de $\alpha = 5\%$.

Para o Polo Universitário, o Modelo 2 foi o escolhido, pois, conforme a Tabela 12, apresenta o maior valor de acurácia e índice Kappa, indicando uma concordância substancial. Também neste modelo foi obtido os maiores valores dos pseudo- r^2 , e o menor valor AIC. A análise multivariada revelou que todas as variáveis categóricas do Modelo 2 são significativas no nível de $\alpha = 5\%$, “Tempo” ($\chi^2 = 98,41$; p -valor = 0,00), “Idade” ($\chi^2 = 33,08$; p -valor = 0,00), “Sexo” ($\chi^2 = 13,12$; p -valor = 0,00), “Renda” ($\chi^2 = 24,17$; p -valor = 0,00), e “Cidade” ($\chi^2 = 49,76$; p -valor = 0,00).

Tabela 11 - Comparação dos modelos para o Polo Universitário

Variáveis	Modelo 1			Modelo 2		
	Coeficientes	Erro padrão	p-valor	Coeficientes	Erro padrão	p-valor
Interceptos: Carro	Referência					
Modo ativo	5,713	2,033	0,005***	1,413	0,665	0,034**
Ônibus	3,898	2,016	0,053*	0,177	0,830	0,831
Transporte público fretado	-442,711	246,765	0,073*	-277,702	307,366	0,366
Distancia: Até 3 km	Referência					
De 4 a 6 km						
Modo ativo	-5,733	1,980	0,004***			
Ônibus	-2,412	1,972	0,221			
Transporte público fretado	-94,609					
De 7 a 8 km						
Modo ativo	-6,698	2,021	0,001***			
Ônibus	-1,874	1,969	0,341			

Transporte público fretado	-220,240					
De 9 a 12 km						
Modo ativo	-6,271	2,037	0,002***			
Ônibus	-1,009	1,962	0,607			
Transporte público fretado	64,393	123,384	0,602			
Mais de 12 km						
Modo ativo	-6,278	2,100	0,003***			
Ônibus	-1,171	1,977	0,554			
Transporte público fretado	-26,677	370,147	0,943			
Tempo: Até 15 min	Referência					
De 16 min a 30 min						
Modo ativo				-0,697	0,484	0,150
Ônibus				1,199	0,671	0,074*
Transporte público fretado				53,540	0,022	0,000***
De 31 min a 60 min						
Modo ativo				0,282	0,780	0,718
Ônibus				4,625	0,888	0,000***
Transporte público fretado				103,378	384,623	0,788
Mais de 60 min						
Modo ativo				-37,493		
Ônibus				6,268	1,875	0,001***
Transporte público fretado				152,774	77,280	0,048
Idade: 18 a 20 anos	Referência					
De 21 a 23 anos						
Modo ativo	0,208	0,703	0,767	0,081	0,575	0,888
Ônibus	-0,482	0,506	0,341	-0,223	0,583	0,702
Transporte público fretado	-155,063			-34,376	4,028	0,000***
Acima de 23 anos						
Modo ativo	-0,266	0,706	0,706	-0,843	0,565	0,136
Ônibus	-1,945	0,527	0,000***	-2,635	0,713	0,000***
Transporte público fretado	-341,404	493,530	0,489	-126,285	76,711	0,010*
Sexo: Feminino	Referência					
Masculino						
Modo ativo	1,470	0,590	0,013**	0,857	0,477	0,072*
Ônibus	-0,043	0,451	0,924	-0,623	0,564	0,269
Transporte público fretado	2,265	1,660	0,173	29,297	89,597	0,744
Renda: Até 2 salários	Referência					
De 3 a 5 salários						
Modo ativo	-1,007	0,724	0,165	-1,552	0,611	0,011**
Ônibus	-0,915	0,567	0,107	-0,979	0,657	0,136
Transporte público fretado	325,808	493,530	0,509	73,347	76,689	0,339
De 6 a 9 salários						
Modo ativo	-1,642	0,893	0,066*	-1,608	0,712	0,024**
Ônibus	-1,341	0,679	0,048**	-1,365	0,794	0,086*
Transporte público fretado	178,494	0,000	0,000***	29,404	0,000	0,000***
Mais de 10 salários						
Modo ativo	-2,242	1,025	0,029**	-1,918	0,716	0,007***
Ônibus	-2,290	0,741	0,002***	-1,371	0,821	0,095*
Transporte público fretado	377,527	123,384	0,002***	163,565	384,054	0,670
Cidade: Joinville	Referência					
Outra						
Modo ativo	-440,992	0,000	0,000***	-217,580	0,000	0,000***
Ônibus	-2,123	1,399	0,129	-3,855	1,694	0,023**
Transporte público fretado	484,217	370,147	0,191	178,173	76,689	0,020**
Significância	*** p-valor<0,01; ** p-valor<0,05; * p-valor<0,10					

Tabela 12 - Medidas de avaliação dos modelos para o Polo Universitário

Medidas de avaliação	Modelo 1				Modelo 2			
	Carro	Modo ativo	Ônibus	Transporte público fretado	Carro	Modo ativo	Ônibus	Transporte público fretado
Sensibilidade	0,5556	0,5778	0,8313	0,9000	0,7037	0,4000	0,9157	1,0000
Especificidade	0,8913	0,9524	0,6789	0,9945	0,8768	0,9252	0,8073	0,9945
Acurácia	0,6979				0,7396			
Índice Kappa	0,5402				0,6070			
Pseudo-R ²								
McFadden	0,5158				0,6343			
Cox & Snell	0,7053				0,7775			
Nagelkerke	0,7781				0,8577			
AIC	336,637				317,333			

Na matriz de confusão (Tabela 13), observa-se que no Modelo 2, o modo de transporte com maior acerto é o transporte particular, seguido pelo ônibus e o carro. Na Tabela 12, o resultado da sensibilidade indicou que o modelo acerta bem os verdadeiros positivos para o modo transporte público fretado (1,0000), ônibus (0,9157) e carro (0,7037), e tem um baixo acerto para o modo ativo (0,4000). Para a especificidade o modelo acerta bem os verdadeiros negativos para todos os modos.

Tabela 13 - Matriz de confusão do Modelo 2 para o Polo Universitário

Previsão / Real	Carro	Modo ativo	Ônibus	Transporte público fretado
Carro	38	13	4	0
Modo ativo	8	18	3	0
Ônibus	7	14	76	0
Transporte público fretado	1	0	0	10

Na Tabela 14, analisando o modelo escolhido (Modelo 3), identifica-se que no nível de $\alpha = 5\%$, a variável “Distancia” é significativa para o modo ativo. Com relação a idade, esta é significativa para o ônibus no nível de $\alpha = 5\%$

Para a Universidade A, o Modelo 3 foi o escolhido, pois, conforme a Tabela 15, apresenta o maior valor de acurácia e índice Kappa, indicando uma concordância moderada. Também neste modelo foi obtido os maiores valores dos pseudo-r². A análise multivariada revelou que todas as variáveis categóricas do Modelo 3 são significativas no nível de $\alpha = 5\%$, “Distancia” ($\chi^2 = 42,81$; p -valor = 0,00), e “Idade” ($\chi^2 = 18,83$; p -valor = 0,00).

Tabela 14 - Comparação dos modelos para a Universidade A

Variáveis	Modelo 3			Modelo 4		
	Coeficientes	Erro padrão	p-valor	Coeficientes	Erro padrão	p-valor
Interceptos: Carro	Referência					
Modo ativo	2,775	1,028	0,007***	1,344	0,740	0,069*
Ônibus	0,633	1,193	0,579	-1,365	1,267	0,281
Distancia: Até 3 km	Referência					
De 4 a 6 km						
Modo ativo	-2,569	0,954	0,007***			
Ônibus	-0,756	1,270	0,552			
De 7 a 8 km						
Modo ativo	-2,176	1,155	0,060*			
Ônibus	1,080	1,339	0,420			
De 9 a 12 km						
Modo ativo	-2,495	1,425	0,080*			
Ônibus	1,853	1,402	0,186			
Mais de 12 km						
Modo ativo	-15,484	355,838	0,965			
Ônibus	0,584	1,253	0,644			
Tempo: Até 15 min	Referência					
De 16 min a 30 min						
Modo ativo				-1,099	0,707	0,120
Ônibus				1,738	1,262	0,168
De 31 min a 60 min						
Modo ativo				-1,162	1,327	0,381
Ônibus				4,552	1,516	0,003***
Mais de 60 min						
Modo ativo				-0,533	1,515	0,725
Ônibus				4,395	1,724	0,011**
Idade: 18 a 20 anos	Referência					
De 21 a 23 anos						
Modo ativo	-0,070	1,107	0,950	0,146	0,988	0,883
Ônibus	0,372	0,980	0,705	0,550	1,107	0,619
Acima de 23 anos						
Modo ativo	-1,574	0,988	0,111	-1,274	0,812	0,117
Ônibus	-2,925	0,976	0,003***	-3,107	1,160	0,007***
Significância	*** p-valor<0,01; ** p-valor<0,05; * p-valor<0,10					

Tabela 15 - Medidas de avaliação do modelo para a Universidade A

Medidas de avaliação	Modelo 3			Modelo 4		
	Carro	Modo ativo	Ônibus	Carro	Modo ativo	Ônibus
Sensibilidade	0,5714	0,7692	0,8077	0,3810	0,6923	0,9231
Especificidade	0,8846	0,8511	0,8511	0,9423	0,7872	0,7872
Acurácia		0,7260			0,6849	
Índice Kappa		0,5853			0,5184	
Pseudo-R ²						
McFadden		0,4220			0,2887	
Cox & Snell		0,6026			0,4681	
Nagelkerke		0,6789			0,5273	
AIC		125,859			126,504	

Na matriz de confusão (Tabela 16) observa-se que no Modelo 3, o modo de transporte com maior acerto é o ônibus, seguido pelo modo ativo. Na Tabela 14, o resultado da sensibilidade indicou que o modelo acerta bem os verdadeiros positivos

para o modo ativo (0,7692) e o ônibus (0,8077), e tem um acerto moderado para o carro (0,5714). Para a especificidade o modelo acerta bem os verdadeiros negativos para todos os modos.

Tabela 16 - Matriz de confusão do Modelo 3 para a Universidade A

Previsão / Real	Carro	Modo ativo	Ônibus
Carro	12	3	3
Modo ativo	5	20	2
Ônibus	4	3	21

Na Tabela 17, analisando modelo escolhido (Modelo 6), identifica-se que no nível de $\alpha = 5\%$, as variáveis “Tempo”, “Idade” e “Renda”, são significativas para todos os modos. Com relação ao gênero, o modelo foi significativo para o modo ativo no nível de $\alpha = 5\%$.

Para a Universidade B, o Modelo 6 foi o escolhido, pois conforme a Tabela 18, apresenta o maior valor de acurácia e índice Kappa, indicando uma concordância substancial. Também neste modelo foi obtido os maiores valores dos pseudo- r^2 , e o menor valor AIC. A análise multivariada revelou que todas as variáveis categóricas do Modelo 6 são significativas no nível de $\alpha = 5\%$, “Tempo” ($\chi^2 = 73,37$; p -valor = 0,00), “Idade” ($\chi^2 = 25,53$; p -valor = 0,00), “Sexo” ($\chi^2 = 10,67$; p -valor = 0,01), “Renda” ($\chi^2 = 17,29$; p -valor = 0,04), e “Cidade” ($\chi^2 = 30,53$; p -valor = 0,00).

Tabela 17 - Comparação dos modelos para a Universidade B

Variáveis	Modelo 5			Modelo 6		
	Coeficientes	Erro padrão	p-valor	Coeficientes	Erro padrão	p-valor
Interceptos: Carro	Referência					
Modo ativo	69,862	1,160	0,000***	1,975	1,047	0,059*
Ônibus	68,274	0,924	0,000***	0,896	1,125	0,426
Transporte público fretado	-81,150	1,176	0,000***	-50,565	0,986	0,000***
Distancia: Até 3 km	Referência					
De 4 a 6 km						
Modo ativo	-71,010	1,180	0,000***			
Ônibus	-65,505	0,805	0,000***			
Transporte público fretado	79,983	1,278	0,000***			
De 7 a 8 km						
Modo ativo	-71,291	1,341	0,000***			
Ônibus	-64,379	0,851	0,000***			
Transporte público fretado	-13,402	0,000	0,000***			
De 9 a 12 km						
Modo ativo	-149,514	0,000	0,000***			
Ônibus	-64,734	0,802	0,000***			
Transporte particular	41,957	0,761	0,000***			
Mais de 12 km						
Modo ativo	-70,211	1,365	0,000***			

Ônibus	-64,533	0,861	0,000***			
Transporte público fretado	29,418	0,717	0,000***			
Tempo: Até 15 min	Referência					
De 16 min a 30 min						
Modo ativo				-1,888	0,809	0,020**
Ônibus				1,331	0,959	0,165
Transporte público fretado				33,414	0,900	0,000***
De 31 min a 60 min						
Modo ativo				-1,151	1,260	0,361
Ônibus				4,235	1,126	0,000***
Transporte público fretado				34,854	0,978	0,000***
Mais de 60 min						
Modo ativo				0,372	0,000	0,000***
Ônibus				85,827	0,000	0,000***
Transporte público fretado				93,419	0,000	0,000***
Idade: 18 a 20 anos	Referência					
De 21 a 23 anos						
Modo ativo	0,532	1,404	0,705	-0,400	0,906	0,659
Ônibus	-2,259	0,745	0,002***	-1,831	0,784	0,020**
Transporte público fretado	-1,469	1,649	0,373	0,192	2,607	0,941
Acima de 23 anos						
Modo ativo	-3,133	2,072	0,131	-3,124	1,208	0,010***
Ônibus	-2,995	0,837	0,000***	-3,332	0,990	0,001***
Transporte público fretado	-2,368	1,739	0,173	-17,008	0,000	0,000***
Sexo: Feminino	Referência					
Masculino						
Modo ativo	2,261	1,150	0,049**	2,332	0,972	0,017**
Ônibus	0,254	0,703	0,718	0,099	0,857	0,908
Transporte público fretado	1,731	1,296	0,182	2,729	1,560	0,080*
Renda: Até 2 salários	Referência					
De 3 a 5 salários						
Modo ativo	0,188	1,118	0,866	-0,535	0,846	0,527
Ônibus	-0,861	0,684	0,208	-0,600	0,810	0,459
Transporte público fretado	0,092	2,292	0,968	14,490	1,585	0,000***
De 6 a 9 salários						
Modo ativo	-4,130	2,526	0,102	-4,056	1,531	0,008***
Ônibus	-2,550	0,921	0,006***	-2,339	1,064	0,028**
Transporte público fretado	-0,287	2,768	0,917	-15,988	0,000	0,000***
Mais de 10 salários						
Modo ativo	-2,910	3,154	0,356	-2,360	1,272	0,063*
Ônibus	-2,403	1,048	0,022**	-1,654	1,087	0,128
Transporte público fretado	38,786	0,761	0,000***	15,029	1,170	0,000***
Cidade: Joinville	Referência					
Outra						
Modo ativo	-87,218	0,000	0,000***	-40,894		
Ônibus	-103,198	0,000	0,000***	-81,976	0,000	0,000***
Transporte público fretado	53,654	0,717	0,000***	0,384	1,942	0,843
Significância	*** p-valor<0,01; ** p-valor<0,05; * p-valor<0,10					

Na matriz de confusão (Tabela 19) observa-se que, no Modelo 6, o modo de transporte com maior acerto é o ônibus, seguido pelo transporte público fretado e o carro. Na Tabela 18, o resultado da sensibilidade indicou que o modelo acerta bem os verdadeiros positivos para o modo transporte público fretado (0,8000), ônibus (0,8966) e carro (0,7353), e tem um acerto moderado para o modo ativo (0,6000). Para a especificidade, o modelo acerta bem os verdadeiros negativos para todos os modos.

Tabela 18 - Medidas de avaliação dos modelos para a Universidade B

Medidas de avaliação	Modelo 5				Modelo 6			
	Carro	Modo ativo	Ônibus	Transporte público fretado	Carro	Modo ativo	Ônibus	Transporte público fretado
Sensibilidade	0,5294	0,7500	0,8448	0,7000	0,7353	0,6000	0,8966	0,8000
Especificidade	0,8750	0,9804	0,7188	0,9821	0,8864	0,9412	0,8594	1,0000
Acurácia	0,7295				0,7951			
Índice Kappa	0,5798				0,6861			
Pseudo-R ²								
McFadden	0,5897				0,7981			
Cox & Snell	0,7497				0,8466			
Nagelkerke	0,8289				0,9360			
AIC	224,205				205,963			

Tabela 19 - Matriz de confusão do Modelo 6 para a Universidade B

Previsão / Real	Carro	Modo ativo	Ônibus	Transporte público fretado
Carro	25	5	4	1
Modo ativo	4	12	2	0
Ônibus	5	3	52	1
Transporte público fretado	0	0	0	8

As razões de chances (*odds ratio*) para os coeficientes significativos dos modelos escolhidos são mostradas na Tabela 20. Caso a razão de chances apresente um valor menor que um, indica ser mais provável a escolha da categoria de referência, o carro. Analisando os resultados para o Polo Universitário, observa-se que estudantes que levam tempos maiores de viagem até a Universidade tem mais chance de utilizar o ônibus do que o carro.

Com relação a idade, estudantes mais velhos (acima de 23 anos) tem maiores chances de utilizarem o carro como modo de transporte. O estudo desenvolvido por Ferreira, Pitombo e Cunha (2016), indicou que usuários mais jovens, com idade de até 30 anos, possuíam menor tendência à utilização do carro como modo de transporte. Também é observado que os homens têm maiores chances de utilizarem o modo ativo do que as mulheres. Zhan *et al.* (2016), verificou que os estudantes do sexo masculino utilizam o modo ativo com mais frequência do que as alunas em suas viagens diárias.

Relativo à renda familiar, conforme ela aumenta, menores são as chances do uso do modo ativo e do ônibus, aumentando a escolha pelo uso do carro. Na pesquisa realizada por Pitombo e Gomes (2014), foi observado que quanto maior a renda familiar, maior a probabilidade de escolher o carro como modo de deslocamento. Por

fim, pode-se constatar que os estudantes que não residem em Joinville, tem maiores chances de utilizarem o transporte público fretado como modo de transporte.

Tabela 20 - Razão de Chance para os modelos escolhidos

Variáveis significativas	Modelo 2 Polo Universitário	Modelo 3 Universidade A	Modelo 6 Universidade B
	Razão de Chance	Razão de Chance	Razão de Chance
Distancia: Até 3 km	Referência		
De 4 a 6 km			
Modo ativo		0,077	
Tempo: Até 15 min	Referência		
De 16 min a 30 min			
Modo ativo			0,151
De 31 min a 60 min			
Ônibus	1,02		
Mais de 60 min			
Modo ativo			
Ônibus	5,27		1,450
Transporte público fretado			
Idade: 18 a 20 anos	Referência		
De 21 a 23 anos			
Ônibus			0,160
Transporte público fretado			
Acima de 23 anos			
Modo ativo			
Ônibus	0,072	0,054	0,044
Transporte público fretado			0,036
Sexo: Feminino	Referência		
Masculino			
Modo ativo	2,355		10,201
Renda: Até 2 salários	Referência		
De 3 a 5 salários			
Modo ativo	0,212		
De 6 a 9 salários			
Modo ativo	0,212		
Ônibus	0,256		0,017
Transporte público fretado			0,096
Mais de 10 salários			
Modo ativo	0,147		
Ônibus	0,254		
Cidade: Joinville	Referência		
Outra			
Transporte público fretado	2,39		

Na Universidade A, o modelo escolhido indicou que estudantes que percorrem maiores distâncias (acima de 4 km) tem menos chance de utilizar o modo ativo em relação ao carro. O trabalho de Nguyen-Phuoc *et al.* (2018) apontou que os estudantes que viajam menos de 2 km estão mais propensos a utilizar modos de transportes ativos. Já a Universidade B apresentou resultados similares ao Polo Universitário.

4.4 COMPARAÇÃO DOS RESULTADOS DA ÁRVORE DE DECISÃO E DA REGRESSÃO MULTINOMIAL

Para comparação dos resultados encontrados para a técnica da Árvore de Decisão e modelo de Regressão Multinomial, foram utilizadas as seguintes medidas de desempenho obtidas através da matriz de confusão: acurácia e índice Kappa, conforme apresentadas na Tabela 21. Desta forma, pode-se observar que em ambas as técnicas os resultados foram próximos para os modelos de escolha do modo de transporte dos estudantes no Polo Universitário.

Com relação a Universidade A, houve uma pequena diferença entre as medidas de desempenho, com melhores resultados obtidos na amostra teste para a regressão multinomial. Este fato pode ser explicado devido a esta instituição de ensino representar aproximadamente apenas 35% da amostra total, e as técnicas de regressão conseguem alcançar melhores previsões com amostras menores, em comparação a técnicas de classificação, como a árvore de decisão. Já na Universidade B, os resultados da amostra de treinamento e teste para ambas as técnicas foram próximos, com comportamento similar para o Polo Universitário.

Tabela 21 - Medidas de desempenho das técnicas abordadas

Técnicas abordadas	Medidas de desempenho			
	Treinamento (70%)		Teste (30%)	
	Acurácia (%)	Índice Kappa	Acurácia (%)	Índice Kappa
Árvore de Decisão				
Polo Universitário	78,95	0,6850	73,75	0,5989
Universidade A	71,23	0,5627	64,29	0,4553
Universidade B	80,83	0,6997	71,43	0,5418
Regressão Multinomial				
Polo Universitário	73,96	0,6070	67,50	0,5080
Universidade A	72,60	0,5853	67,86	0,5135
Universidade B	79,51	0,6861	73,47	0,5801

A Tabela 22 mostra as variáveis significativas encontradas nos modelos de Árvore de Decisão e Regressão Multinomial. Comparando os resultados em cada técnica, observa-se que no Polo Universitário foram identificados como fatores que influenciam a escolha modal para ambos os modelos, o tempo, a idade e cidade. A variável distância também foi significativa para as duas técnicas, mas na Regressão Multinomial optou-se somente pela variável tempo, pois como já comentado anteriormente, os fatores tempo e distância são altamente correlacionados, sendo

indicado a escolha de apenas um deles para criação de modelos de regressão multinomial. Os fatores gênero e renda familiar foram encontrados como significativos apenas no modelo de regressão multinomial.

Para a Universidade A, em ambos os modelos foram identificados como relevantes para a escolha modal, a distância e a idade. Já para a Universidade B, ambos os modelos apresentaram as variáveis tempo, idade, renda familiar e cidade, como influentes na escolha modal. O fator distância também foi significativo em ambos os modelos, porém, assim como no Polo Universitário, optou-se por sua retirada no modelo de regressão multinomial. A variável gênero apareceu como relevante para a escolha do modo de transporte na Universidade B, apenas no modelo de regressão multinomial.

Tabela 22 - Variáveis significativas encontradas em cada técnica abordada na pesquisa

Objetos de estudo	Variáveis significativas encontradas na pesquisa	
	Árvore de Decisão	Regressão Multinomial
Polo Universitário	Tempo, distância, idade e cidade	Tempo, idade, gênero, renda e cidade
Universidade A	Distância e idade	Distância e idade
Universidade B	Tempo, distância, idade, renda e cidade	Tempo, idade, gênero, renda e cidade

Portanto, pode-se concluir que os fatores identificados como influentes na escolha do modo de transporte para os objetos de estudo foram similares, sendo que na Regressão Multinomial foi encontrada como diferente a variável gênero, que não apareceu em nenhum dos modelos da Árvore de Decisão.

4.5 COMPARAÇÃO DOS RESULTADOS DOS MÉTODOS UTILIZADOS NA PESQUISA COM A LITERATURA

A Tabela 23 apresenta um resumo dos fatores encontrados na pesquisa e na literatura como influentes na escolha do modo de transporte de estudantes universitários. Comparando os resultados, observa-se que as variáveis identificadas como significativas para a escolha modal nos modelos adotados utilizando as técnicas de Árvore de Decisão e Regressão Multinomial, também apareceram como relevantes em outros estudos envolvendo estudantes universitários.

Tabela 23 - Variáveis influentes na escolha modal encontradas na pesquisa e na literatura

Técnicas abordadas	Variáveis influentes na escolha modal	
	Pesquisa	Literatura
Árvore de Decisão		Pitombo e Gomes (2014), renda familiar
Polo Universitário	Tempo, distância, idade, cidade	Ferreira, Pitombo e Cunha (2016), idade
Universidade A	Distância, idade	Zhan <i>et al.</i> (2016), gênero
Universidade B	Tempo, distância, cidade, idade, renda	Nguyen-Phuoc <i>et al.</i> (2018) idade, gênero e renda
Regressão		Nasrin (2017), tempo
Multinomial		Kotoula <i>et al.</i> (2018) tempo e distância
Polo Universitário	Tempo, idade, gênero, renda, cidade	Romanowska, Okraszewska e Jamroz (2019) tempo, distância e gênero
Universidade A	Distância, idade	
Universidade B	Tempo, idade, gênero, renda, cidade	

Portanto, pode-se concluir que fatores socioeconômicos, como idade, gênero e renda familiar, podem ser influentes na escolha do modo de transporte de estudantes universitários. Também se observa que a escolha modal pode sofrer influência de outros fatores como tempo de deslocamento e distância percorrida. A variável cidade onde mora, fator não encontrado como relevante na literatura, apareceu neste estudo, pois como já mencionado anteriormente, a Universidade B (privada), apresenta como alternativa de modo de transporte para os estudantes oriundos de municípios vizinhos, o transporte público fretado, modo este, que está relacionado diretamente com esses estudantes.

Outros fatores também foram encontrados como influentes na escolha do modo de transporte na literatura, como apresentado no capítulo Revisão da Literatura no item 2.2.2, mas não estão na Tabela 23, pois o intuito desta foi identificar que as variáveis encontradas na presente pesquisa, também já foram constatadas como relevantes em outros estudos similares envolvendo a escolha do modo de transporte de estudantes universitários.

4.6 MODELOS OBTIDOS CONSIDERANDO APENAS AS CATEGORIAS TRANSPORTE PÚBLICO FRETADO, CARONA E MOTO

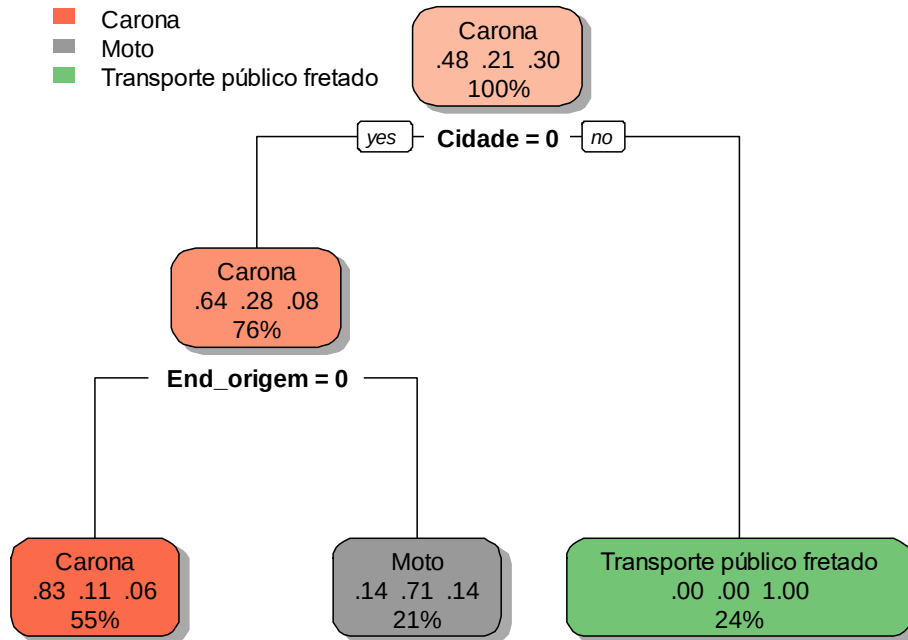
Nesta seção são apresentados os modelos obtidos para o Polo Universitário, considerando as categorias transporte particular, carona e moto, sendo que as duas

últimas categorias não aparecem nos modelos principais. Espera-se que esta seção possa despertar o interesse para o desenvolvimento de outras pesquisas para uma melhor compreensão dos fatores influentes na escolha de outros modos de transporte, além do carro, ônibus e modo ativo. A Tabela 24 apresenta a categorização das variáveis utilizadas para criação dos modelos de árvore de decisão e regressão multinomial.

Tabela 24 - Descrição das variáveis utilizadas nos modelos

Variáveis	Categorias
Dependente:	
Modo de transporte diário (Modo_transp)	Carona – referência Moto Transporte público fretado
Independentes:	
Endereço de origem (End_origem)	Casa – referência Trabalho
Cidade onde mora (Cidade)	Joinville – referência Outro

A Figura 20 representa a árvore de decisão para o Polo Universitário, sendo possível observar que as variáveis representativas do modelo foram a cidade onde mora e o endereço de origem. Assim, a primeira variável de divisão da árvore podada foi a cidade onde mora (0 significa reside em Joinville), com isso, o estudante que reside em Joinville tem probabilidade de 64% de utilizar como modo de transporte a carona, caso contrário tem 100% de probabilidade de utilizar o transporte público fretado. Para os estudantes que moram em Joinville aparece outra divisão, a variável endereço de origem (0 significa casa). O estudante que tem como endereço de partida sua residência tem 83% de probabilidade de fazer uso do modo carona, mas caso o estudante tenha como endereço de partida seu local de trabalho, tem a probabilidade de 71% de utilizar a moto como modo de transporte.

Figura 20 - Árvore de decisão para o Polo Universitário para outras categorias

Nas Tabelas 25 e 26 são apresentados, respectivamente, a matriz de confusão e os valores obtidos para avaliação do desempenho do modelo de árvore de decisão para o Polo Universitário. Na matriz de confusão observa-se que o modelo acertou bem os três modos de transporte. O modelo apresentou uma boa acurácia de 0,9091 (90,91%) e um índice Kappa de 0,8383, indicando que este modelo tem uma concordância quase perfeita. O resultado da sensibilidade indicou que o modelo acertou bem os verdadeiros positivos para o modo carona (1,0000) e transporte público fretado (1,0000), e moderadamente para modo moto (0,5000). Para a especificidade o modelo acertou bem os verdadeiros negativos para os modos moto (1,0000), transporte público fretado (1,0000), e carona (0,8000).

Tabela 25 - Matriz de confusão da amostra teste

Previsão / Real	Carona	Moto	Transporte público fretado
Carona	6	1	0
Moto	0	1	0
Transporte público fretado	0	0	3

Tabela 26 - Medidas de desempenho da amostra teste

Medidas de desempenho	Modo de transporte		
	Carona	Moto	Transporte público fretado
Sensibilidade	1,0000	0,5000	1,0000
Especificidade	0,8000	1,0000	1,0000
Acurácia		0,9091	
Índice Kappa		0,8382	

Utilizando a regressão multinomial, criou-se o Modelo 7 para o Polo Universitário, apresentado na Tabela 27, com as mesmas variáveis identificadas na árvore decisão, cidade onde mora e endereço de origem. A análise multivariada revelou que as variáveis categóricas do Modelo 7 são significativas no nível de $\alpha = 5\%$, “End_origem” ($\chi^2 = 9,95$; p -valor = 0,01) e “Cidade” ($\chi^2 = 34,83$; p -valor = 0,00). Com relação a matriz de confusão e desempenho do modelo, estes são apresentados na Tabela 28, sendo possível observar que os resultados são parecidos com os encontrados no modelo de árvore de decisão.

Na Tabela 27 analisando o modelos identifica-se que no nível de $\alpha = 5\%$, a variável “End_origem” é significativa para a moto, e o fator “Cidade” para a moto e transporte público fretado.

Tabela 27 - Modelo para o Polo Universitário

Variáveis		Modelo 7		
		Coeficientes	Erro padrão	p-valor
Interceptos				
	Carona	Referência		
	Moto	-1,609	0,632	0,011**
	Transporte público fretado	-11,617	86,000	0,893
End_origem: Casa		Referência		
	Trabalho			
	Moto	2,996	1,284	0,020**
	Transporte público fretado	11,617	86,012	0,893
Cidade: Joinville		Referência		
	Outra			
	Moto	-5,105	0,000	0,000***
	Transporte público fretado	36,749	0,000	0,000***
Significância		*** p-valor<0,01; ** p-valor<0,05; * p-valor<0,10		

Tabela 28 - Medidas de avaliação do modelo para Polo Universitário

Medidas de avaliação	Modelo 7		
	Carona	Moto	Transporte público fretado
Sensibilidade	1,0000	1,0000	0,6667
Especificidade	0,8000	1,0000	1,0000
Acurácia	0,9091		
Índice Kappa	0,8406		
Pseudo-R ²			
McFadden	0,7377		
Cox & Snell	0,7696		
Nagelkerke	0,8915		
AIC	38,631		

As razões de chances (*odds ratio*) para os coeficientes significativos dos modelos escolhidos são mostradas na Tabela 29. Caso a razão de chances apresente um valor menor que um, indica ser mais provável a escolha da categoria de referência, o modo carona. Analisando os resultados para o Polo Universitário, observa-se que os estudantes que tem como endereço de origem o trabalho tem maior chance de utilizar a moto como modo de transporte, em comparação ao modo carona. Também pode-se constatar que os alunos que não residem em Joinville, tem maiores chances de utilizarem o transporte público fretado como modo de transporte.

Tabela 29 - Razão de Chance para o modelo escolhido

Variáveis significativas	Modelo 7	
	Polo Universitário	
	Razão de Chance	
End_origem: Casa	Referência	
Trabalho	Moto	19,998
Cidade: Joinville	Referência	
Outra	Moto	
Transporte público fretado		9,115

5 CONCLUSÕES

Esta dissertação teve por objetivo identificar possíveis fatores relevantes para a escolha do modo de transporte de estudantes universitários em um Polo Universitário localizado no município de Joinville.

As instituições de ensino superior são consideradas como polos geradores de viagens, por este motivo, é importante a compreensão das variáveis que afetam na escolha do modo de transporte dos estudantes. O Polo Universitário estudado nesta pesquisa gera um grande deslocamento de pessoas nos horários de chegada e saída dos estudantes, ocasionando congestionamentos nas vias de acesso e entorno. Assim, entender quais fatores podem influenciar na escolha do modo de transporte dos estudantes universitários, pode auxiliar nas políticas de desenvolvimento dos sistemas viários e de transportes, além de contribuir para a busca de uma mobilidade mais sustentável e eficiente.

Com revisão da literatura, pode-se observar quais as variáveis mais frequentes que influenciam na escolha do modo de transporte, contribuindo para o desenvolvimento dos métodos de análise propostos neste estudo: a Árvore de Decisão e a Regressão Multinomial. A partir disso, foram criados vários modelos, sendo escolhido o melhor modelo para cada um dos métodos, com base em medidas de desempenho definidas na metodologia. Assim, foram selecionados três modelos de árvores de decisão (Polo Universitário, Universidade A e Universidade B) e três modelos de regressão multinomial (Polo Universitário, Universidade A e Universidade B), realizando ao final a comparação entre os modelos obtidas pelas duas técnicas e com a literatura, conforme proposto nos objetivos específicos da pesquisa.

Os modelos apresentaram boas medidas de desempenho para as duas técnicas, obtendo para o Polo Universitário valores de acurácias para a amostra teste de 73,75% e 67,50% para a árvore de decisão e regressão multinomial, respectivamente.

Os resultados apontaram que fatores como tempo de deslocamento, distância percorrida, idade, gênero, renda familiar e cidade onde mora, são variáveis significativas para a escolha do modo de transporte dos estudantes do Polo Universitário em estudo. Considerando cada método adotado, na técnica de Regressão Multinomial foi encontrada como influente na escolha modal a variável gênero, que não apareceu em nenhum dos modelos da Árvore de Decisão. As demais

variáveis apareceram como relevantes para a escolha modal nos dois métodos utilizados, em pelo menos um dos modelos.

Analisando separadamente as duas instituições de ensino, verificou-se que, na instituição pública, apenas a distância percorrida e a idade foram identificados como fatores que influenciam a escolha do modo de transporte dos estudantes desta universidade. Já para a instituição privada foram encontrados os mesmos fatores apontados como relevantes para a escolha modal nos modelos para o Polo Universitário, tempo de deslocamento, distância percorrida, idade, gênero, renda familiar e cidade onde mora. Este fato pode ser explicado devido a esta instituição de ensino representar aproximadamente 65% da amostra total, tendo assim, maior influência nos resultados do Polo Universitário.

A pesquisa também indicou que a distância percorrida é uma variável influente na escolha pelo modo ativo, e a cidade onde mora é influente na escolha pelo transporte público fretado. Já com relação ao carro e o ônibus, as principais variáveis influentes na escolha são tempo de deslocamento, idade e renda familiar. Com relação a idade, a pesquisa sugeriu que estudantes mais jovens possuem menor tendência ao uso do carro como modo de transporte. Relativo à renda familiar, o estudo indicou que conforme ela aumenta menores são as chances do uso do modo ativo e do ônibus, aumentando a escolha pelo uso do carro. Já a variável gênero indicou que os homens estão mais propensos a utilizarem o modo ativo como modo de deslocamento do que as mulheres.

Nesta pesquisa também pode ser observado que as variáveis identificadas como significativas para a escolha modal nos modelos adotados utilizando as técnicas de Árvore de Decisão e Regressão Multinomial, também apareceram como relevantes em outros estudos envolvendo estudantes universitários, conforme apresentado na revisão da literatura.

Portanto, a melhor compreensão dos fatores que influenciam a escolha modal pode fornecer subsídios para os órgãos públicos, auxiliando na tomada de decisões que envolvam o planejamento e investimento no sistema viário e de transportes para a área de estudo, e contribuindo para a busca de políticas que incentivem a escolha por modos de transporte mais sustentáveis.

Este trabalho também trouxe a aplicação de duas técnicas na análise da escolha modal, a Árvore de Decisão e a Regressão Logística Multinomial, contribuindo

para o estudo técnico científico da análise de demanda por transporte, principalmente no que diz respeito à modelagem de escolha modal.

Durante o desenvolvimento do trabalho foram encontradas algumas limitações no desenvolvimento dos modelos, como por exemplo com relação ao modo ativo, categoria esta, criada a partir da junção dos modos bicicleta e a pé, pois estes modos separadamente não apresentaram parâmetros significativos para os modelos, apenas em conjunto. Além disto, também foram retirados dos modelos principais os modos carona e moto, pois nas análises dos modelos iniciais os parâmetros para esses modos não foram significativos.

Como sugestão para trabalhos futuros, pode-se buscar a coleta de uma amostra com maior quantidade de respondentes, e verificar a influência na escolha de outros modos de transportes além dos identificados nesta pesquisa, como uso de aplicativos de transporte. Também pode-se criar um modelo de regressão espacial e compará-lo com os modelos obtidos neste trabalho.

REFERÊNCIAS

ABDALA, Ivanilde Maria de Rezende; PASQUALETTO, Antônio. Índice de Mobilidade Urbana Sustentável em Goiânia como ferramenta para políticas públicas. **Cadernos Metrópole**, São Paulo, v. 15, n. 30, p. 489-511, 2013.

ALVARENGA, Ana Maria Tavares. **Modelos lineares generalizados: aplicação a dados de acidentes rodoviários**. 101 f. Dissertação (Mestrado em Gestão de Informação) – Universidade de Lisboa. 2015. Disponível em: https://repositorio.ul.pt/bitstream/10451/22827/1/ulfc117306_tm_Ana_Maria_Alvarenga.pdf. Acesso em: 10 nov. 2019.

ALVARENGA JÚNIOR, Wagner José de. **Métodos de otimização hiperparamétrica: um estudo comparativo utilizando árvores de decisão e florestas aleatórias na classificação binária**. 2018. 82 f. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Minas Gerais, Belo Horizonte. 2018. Disponível em: <http://hdl.handle.net/1843/BUBD-AX2NLF>. Acesso em: 18 nov. 2019.

BASGALUPP, Márcio Porto. **LEGAL-Tree: um algoritmo genético multi-objeto lexicográfico para indução de árvores de decisão**. 2010. 116 f. Tese (Doutorado em Ciência de Computação e Matemática Computacional) – Universidade de São Paulo, São Carlos. 2010. Disponível em: <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-12052010-165344/publico/teseMarcioBasgalupp.pdf> >. Acesso em: 02 mar. 2019.

BATTISTON, Marcia; DEBATIN NETO, Arnaldo; OLEKSZECHEN, Nikolas. Barreira e facilitadores no uso da bicicleta em deslocamentos diários: alternativas para a mobilidade urbana. **Revista de Ciências HUMANAS**, Florianópolis, v. 51, n. 1, p. 269-286, jan. - jun. 2017.

BITTENCOURT, Hélio Radke. Regressão logística politômica: revisão teórica e aplicações. **Acta Scientiae**, Canoas, v. 5, n. 1, p. 77-86, 2013.

BRAMER, Max. **Principles of data mining**. Londres: Springer, 2007. DOI: 10.1007/978-1-84628-766-4.

BRASIL, Lei Federal nº 12.587, de 03 de janeiro de 2012. **Institui as diretrizes da Política Nacional de Mobilidade Urbana**. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12587.htm#art28>. Acesso em: 29 mar. 2018.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. **Classification and regression trees**. Wadsworth, 1984.

DETRAN/SC, Departamento Estadual de Trânsito do Estado de Santa Catarina. **Estatísticas: veículos em circulação em Santa Catarina**. 2019. Disponível em: <http://www.detransc.gov.br/estatisticas/veiculos>. Acesso em: 22 out. 2019.

DEVORE, Jay L. **Probabilidade e estatística: para engenharia e ciências**. São Paulo: Cengage Learning, 2012.

FAVERO, L. P.; BELFIORE, P.; SILVA, F. L.; CHAN, B. L. **Análise de dados: modelagem multivariada para tomada de decisões**. Rio de Janeiro: Campus/Elsevier, 2009.

FERNANDES, V. O.; PITOMBO, C. S.; ALIXANDRINI JUNIOR, M. J.; SANTOS, D. V. C. Aplicação conjunta de técnicas de análise multivariada de dados e análise espacial exploratória para avaliação do desempenho de transporte público. In: SIMPÓSIO BRASILEIRO DE CIÊNCIAS GEODÉSICAS E TECNOLOGIAS DA GEOINFORMAÇÃO, 4., 2012, Recife. **Anais eletrônicos....** Recife: UFPE, 2012. p. 1-9. Disponível em: <https://www3.ufpe.br/cgtg/SIMGEOIV/CD/artigos/SIG/023_5.pdf>. Acesso em: 06 mar. 2019.

FERREIRA, F. A.; PITOMBO, C. S.; CUNHA, A. L. Previsão de escolha modal em um campus universitário utilizando a regressão logística binomial. In: CONGRESSO LUSO BRASILEIRO PARA O PLANEJAMENTO URBANO, REGIONAL, INTEGRADO E SUTENTÁVEL, 7., 2016, Maceió. **Anais eletrônicos....** Maceió: Pluris, 2016. p. 725-736. Disponível em: <<http://www.fau.ufal.br/evento/pluris2016/files/Tema%203%20-%20Mobilidade%20e%20Transportes/Paper725.pdf>>. Acesso em: 18 nov. 2018.

FIGUEIRA, Cleonis Viater. **Modelos de regressão logística**. 2006. 149 f. Dissertação (Mestrado em Matemática) – Programa de Pós-graduação em Matemática do Instituto de Matemática da Universidade Federal do Rio Grande do Sul, Porto Alegre. 2006. Disponível em: <<http://hdl.handle.net/10183/8192>>. Acesso em: 15 maio 2018.

GARCIA, Simone Carboni. **O uso de árvores de decisão na descoberta de conhecimento na área da saúde**. 2003. 87 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre. 2003. Disponível em: <<https://www.lume.ufrgs.br/bitstream/handle/10183/4703/000503532.pdf?sequence=1>>. Acesso em: 10 mar. 2019.

GONÇALVES, Jorge Humberto Gaspar. **Fatores relevantes para a escolha modal em áreas urbanas**. 2012. 576 f. Tese (Doutorado em Engenharia Civil) – Universidade da Beira Interior, Covilhã. 2012. Disponível em: <https://ubibliorum.ubi.pt/bitstream/10400.6/3606/1/Tese_Fatores%20relevantes%20para%20a%20escolha%20modal%20em%20%C3%A1reas%20urbanas_JHGG_revisto.pdf>. Acesso em: 02 abr. 2018.

GOOGLE. **Google My Maps**. 2019. Disponível em: <<https://www.google.com/mymaps>>. Acesso em: 15 jun. 2019.

HADDAD, Eduardo A.; VIEIRA, Renato S. Mobilidade, acessibilidade e produtividade: nota sobre a valoração econômica do tempo de viagem na região metropolitana de São Paulo. **Revista de Economia Contemporânea**, Rio de Janeiro, v. 19, n. 3, p. 343-365, 2015.

HAIR JUNIOR, Joseph; BLACK, William C.; BABIN, Barry J.; ANDERSON, Rolph E.; TATHAM, Ronald L. **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009.

HARRIS, Sabrina. **Estrutural espacial urbana e mobilidade: o caso da região metropolitana de São Paulo**. 2015. 118 f. Dissertação (Mestrado em Arquitetura e Urbanismo) – Faculdade de Arquitetura e Urbanismo da Universidade de São Paulo, São Paulo. 2015. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/16/16132/tde-29102015-092157/publico/sabrinaharris.pdf>>. Acesso em: 14 maio 2018.

IBGE, Instituto Brasileiro de Geografia e Estatística. **Indicadores Sociais Municipais: uma análise dos resultados do universo do censo demográfico 2010**. Rio de Janeiro: IBGE, 2011. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv54598.pdf>>. Acesso em: 05 jun. 2018.

IBGE, Instituto Brasileiro de Geografia e Estatística. 2019. **Estatísticas por Cidade e Estado**. Disponível em: <<https://www.ibge.gov.br/estatisticas-novoportal/por-cidade-estado-estatisticas.html>>. Acesso em: 16 nov. 2019.

IEMA, Instituto de Energia e Meio Ambiente. **A bicicleta e as cidades: como inserir a bicicleta na política de mobilidade urbana**. 2010. Disponível em: https://iema-site-staging.s3.amazonaws.com/a_bicicleta_e_as_cidades.pdf. Acesso em: 22 out. 2019.

INEP, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Sinopse Estatística da Educação Superior 2018**. 2019. Disponível em: <http://portal.inep.gov.br/web/guest/sinopses-estatisticas-da-educacao-superior>. Acesso em: 15 dez. 2019.

IZENMAN, Alan Julian. **Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning**. Springer, 2008.

JOINVILLE, Decreto nº 24.181, de 27 de março de 2015. **Aprova o Plano de Mobilidade Sustentável de Joinville – PlanMOB**. Disponível em: <<https://leismunicipais.com.br/a1/sc/j/joinville/decreto/2015/2418/24181/decreto-n-24181-2015-aprova-o-plano-de-mobilidade-sustentavel-de-joinville-planmob>>. Acesso em: 01 abr. 2018.

JOINVILLE, Lei Complementar nº 470, de 09 de janeiro de 2017. **Redefine e institui, respectivamente, os Instrumentos de Controle Urbanístico - Estruturação e Ordenamento Territorial do Município de Joinville, partes integrantes do Plano Diretor de Desenvolvimento Sustentável do Município de Joinville e dá outras providências**. Disponível em: <<https://leismunicipais.com.br/a1/plano-de-zoneamento-uso-e-ocupacao-do-solo-joinville-sc>>. Acesso em: 30 mar. 2018.

KOTOULA, Kornilia Maria; SIALDAS, Alexandros; BOTZORIS, George; CHANIOTAKIS, Emmanouil; GRAU, Josep Maria Salanova. Exploring the Effects of University Campus Decentralization to Students' Mode Choice. **Periodica Polytechnica Transportation Engineering**, Budapeste, v. 46, n. 4, p. 207-2014, 2018.

KUHN, Max. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris

Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan and Tyler Hunt. (2019). **caret: Classification and Regression Training**. R package version 6.0-84. Disponível em: <https://CRAN.R-project.org/package=caret>.

LANDIS, J. R.; KOCH, G. G. The Measurement of Observer Agreement for Categorical Data. **Biometrics**, vol. 33, n. 1. p.159-174, 1977.

LARGURA, Aline Estela. **Fatores que influenciam o uso de bicicleta em cidades de médio porte: estudo de caso em Balneário Camboriú/SC**. 2012. 111 f. Dissertação (Mestrado) - Curso de Arquitetura e Urbanismo, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2012. Disponível em: . Acesso em: 22 dez. 2015.

LAURETTO, Marcelo S. **Árvores de Decisão**. 27f. Notas de Aula – Universidade de São Paulo. 2010. Disponível em: https://edisciplinas.usp.br/pluginfile.php/4469825/mod_resource/content/1/ArvoresDecisao_normalsize.pdf. Acesso em: 20 out. 2019.

LIBARDI, Rafaela. Mobilidade urbana frente à complexidade urbana. **EURE**, Santiago, v. 40, n. 121, p. 273-276, 2014.

LINDNER, Anabele. **Análise desagregada de dados de demanda por transporte através de modelagem geoestatística e tradicional**. 2015. 106 f. Dissertação (Mestrado em Ciências) – Departamento de Engenharia de Transportes, Escola de Engenharia de São Carlos da Universidade de São Paulo, São Carlos. 2015. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/18/18144/tde-09042015-090539/publico/ANABELELINDNER.pdf>>. Acesso em: 01 abr. 2018.

LINDNER, Anabele; PITOMBO, Cira Souza. Modelo logit binomial com componentes principais para estimação de preferência por modo de transporte motorizado. **The Journal of Transport Literature**, Manaus, v. 10, n. 3, p. 5-9, 2016.

LOPES, Simone Becker. **Uma ferramenta para planejamento da mobilidade sustentável com base em modelo de uso do solo e transportes**. 2010. 200 f. Tese (Doutorado em Ciências) – Programa de Pós-Graduação em Engenharia de Transportes, Escola de Engenharia de São Carlos da Universidade de São Paulo, São Carlos. 2010. Disponível em: <http://www.teses.usp.br/teses/disponiveis/18/18144/tde-13122010_161312/publico/TeseSBL_291210.pdf>. Acesso em: 22 abr. 2018.

MACHADO, Laura; PICCININI, Lívia Salomão. Os desafios para a efetividade da implementação dos planos de mobilidade urbana: uma revisão sistemática. **Urbe, Revista Brasileira de Gestão Urbana**, Curitiba, v. 10, n. 1, p. 72-94, 2018.

MAIA, C. A.; RODRIGUES, L. A.; MOREIRA, M. E. P.; MONTENEGRO, N. F. S. D. Análise das características relevantes no deslocamento dos usuários do campus do Pici – UFC. Congresso Nacional de Ensino e Pesquisa em Transportes, 26. **Anais....** Joinville: ANTP, 2012.

MESQUITA, Paulo Sérgio Belchior. **Um modelo de regressão logística para avaliação dos programas de pós-graduação no Brasil**. 2014. 90 f. Dissertação (Mestrado em Engenharia de Produção) – Centro de Ciência e Tecnologia da Universidade Estadual do Norte Fluminense, Campo dos Goytacazes. 2014. Disponível em: <<http://uenf.br/posgraduacao/engenharia-de-producao/wp-content/uploads/sites/13/2013/04/Disserta%C3%A7%C3%A3o-Paulo-Mesquita.pdf>>. Acesso em: 15 maio 2018.

MITCHELL, Tom M. **Machine Learning**. Nova Iorque: McGraw-Hill, 1997.

MONTGOMERY, Douglas C.; RUNGER, George C. **Estatística aplicada e probabilidade para engenheiros**. Tradução e revisão técnica Verônica Calado. 5. ed. Rio de Janeiro: LTC, 2012.

MOTTA, Porthos Ribeiro de Albuquerque. **Estudo exploratório do uso de classificadores para a predição de desempenho e abandono em universidades**. 2016. 156 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Goiás, Goiânia. 2016. Disponível em: <http://repositorio.bc.ufg.br/tede/handle/tede/6563>. Acesso em: 5 jan. 2020.

NASRIN, Sharmin. Private University Students' Mode Choice Behaviour for Travel to University: Analysis in the Context of Dhaka City. In: CONFERÊNCIA SOBRE O GRUPO DE PESQUISA EM TRANSPORTE DA ÍNDIA, 4. 2017, Mumbai. **Anais eletrônicos....** Mumbai: IIT, 2017. Disponível em: <https://www.researchgate.net/publication/321137364>. Acesso em: 20 nov. 2019.

NGUYEN-PHUOC, Duy Quy; AMOAH-GYIMAH, Richard; TRAN, Anh T. P.; PHAN, Cao Tho. Mode choice among university students to school in Danang, Vietnam. **Travel Behaviour and Society**, Hong Kong, v. 13, p. 1-10, 2018.

NSC TOTAL. **Cotidiano: Gargalhos do trânsito**. 2019. Disponível em: <<https://www.nsctotal.com.br/noticias/tres-ruas-concentram-os-pontos-de-congestionamento-na-zona-norte-de-joinville>>. Acesso em: 27 abr. 2019.

NUTSUGBODO, Ricky Yao; AMENUMEY, Edem Kwesi; MENSAH, Collins Adjei. Public transport mode preferences of international tourists in Ghana: Implications for transport planning. **Travel Behaviour and Society**, Hong Kong, v. 11, p. 1-8, 2018.

OLIVEIRA, Angélica Meireles de. **Um índice para o planejamento de mobilidade com foco em Grandes Polos Geradores de Viagens – Desenvolvimento e aplicação em um campus universitário**. 2015. 197f. Dissertação (Mestrado em Ciências) – Programa de Pós-Graduação em Engenharia de Transportes, Escola de Engenharia de São Carlos da Universidade de São Paulo, São Carlos. 2015. Disponível em: < <http://www.teses.usp.br/teses/disponiveis/18/18144/tde-23022016-143827/publico/DissertOliveiraAngelicaMeirelesdeCorrig.pdf> >. Acesso em: 22 fev. 2019.

ORAL, L. O.; Tecim, V. Using decision tress for estimating mode choice of trips in Buca – Izmir. In: Simpósio de Gerenciamento de Dados Urbanos, 29., 2013, Londres. **Anais**

eletrônicos.... Londres: ISPRS, 2013. p. 139-145. DOI: 10.5194/isprsarchives-XL-4-W1-139-2013.

ORTÚZAR, Juan de Dios; WILLUMSEN, Luis G. **Modelling Transport**. 4 ed. Londres: Wiley, 2011.

PARRA, Marsela Caipa. **Gerenciamento da mobilidade em campi universitários: problemas, dificuldades e possíveis soluções no caso Ilha do Fundão – UFRJ**. 2006. 109 f. Dissertação (Mestrado em Engenharia de Transportes) – Programa de Pós-Graduação de Engenharia da Universidade Federal do Rio de Janeiro, Rio de Janeiro. 2006. Disponível em: <http://objdig.ufrj.br/60/teses/coppe_m/MarselaCaipaParra.pdf>. Acesso em: 16 maio 2018.

PASSOS, Luciana Andrade dos; SILVEIRA, F. A.; PITA, A. L. L. R.; BRAGA, C. F. C.; SILVEIRA, J. A. R. Processo de expansão *versus* sustentabilidade urbana: reflexão sobre as alternativas de deslocamento na cidade de João Pessoa, PB. **Urbe, Revista Brasileira de Gestão Urbana**, Curitiba, v. 4, n. 1, p. 47-59, 2012.

PERO, Valéria; STEFANELLI, Victor. A questão da mobilidade urbana nas metrópoles brasileiras. **Revista de Economia Contemporânea**, Rio de Janeiro, v. 19, n. 3, p. 366-402, 2015.

PIANUCCI, M. N.; PITOMBO, C. S. Uso de árvore de decisão para previsão de geração de viagens como alternativa ao método de classificação cruzada. **Revista de Engenharia Civil**, Braga, n. 56, p. 5-13, 2019. Disponível em: <http://www.civil.uminho.pt/revista/artigos/n56/Pag.5-13.pdf>. Acesso em: 06 dez 2019.

PITOMBO, Cira Souza; COSTA, Aline Schindler Gomes da. Aplicação conjunta de modelos não paramétricos e paramétricos para previsão de escolha modal. **The Journal of Transport Literature**, Manaus, v. 9, n. 1, p. 30-34, 2015.

PITOMBO, Cira Souza; COSTA, Aline Schindler Gomes da; SALGUEIRO, Ana Rita. Proposal of a sequential method for spatial interpolation of mode choice. **Boletim de Ciências Geodésicas**, Curitiba, v. 21, n. 2, p. 274-289, 2015.

PITOMBO, Cira Souza; GOMES, Monique Martins. Study of work-travel related behavior using principal component analysis. **Open Journal of Statistics**, Wuhan, v. 4, n. 11, p. 889-901, 2014.

PITOMBO, Cira Souza; SOUZA, Andreza Dornelas de; LINDNER, Anabele. Comparing decision tree algorithms to estimate intercity trip distribution. **Transportation Research Part C**, v. 77, p. 16-32, 2017.

QUINLAN, J. R. Induction of decision trees. **Machine Learning**, v. 1, n. 1, p. 81-106, mar. 1986.

QUINLAN, J. (1988). Decision trees and multivalued attributes. **Machine Intelligence**, v. 11, p. 305-318, set. 1988.

QUINLAN, J. R. **C4.5: programs for machine learning**. San Francisco: Morgan Kaufmann Publishers Inc., 1993.

R Core Team (2019). R: **A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <https://www.R-project.org/>.

RAIMUNDO, Lidiane Rosso; MATTOS; Merisandra Côrtes de; SIMÕES, Priscyla Waleska Targino de Azevedo; CECHINEL, Cristian. O Algoritmo de Classificação CART em uma Ferramenta de Data Mining. In: CONGRESSO SUL BRASILEIRO DE COMPUTAÇÃO, 4. 2008, Criciúma. **Anais eletrônicos....** Criciúma: SULCOMP, 2008. Disponível em: <http://periodicos.unesc.net/sulcomp/article/view/1994/1888>. Acesso em: 05 dez. 2019.

RAMIS, Jacqueline Elhage; SANTOS, Emmanuel Antônio dos. Uso de automóveis e o caos urbano: considerações sobre o planejamento de transportes das grandes cidades. **The Journal of Transport Literature**, Manaus, v. 6, n. 4, p. 164-177, 2012.

RENCER, Alvin C. **Methods of multivariate analysis**. 2 ed. Londres: Wiley-Interscience: Nova Jersey, 2012.

RESENDE, Ubiratan Pereira de. **As contradições e os desafios da mobilidade urbana de Goiânia no contexto de sua região metropolitana**. 2017. 328 f. Tese (Doutorado em Geografia) – Programa de Pesquisa e Pós- Graduação em Geografia, do Instituto de Estudos Sócio-Ambientais, da Universidade Federal de Goiás, Goiânia. 2017. Disponível em: <<https://repositorio.bc.ufg.br/tede/bitstream/tede/7116/5/Tese%20-%20Ubiratan%20Pereira%20de%20Resende%20-%202017.pdf>>. Acesso em: 15 maio 2018.

ROMA, Andreza Dornelas de Souza; PITOMBO, Cira Souza; GUIMARÃES, Henrique Stramandinoli; COSTA, Luiz Henrique Magalhães. Análise de desempenho de algoritmos de aprendizagem de máquinas para análise desagregada de viagens intermunicipais. **Transportes**, ANPET, v. 26, n. 3, p. 159-175, nov. 2018.

ROMANOWSKA, Aleksandra; OKRASZEWSKA, Romanika; JAMROZ, Kazimierz. A Study of Transport Behaviour of Academic Communities. **Sustainability**, v. 11, n. 13, p. 1-18, 2019.

RUBIM, Barbara; LEITÃO, Sérgio. O plano de mobilidade urbana e o futuro das cidades. **Estudos Avançados**, São Paulo, v. 27, n. 79, p. 55-66, 2013.

SAFITRI, Dian Mardi; Surjandari, Isti. Travel Mode Switching Prediction using Decision Tree in Jakarta Greater Area. In: CONFERÊNCIA INTERNACIONAL SOBRE SISTEMAS DE TECNOLOGIA DA INFORMAÇÃO E INOVAÇÃO. 2017, Badung. **Anais eletrônicos....** Badung: 2017.

SANCHES, S. P.; FERREIRA, M. A. G. Mobilidade em um Campus Universitário. CONGRESSO BRASILEIRO DE TRANSPORTES, 19. **Anais eletrônicos....** Brasília: ANTP, 2013. Disponível em: <<http://redpgv.coppe.ufrj.br/index.php/pt-BR/producao>>.

da-rede/universidades-com-pgvs/959-mobilidade-em-um-campus-universitario/file>. Acesso em: 25 nov. 2018.

SCHUBERT, Thamires Ferreira. **Análise da possibilidade de mudança da escolha do modo de transporte: estudo de caso dos deslocamentos de estudantes universitários**. 124 f. Dissertação (Mestrado em Engenharia Civil) – Universidade do Estado de Santa Catarina, Joinville. 2019. Disponível em: https://www.udesc.br/arquivos/cct/id_cpmenu/706/2019_05_15_Thamires_Ferreira_Schubert_15689962257242_706.pdf. Acesso em: 15 set. 2019.

SEPUD - SECRETARIA DE PLANEJAMENTO URBANO E DESENVOLVIMENTO SUSTENTÁVEL. 2018. **Joinville Cidade em Dados 2018**. Disponível em: <<https://www.joinville.sc.gov.br/wp-content/uploads/2018/09/Joinville-Cidade-em-Dados-2018-Character%C3%ADsticas-Gerais.pdf>>. Acesso em: 16 nov. 2018.

SILVA, Fernando Nunes da. Mobilidade urbana: os desafios do futuro. **Cadernos Metr pole**, S o Paulo, v. 15, n. 30, p. 377-388, 2013.

SILVA, Ricardo Jorge Perdig o da. **Mobilidade Urbana: A bicicleta como meio de transporte di rio**. 2012. 124 f. Disserta o (Mestrado) - Curso de Design Industrial, Departamento em Engenharia Mec nica e Gest o Industrial, Universidade do Porto, Porto, 2012.

SILVA, Tha s. **An lise da escolha modal binomial com base no modelo logit**. 2010. 127 f. Disserta o (Mestrado em Engenharia Civil) – Universidade Federal de Uber ndia, Uber ndia. 2010. Disponível em: <<https://repositorio.ufu.br/bitstream/123456789/14142/1/Diss%20Thais.pdf>>. Acesso em: 30 mar. 2018.

SILVEIRA, Igor Machado da. **Avalia o da mobilidade urbana com  nfase na condi o da acidentalidade do usu rio vulner vel**. 2016. 93 f. Disserta o (Mestrado em Engenharia Civil) – Escola de Engenharia de S o Carlos da Universidade de S o Paulo, S o Carlos. 2016. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/18/18144/tde-16062016-171651/publico/DissertacaoSilveiraIM.pdf>>. Acesso em: 03 abr. 2018.

SIMGEO, Sistema de Informa  es Municipais Georreferenciadas PMJ. **Mapa base**. 2019. Disponível em: <https://simgeo.joinville.sc.gov.br/>. Acesso em: 14 set. 2019.

SOUZA, Marcello Victorino Junqueira de. **O papel da modelagem da divis o modal na elabora o de planos de mobilidade urbana sustent vel**. 2016. 150 f. Disserta o (Mestrado em Engenharia de Transportes) – Instituto Alberto Luiz Coimbra de P s-gradua  o e Pesquisa de Engenharia da Universidade Federal do Rio de Janeiro, Rio de Janeiro. 2016. Disponível em: <http://www.pet.coppe.ufrj.br/index.php/producao/teses-de-dsc/doc_download/330--o-papel-da-modelagem-da-divisao-modal-na-elaboracao-de-planos-de-mobilidade-urbana-sustentavel>. Acesso em 03 abr. 2018.

SPECHT, L. P.; BRANDLI, L. L.; SILVA, R. Modelagem da preferência do usuário para a escolha do transporte público universitário. **Revista Produção Online**, Florianópolis, v. 9, n. 2, p. 303-318, 2009.

STEIN, Peolla Paula. **Barreiras, motivações e estratégias para mobilidade sustentável no campus São Carlos USP**. 2013. 269 f. Dissertação (Mestrado em Ciências) – Programa de Pós-Graduação em Engenharia de Transportes, Escola de Engenharia de São Carlos da Universidade de São Paulo, São Carlos. 2013. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/18/18144/tde-27062013-163702/publico/PeolaPaulaStein.pdf>>. Acesso em: 11 maio 2018.

TEIXEIRA, Sivanilza Machado; REIS, João Gilberto Mendes dos; SANTOS, Rodrigo Couto; OLIVEIRA, Rone Vieira; VERGARA, Walter Hernandez; JORDAN, Rodrigo Aparecido. Qualidade do transporte urbano de passageiros: uma avaliação do nível de serviço do sistema do metropolitano de São Paulo. **Revista Metropolitana de Sustentabilidade**, São Paulo, v. 4, n. 1, p. 3-20, 2014.

THERNEAU, Terry; ATKINSON, Beth Atkinson. **rpart: Recursive Partitioning and Regression Trees**. 2018. R package version 4.1-13. Disponível em: <https://CRAN.R-project.org/package=rpart>.

UDESC, Universidade do Estado de Santa Catarina. **UDESC em números**. 2019. Disponível em: <https://www.udesc.br/numeros>. Acesso em: 10 dez. 2019.

VENABLES, W. N. & Ripley, B. D. (2002) **Modern Applied Statistics with S**. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.

WHALEN, Kate E.; PÁEZ, Antonio; CARRASCO, Juan A. Mode choice of university students commuting to school and the role of active travel. **Journal of Transport Geography**, Elsevier, v. 31, p. 132-142, 2013.

ZHAN, G.; YAN, X.; ZHU, S.; WANG, Y. Using hierarchical tree-based regression model to examine University student travel frequency and mode choice patterns in China. **Transport Policy**, v. 45, pp. 55–65, jan. 2016.

ZHANG, Rui; YAO, Enjian; LIU, Zhili. School travel mode choice in Beijing, China. **Journal of Transport Geography**, v. 62, p. 98-110, jun. 2017.

APÊNDICE A - Testes t e Wilcoxon realizados entre os modos de transporte

A seguir são apresentados os resultados do teste t e Wilcoxon entre as médias de idade dos estudantes que utilizam os modos de transporte contemplados na presente pesquisa. Caso o *p*-valor encontrado para os testes seja menor que 0,05, considerando $\alpha = 5\%$, indica que as médias são diferentes entre si, caso contrário, as médias não diferem entre si.

1 – Teste entre as médias de idade dos estudantes que utilizam o carro e o transporte público fretado:

```
> t.test(media_carro$Idade,media_fretado$Idade)

welch Two Sample t-test

data: media_carro$Idade and media_fretado$Idade
t = 4.6728, df = 43.385, p-value = 2.888e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.782999 7.007211
sample estimates:
mean of x mean of y
 25.81818  20.92308

> wilcox.test(media_carro$Idade,media_fretado$Idade)

wilcoxon rank sum test with continuity correction

data: media_carro$Idade and media_fretado$Idade
W = 753, p-value = 0.003702
alternative hypothesis: true location shift is not equal to 0
```

2 – Teste entre as médias de idade dos estudantes que utilizam o carro e o modo ativo:

```
> t.test(media_carro$Idade,media_modoativo$Idade)

welch Two Sample t-test

data: media_carro$Idade and media_modoativo$Idade
t = 3.5664, df = 132.18, p-value = 0.0005047
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.470802 5.134311
sample estimates:
mean of x mean of y
 25.81818  22.51562

> wilcox.test(media_carro$Idade,media_modoativo$Idade)

wilcoxon rank sum test with continuity correction

data: media_carro$Idade and media_modoativo$Idade
W = 3309.5, p-value = 0.0004411
```

```
alternative hypothesis: true location shift is not equal to 0
```

3 – Teste entre as médias de idade dos estudantes que utilizam o carro e o ônibus:

```
> t.test(media_carro$Idade,media_onibus$Idade)

      welch Two Sample t-test

data:  media_carro$Idade and media_onibus$Idade
t = 5.6368, df = 104.32, p-value = 1.486e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.991361 6.238223
sample estimates:
mean of x mean of y
 25.81818  21.20339

> wilcox.test(media_carro$Idade,media_onibus$Idade)

      wilcoxon rank sum test with continuity correction

data:  media_carro$Idade and media_onibus$Idade
W = 7007, p-value = 1.213e-10
alternative hypothesis: true location shift is not equal to 0
```

4 – Teste entre as médias de idade dos estudantes que utilizam o carro e o modo outro:

```
> t.test(media_carro$Idade,media_outro$Idade)

      welch Two Sample t-test

data:  media_carro$Idade and media_outro$Idade
t = 4.316, df = 89.209, p-value = 4.101e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.425984 6.565218
sample estimates:
mean of x mean of y
 25.81818  21.32258

> wilcox.test(media_carro$Idade,media_outro$Idade)

      wilcoxon rank sum test with continuity correction

data:  media_carro$Idade and media_outro$Idade
W = 1738, p-value = 0.000209
alternative hypothesis: true location shift is not equal to 0
```

5 – Teste entre as médias de idade dos estudantes que utilizam o ônibus e o transporte público fretado:

```
> t.test(media_fretado$Idade,media_onibus$Idade)

      welch Two Sample t-test

data:  media_fretado$Idade and media_onibus$Idade
```

```
t = -0.35151, df = 17.108, p-value = 0.7295
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.961997  1.401371
sample estimates:
mean of x mean of y
 20.92308  21.20339
```

```
> wilcox.test(media_fretado$Idade,media_onibus$Idade)
```

```
Wilcoxon rank sum test with continuity correction
```

```
data: media_fretado$Idade and media_onibus$Idade
W = 778, p-value = 0.9348
alternative hypothesis: true location shift is not equal to 0
```

6 – Teste entre as médias de idade dos estudantes que utilizam o modo outro e o transporte público fretado:

```
> t.test(media_fretado$Idade,media_outro$Idade)
```

```
Welch Two Sample t-test
```

```
data: media_fretado$Idade and media_outro$Idade
t = -0.38975, df = 33.936, p-value = 0.6992
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.482743  1.683736
sample estimates:
mean of x mean of y
 20.92308  21.32258
```

```
> wilcox.test(media_fretado$Idade,media_outro$Idade)
```

```
Wilcoxon rank sum test with continuity correction
```

```
data: media_fretado$Idade and media_outro$Idade
W = 189.5, p-value = 0.7651
alternative hypothesis: true location shift is not equal to 0
```

7 – Teste entre as médias de idade dos estudantes que utilizam o modo ativo e o transporte público fretado:

```
> t.test(media_fretado$Idade,media_modoativo$Idade)
```

```
Welch Two Sample t-test
```

```
data: media_fretado$Idade and media_modoativo$Idade
t = -1.7552, df = 27.22, p-value = 0.09047
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.4535000  0.2684038
sample estimates:
mean of x mean of y
 20.92308  22.51562
```

```
> wilcox.test(media_fretado$Idade,media_modoativo$Idade)
```

```
Wilcoxon rank sum test with continuity correction
```

```
data: media_fretado$Idade and media_modoativo$Idade
w = 338, p-value = 0.2884
alternative hypothesis: true location shift is not equal to 0
```

8 – Teste entre as médias de idade dos estudantes que utilizam o ônibus e o modo ativo:

```
> t.test(media_onibus$Idade,media_modoativo$Idade)

welch Two Sample t-test

data: media_onibus$Idade and media_modoativo$Idade
t = -2.0852, df = 108.67, p-value = 0.394
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.55957375 -0.06489658
sample estimates:
mean of x mean of y
 21.20339  22.51562

> wilcox.test(media_onibus$Idade,media_modoativo$Idade)

wilcoxon rank sum test with continuity correction

data: media_onibus$Idade and media_modoativo$Idade
w = 1286.5, p-value = 0.6495
alternative hypothesis: true location shift is not equal to 0
```

9 – Teste entre as médias de idade dos estudantes que utilizam o ônibus e o modo outro:

```
> t.test(media_onibus$Idade,media_outro$Idade)

welch Two Sample t-test

data: media_onibus$Idade and media_outro$Idade
t = -0.15095, df = 42.834, p-value = 0.8807
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.711785  1.473403
sample estimates:
mean of x mean of y
 21.20339  21.32258

> wilcox.test(media_onibus$Idade,media_outro$Idade)

wilcoxon rank sum test with continuity correction

data: media_onibus$Idade and media_outro$Idade
w = 1738.5, p-value = 0.6702
alternative hypothesis: true location shift is not equal to 0
```

10 – Teste entre as médias de idade dos estudantes que utilizam o modo ativo e o modo outro:

```
> t.test(media_outro$Idade,media_modoativo$Idade)

welch Two Sample t-test
```

```

data: media_outro$Idade and media_modoativo$Idade
t = -1.325, df = 63.618, p-value = 0.1899
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.9920688  0.6059801
sample estimates:
mean of x mean of y
 21.32258  22.51562

> wilcox.test(media_outro$Idade,media_modoativo$Idade)

      wilcoxon rank sum test with continuity correction

```

```

data: media_outro$Idade and media_modoativo$Idade
W = 837.5, p-value = 0.2186
alternative hypothesis: true location shift is not equal to 0

```

APÊNDICE B – Árvore de Decisão no R

##Dados Polo Universitário

```
library(readxl)
library(caret)
library(rpart)
dados_tempo_distancia_categorizado_novo_polo_sem_outro <-
read_excel("C:/Users/Tiago/Dropbox/MEP/dados-tempo-distancia-categorizado-
novo-polo-sem-outro.xlsx")
View(dados_tempo_distancia_categorizado_novo_polo_sem_outro)
attach(dados_tempo_distancia_categorizado_novo_polo_sem_outro)
```

##Árvore de Decisão Polo Universitário

```
set.seed(1)
Treinoindice<-createDataPartition(Modo_transp, p=0.7, list = FALSE)
polo_treino<-
dados_tempo_distancia_categorizado_novo_polo_sem_outro[Treinoindice,]
polo_teste<-dados_tempo_distancia_categorizado_novo_polo_sem_outro[-
Treinoindice,]
arvore_treino<-
rpart(Modo_transp~Inst_ensino+Distancia_ida+Tempo_ida+Periodo_estudo+Idade+
Sexo+Renda_fam+Est_civil+Cidade+End_origem+Filho+Trabalho, data =
polo_treino)
rpart.plot::rpart.plot(arvore_treino, shadow.col = "darkgray", cex = 0.55)
```

##Matriz de confusão

```
probab<-predict(arvore_treino,polo_treino,type='class')
confusionMatrix(table(probab, polo_treino$Modo_transp))
```

##Poda Árvore de Decisão Polo Universitário

```
printcp(arvore_treino)
rpart.plot::rpart.plot(prune(arvore_treino,cp=0.074766),cex = 0.55)
```

##Teste Árvore de Decisão Polo Universitário

```
arvore_teste<-
rpart(Modo_transp~Inst_ensino+Distancia_ida+Tempo_ida+Periodo_estudo+Idade+
Sexo+Renda_fam+Est_civil+Cidade+End_origem+Filho+Trabalho, data =
polo_teste)
rpart.plot::rpart.plot(arvore_teste, shadow.col = "darkgray")
```

##Matriz de confusão

```
teste_predicao<-predict(arvore_teste, newdata = polo_teste, type = 'class')
confusionMatrix(table(teste_predicao, polo_teste$Modo_transp))
```

##Poda Teste Árvore de Decisão Polo Universitário

```
printcp(arvore_teste)
rpart.plot::rpart.plot(prune(arvore_teste,cp=0.022222))
```

APÊNDICE C – Regressão Multinomial no R

Polo Universitário

```
library(readxl)
library(caret)
library(nnet)
library(foreign)
library(ggplot2)
library(reshape2)
dados_tempo_distancia_sem_outro_novo <- read_excel("C:/Users/Tiago/Dropbox/MEP/dados-tempo-distancia-sem outro-novo.xlsx")
View(dados_tempo_distancia_sem_outro_novo)
attach(dados_tempo_distancia_sem_outro_novo)
set.seed(5)
Treinoindice<-createDataPartition(Modo_transp, p=0.7, list = FALSE)
polo_treino<-dados_tempo_distancia_sem_outro_novo[Treinoindice,]
polo_teste<-dados_tempo_distancia_sem_outro_novo[-Treinoindice,]
Idade <- factor(Idade, levels=sort(unique(Idade)))
Tempo_ida <- factor(Tempo_ida, levels=sort(unique(Tempo_ida)))
Distancia_ida <- factor(Distancia_ida, levels=sort(unique(Distancia_ida)))
Renda_fam <- factor(Renda_fam, levels=sort(unique(Renda_fam)))
Sexo <- factor(Sexo, levels=sort(unique(Sexo)))
Cidade <- factor(Cidade, levels=sort(unique(Cidade)))
Modo_transp <- factor(Modo_transp, levels=sort(unique(Modo_transp)))
relevel(Modo_transp, ref = "Carro (motorista)")
relevel(Idade, ref = "A")
relevel(Tempo_ida, ref = "A")
relevel(Renda_fam, ref = "Até dois salários mínimos")
relevel(Sexo, ref = "Feminino")
relevel(Cidade, ref = "Em Joinville")
```

##Modelo regressão multinomial Polo Universitário

```
modelo1<-multinom(Modo_transp~Tempo_ida+Idade+Sexo+Renda_fam+Cidade, data = polo_treino)
summary(modelo1)
library(stargazer)
stargazer::stargazer(modelo1, type = "text")
modelo1.rrr=exp(coef(modelo1))
modelo1.rrr
stargazer(modelo1, type = "text", coef = list(modelo1.rrr), p.auto = FALSE)
```

##Matriz de Confusão modelo treinamento

```
probab<-predict(modelo1,polo_treino,type='class')
confusionMatrix(table(probab, polo_treino$Modo_transp))
```

##Matriz de Confusão modelo teste

```
teste_predicao<-predict(modelo1, newdata = polo_teste, type = 'class')
confusionMatrix(table(teste_predicao, polo_teste$Modo_transp))
```


##Teste Chi-quadrado variáveis do modelo

```
library(car)
Anova(modelo1, type = "II", test="Chisq")
modelot<-multinom(Modo_transp~Distancia_ida+Idade+Sexo+Renda_fam+Cidade,
data = polo_teste)
```

##Pseudo-r² modelo

```
library(DescTools)
PseudoR2(modelo1, c("McFadden", "Nagel", "CoxSnell", "McFaddenAdj"))
PseudoR2(modelot, c("McFadden", "Nagel", "CoxSnell", "McFaddenAdj"))
```

APÊNDICE D – Regressão Multinomial Modelo 1

Dependent variable:			
	Modo ativo	Transporte público fretado	Transporte público (ônibus)
	(1)	(2)	(3)
Distancia_idaB	-5.733*** (1.980)	-94.609	-2.412 (1.972)
Distancia_idaC	-6.698*** (2.021)	-220.240	-1.874 (1.969)
Distancia_idaD	-6.271*** (2.037)	64.393 (123.384)	-1.009 (1.962)
Distancia_idaE	-6.278*** (2.100)	-26.677 (370.147)	-1.171 (1.977)
IdadeB	0.208 (0.703)	-155.063	-0.482 (0.506)
IdadeC	-0.266 (0.706)	-341.404 (493.530)	-1.945*** (0.527)
SexoMasculino	1.470** (0.590)	2.265 (1.660)	-0.043 (0.451)
Renda_famDe seis a nove	-1.642* (0.893)	178.494*** (0.00000)	-1.341** (0.679)
Renda_famDe três a cinco	-1.007 (0.724)	325.808 (493.530)	-0.915 (0.567)
Renda_famMais de dez	-2.242** (1.025)	377.527*** (123.384)	-2.290*** (0.741)
CidadeOutra	-440.992*** (0.00000)	484.217 (370.147)	-2.123 (1.399)
Constant	5.713*** (2.033)	-442.711* (246.765)	3.898* (2.016)
Akaike Inf. Crit.	336.637	336.637	336.637
Note:			
*p<0.1; **p<0.05; ***p<0.01			

APÊNDICE E – Regressão Multinomial Modelo 2

## =====			
##	Dependent variable:		
##	-----		
##	Modo ativo	Transporte público fretado	Transporte público (ônibus)
##	(1)	(2)	(3)
## -----			
## Tempo_idaB	-0.697	53.540***	1.199*
##	(0.484)	(0.022)	(0.671)
##			
## Tempo_idaC	0.282	103.378	4.625***
##	(0.780)	(384.623)	(0.888)
##			
## Tempo_idaD	-37.493	152.774**	6.268***
##		(77.280)	(1.875)
##			
## IdadeB	0.081	-34.376***	-0.223
##	(0.575)	(4.028)	(0.583)
##			
## IdadeC	-0.843	-126.285*	-2.635***
##	(0.565)	(76.711)	(0.713)
##			
## SexoMasculino	0.857*	29.297	-0.623
##	(0.477)	(89.597)	(0.564)
##			
## Renda_famDe seis a nove	-1.608**	29.404***	-1.365*
##	(0.712)	(0.000)	(0.794)
##			
## Renda_famDe três a cinco	-1.552**	73.347	-0.979
##	(0.611)	(76.689)	(0.657)
##			
## Renda_famMais de dez	-1.918***	163.565	-1.371*
##	(0.716)	(384.054)	(0.821)
##			
## CidadeOutra	-217.580***	178.173**	-3.855**
##	(0.00000)	(76.689)	(1.694)
##			
## Constant	1.413**	-277.702	0.177
##	(0.665)	(307.366)	(0.830)
##			
## -----			
## Akaike Inf. Crit.	317.333	317.333	317.333
## =====			
## =====			
## Note:	*p<0.1; **p<0.05; ***p<0.01		

APÊNDICE F – Regressão Multinomial Modelo 3

```

## =====
##           Dependent variable:
##           -----
##           Modo ativo Transporte público (ônibus)
##           (1)         (2)
## -----
## Distancia_idaB -2.569***      -0.756
##                (0.954)      (1.270)
##
## Distancia_idaC -2.176*        1.080
##                (1.155)      (1.339)
##
## Distancia_idaD -2.495*        1.853
##                (1.425)      (1.402)
##
## Distancia_idaE -15.484         0.580
##                (355.838)     (1.253)
##
## IdadeB         -0.070         0.372
##                (1.107)      (0.980)
##
## IdadeC         -1.574        -2.925***
##                (0.988)      (0.976)
##
## Constant      2.775***        0.663
##                (1.028)      (1.193)
##
## -----
## Akaike Inf. Crit. 125.859      125.859
## =====
## Note:          *p<0.1; **p<0.05; ***p<0.01

```

APÊNDICE G – Regressão Multinomial Modelo 4

```

## =====
##           Dependent variable:
##           -----
##           Modo ativo  Transporte público (ônibus)
##           (1)         (2)
## -----
## Tempo_idaB -1.099      1.738
##              (0.707)    (1.262)
##
## Tempo_idaC -1.162      4.552***
##              (1.327)    (1.516)
##
## Tempo_idaD -0.533      4.395**
##              (1.515)    (1.724)
##
## IdadeB      0.146      0.550
##              (0.988)    (1.107)
##
## IdadeC      -1.274     -3.107***
##              (0.812)    (1.160)
##
## Constant    1.344*     -1.365
##              (0.740)    (1.267)
##
## -----
## Akaike Inf. Crit. 126.504      126.504
## =====
## Note:          *p<0.1; **p<0.05; ***p<0.01

```

APÊNDICE H – Regressão Multinomial Modelo 5

##	=====			
##	=====			
##		Dependent variable:		
##		-----		
##		Modo ativo	Transporte público fretado	Transporte público (ônibus)
##		(1)	(2)	(3)
##	-----			
## Distancia_idaB	-71.010***	79.983***	-65.505***	
##	(1.180)	(1.278)	(0.805)	
##				
## Distancia_idaC	-71.291***	-13.402***	-64.379***	
##	(1.341)	(0.000)	(0.851)	
##				
## Distancia_idaD	-149.514***	41.957***	-64.734***	
##	(0.000)	(0.761)	(0.802)	
##				
## Distancia_idaE	-70.211***	29.418***	-64.533***	
##	(1.365)	(0.717)	(0.861)	
##				
## IdadeB	0.532	-1.469	-2.259***	
##	(1.404)	(1.649)	(0.745)	
##				
## IdadeC	-3.133	-2.368	-2.995***	
##	(2.072)	(1.739)	(0.837)	
##				
## CidadeOutra	-87.218***	53.654***	-103.198***	
##	(0.000)	(0.717)	(0.000)	
##				
## Renda_famDe seis a nove	-4.130	-0.287	-2.550***	
##	(2.526)	(2.768)	(0.921)	
##				
## Renda_famDe três a cinco	0.188	0.092	-0.861	
##	(1.118)	(2.292)	(0.684)	
##				
## Renda_famMais de dez	-2.910	38.786***	-2.403**	
##	(3.154)	(0.761)	(1.048)	
##				
## SexoMasculino	2.261**	1.731	0.254	
##	(1.150)	(1.296)	(0.703)	
##				
## Constant	69.862***	-81.150***	68.274***	
##	(1.160)	(1.176)	(0.924)	
##				
##	-----			
## Akaike Inf. Crit.	224.205	224.205	224.205	
##	=====			
##	=====			
## Note:		*p<0.1; **p<0.05; ***p<0.01		

APÊNDICE I – Regressão Multinomial Modelo 6

##	=====			
##	=====			
##		Dependent variable:		
##		-----		
##		Modo ativo Transporte público fretado Transporte público (ônibus)		
##		(1)	(2)	(3)
##	-----			
##	Tempo_idaB	-1.888**	33.414***	1.331
##		(0.809)	(0.900)	(0.959)
##				
##	Tempo_idaC	-1.151	34.854***	4.235***
##		(1.260)	(0.978)	(1.126)
##				
##	Tempo_idaD	0.372***	93.419***	85.827***
##		(0.000)	(0.000)	(0.000)
##				
##	IdadeB	-0.400	0.192	-1.831**
##		(0.906)	(2.607)	(0.784)
##				
##	IdadeC	-3.124***	-17.008***	-3.332***
##		(1.208)	(0.00001)	(0.990)
##				
##	CidadeOutra	-40.894	0.384	-81.976***
##		(1.942)		(0.000)
##				
##	Renda_famDe seis a nove	-4.056***	-15.988***	-2.339**
##		(1.531)	(0.000)	(1.064)
##				
##	Renda_famDe três a cinco	-0.535	14.490***	-0.600
##		(0.846)	(1.585)	(0.810)
##				
##	Renda_famMais de dez	-2.360*	15.029***	-1.654
##		(1.272)	(1.170)	(1.087)
##				
##	SexoMasculino	2.322**	2.729*	0.099
##		(0.972)	(1.560)	(0.857)
##				
##	Constant	1.975*	-50.565***	0.896
##		(1.047)	(0.986)	(1.125)
##				
##	-----			
##	Akaike Inf. Crit.	205.963	205.963	205.963
##	=====			
##	=====			
##	Note:	*p<0.1; **p<0.05; ***p<0.01		

APÊNDICE J – Regressão Multinomial Modelo 7

##	=====		
##		Dependent variable:	
##		-----	
##		Moto	Transporte público fretado
##		(1)	(2)
##	-----		
##	CidadeOutra	-5.105***	36.749***
##		(0.00000)	(0.0001)
##			
##	End_origemTrabalho	2.996**	11.617
##		(1.284)	(86.012)
##			
##	Constant	-1.609**	-11.617
##		(0.632)	(86.000)
##			
##	-----		
##	Akaike Inf. Crit.	38.631	38.631
##	=====		
##	Note:	*p<0.1; **p<0.05; ***p<0.01	