An investigation of trip-chaining behaviour based on activity participation, socioeconomic variables and aggregated characteristics of modal alternatives

Investigação do comportamento relativo a viagens encadeadas a partir de participação em atividades, variáveis socioeconômicas e características agregadas das alternativas modais

Viviani Antunes Gomes[1], **Marina Urano de Carvalho Caldas**[2], **Cira Souza Pitombo**[3]

[1]University of São Paulo, São Paulo – Brasil, viviani_antunes@yahoo.com.br
[2]University of São Paulo, São Paulo – Brasil, marina.urano.caldas@usp.br
[3]University of São Paulo, São Paulo – Brasil, cirapitombo@usp.br

ABSTRACT
This study investigates the individual trip-chaining behaviour for a sample of workers considering activity participation and socioeconomic characteristics. It proposes the inclusion of aggregated travel mode characteristics obtained from the Classification and Regression Tree (CART) algorithm and Origin-Destination (OD) Survey (Revealed Preference data RP) using data from São Paulo city, collected in 2007. Specifically, it proposes to: (1) classify individuals into clusters that have similar trip-chaining patterns; (2) perform a characterisation of modal alternatives from RP data; (3) propose an auxiliary criterion for formulating the Multinomial Logit model and reducing the number of parameters to be estimated; and (4) measure the improvement of estimates by including alternative characteristics (average travel times). Thus, this paper is associated with the following research gaps (1) the absence of data, associated to alternatives in an RP survey; (2) the lack of a criterion for utility function composition for the case of a large set choices; and (3) modelling based on the activity-based approach. Taking into account the findings, the feasibility of the proposed methodological procedure was verified despite of the technique constraints. Additionally, the model improved by including the alternative variables and corroborated relationships based on the literature.

RESUMO
Este trabalho investiga o comportamento individual, relativo a viagens encadeadas, para uma amostra de trabalhadores através de participação em atividades e de características socioeconômicas. Além disso, este trabalho propõe a inclusão de características modais agregadas, obtidas através do algoritmo Classification And Regression Tree (CART) e da Pesquisa Origem-Destino (Preferência Revelada - PR) de 2007 associada ao município de São Paulo. Especificamente, se propõe a: (1) classificar os indivíduos em grupos que apresentem similares padrões de viagens encadeadas; (2) realizar uma caracterização das alternativas modais a partir de dados de PR; (3) propor um critério para formulação do modelo Logit Multinomial a partir da redução de parâmetros a serem estimados; e (4) mensurar o aprimoramento das estimativas através da inclusão de características das alternativas (tempos médios de viagens). Assim, este estudo está associado às seguintes lacunas: (1) a ausência de dados relativos às alternativas em Pesquisa de PR; (2) a falta de um critério para composição das funções utilidade para o caso de grande conjunto de alternativas e (3) modelagem com base na abordagem de demanda por viagens baseadas em atividades. Levando-se em conta os resultados obtidos, verificou-se a viabilidade do procedimento metodológico proposto considerando as restrições da ferramenta utilizada. Adicionalmente, a modelagem paramétrica apresentou melhoria no modelo a partir da inclusão da variável que caracteriza a alternativa, além de corroborar relações embasadas pela literatura.

1. INTRODUCTION

The need for realistic representation of individual travel behaviour is well known in the current literature. Over the years, some approaches have contributed to better behavioural representation regarding individual displacements, considering two main ones: (1) the trip-based approach and (2) the activity-based approach (Bhat and Koppelman, 1999).

The first approach, usually associated with the traditional four-step model, considers trips as units of analysis, not representing the dependency between all trips made by the same individual over the course of a day, in addition to disregarding the schedule (spatial and temporal) of the trips sequence. Thus, a new approach was initiated, based on the assumption that travel demand is derived from a schedule of activities to be carried out, also considering a time-space organisation (Jones, 1983; Jones *et al.*, 1990; Ettema, 1996; Hao *et al*, 2010; Bradley *et al*, 2010; Pinjari and Bhat, 2011). Travel is often interconnected and interdependent. The sequence in which the trips are made, known as trip chaining, relates to the individual's lifestyle, socioeconomic characteristics and household, in addition to the daily activities to be performed and the factors related to travel and/or transportation systems (Bhat and Koppelman, 1999; Primerano *et al.*, 2008; Pitombo *et al.*, 2011). The trip-chaining behaviour analysis, which is the focus of this paper, is included in the activity-based approach.

Discrete choice models, one of the tools used in the activity-based approach, are primarily focused on individual behaviour and seek to identify the variables that influence the individual's decision in a choice process (McFadden, 1973).

An individual's decision-making process is complex and based on choices that can be characterised through utility functions that consider individuals´ and alternatives´ characteristics. Origin-Destination (OD) surveys, as an important source of Revealed Preference (RP) data, are instruments for numerous studies in the transport planning field, as they describe the real choices of individuals, thus allowing for future projections. However, the information obtained in the OD Survey refers only to data on the trips made, not characterising the other alternatives available to the individual. In order to improve the model, it is essential to insert variables that characterise the possible alternatives (Souza *et al.*, 2017; Fezzi *et al.*, 2014). A model can be estimated only considering variables related to the individuals, varying the values of the parameters associated with the individuals in each utility function. However, this would not be the most suitable condition for estimating the probabilities of the choices.

Thus, some studies have proposed the aggregated characterisation of alternatives based on RP data and empirical criteria, associated to the choice of variables and cut-off values (Kato *et al.*, 2013; Fezzi *et al.*, 2014; Souza *et al.*, 2017). It can be observed that usually the travel time and travel cost are used, and that the clustering criteria in these studies were defined often subjectively.

In addition, using models such as *Logit Multinomial* in problems related to the activity-based approach, is linked to some constraints, such as: a large number of alternatives to be analysed, possibility of violating of independence of irrelevant alternatives (IIA) property and unobserved heterogeneity (Bhat and Koppelman, 1999). The main constraint, associated with this study, is the large choice set. These problems generally lead to extensive collections of information related to the decision maker and to the choice set and computational constraints related to the calibration of such models.

Thus, the present research is contextualised with the following research problems, mentioned in the previous paragraphs: (1) need for a better representation of individual travel behaviour using the activity-based approach and (2) application constraints of *Logit Multinomial* models associated with both the available data and the approach. Thus, the second problem deals with the constraint focused on the use of RP survey and the analysis of different aspects of trip-chaining, resulting in the absence of data related to the alternatives and a large choice set. Figure 1 summarises and illustrates the research problems involved, contributions and the research gaps.
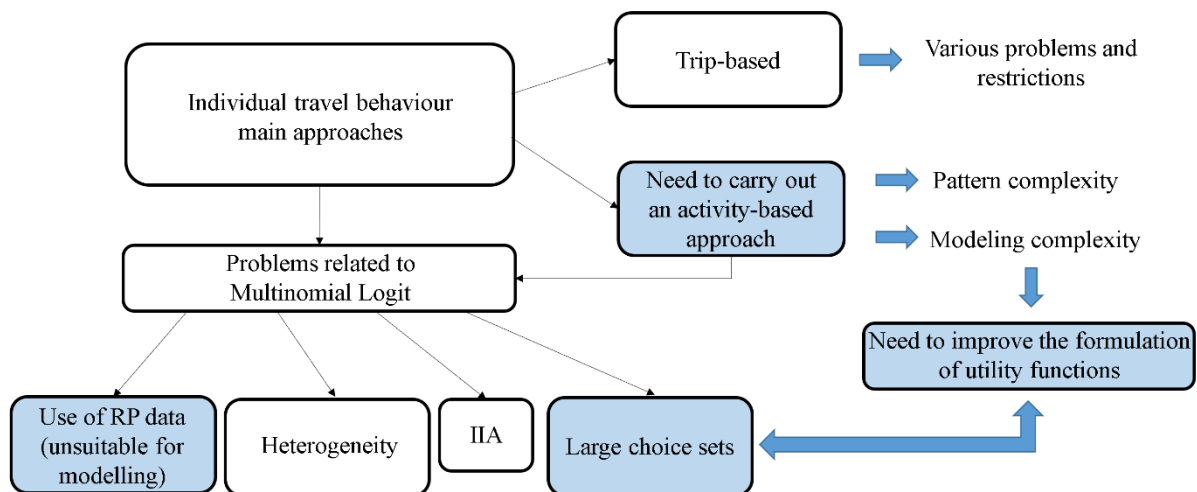

Figure 1. Summary of the research problems involved

Considering this background, this research investigates the trip-chaining behaviour for a sample of workers, by activity participation, socioeconomic characteristics, as well as proposing the inclusion of aggregated modal characteristics, obtained through the Classification and Regression Tree (CART) algorithm and the OD Survey conducted in the São Paulo Metropolitan Area (SPMA) in 2007.

In order to meet the main objective, this article presents some specific objectives: (1) to classify individuals into clusters that present similar travel patterns, using the travel modes and travel purposes sequences (dependent variable/choice set); (2) to characterise the alternatives based on RP data; (3) to propose an auxiliary criterion to formulate the *Logit Multinomial* model ; and (4) to incorporate the aggregated variables, obtained in (2), and measure the improvement in estimates through the *Logit Multinomial* model. It is noteworthy that the clusters obtained in (1) constitute the choice set and the aggregated travel mode characteristic was the average travel time for five travel modes.

Through this study, we aimed to solve three important issues in travel behaviour studies: (1) the absence of data related to alternatives in RP Survey; (2) the lack of a criterion for the composition of utility functions in the case of a large choice set and (3) modelling based on the activity-based approach.

The study was based on data from the OD household survey conducted in the São Paulo Metropolitan Area (SPMA) in 2007. It was used the area of the city of São Paulo and a sample consisting exclusively of workers.

This article has four sections, as well as this introduction. Section 2 provides a brief literature review considering the two research problems linked to this study. Section 3 describes the database, computational packages used and the methodological steps. Section 4 presents the results obtained and the discussions based on the literature. Finally, Section 5 draws the conclusions regarding this study.

2. LITERATURE REVIEW

2.1. Activity-based approach and individual travel behaviour

Activity-based models have emerged as a need to review the trip-based conventional modelling, that do not represent the dependency between all trips made by an individual throughout the day. Theories, the conceptual basis, methods and empirical studies began to be developed from studies carried out by Hägerstrand (1970), Chapin (1971) and Cullen and Godson (1975).

Bhat and Koppelman (1999) relate some conceptual issues of the trip-based approach that limit its application. In these models, travel is the unit of analysis and the models are developed for isolated trips, without taking into account a spatiotemporal organisation of trips. Thus, the modal or departure times choices, in addition to the various stops made, are interrelated.

Regarding trip-based studies, aspects of the choices of a single isolated trip are considered as a dependent variable. Traditional examples of variables that represent the individual travel behaviour are: (1) modal choice through categorical variables (Bhat, 1995; Madan and Groenhout, 1987; Kim and Ulfarsson, 2004); (2) route and departure time choices (Ben-akiva and Bierlaire, 2003; Bajwa *et al.*; 2008); (3) total distance covered in one day; (5) discrete destinations choices (Caldas *et al.*, 2019; Kitamura, 1984; Shrewsbury, 2012).

In an activity-based approach, several authors began to analyse aspects of choices aimed at the daily schedule of activities performed by individuals. Thus, individual behaviour began to be analysed by trip-chaining. Bhat and Koppelman (1991) defined trip chaining as an appropriate sequence of trips, carried out with the proposal of fulfilling various activities related to an individual's program of activities. Figure 2 illustrates examples of two trip chaining patterns, characterising trip purposes, travel modes and periods of the day.
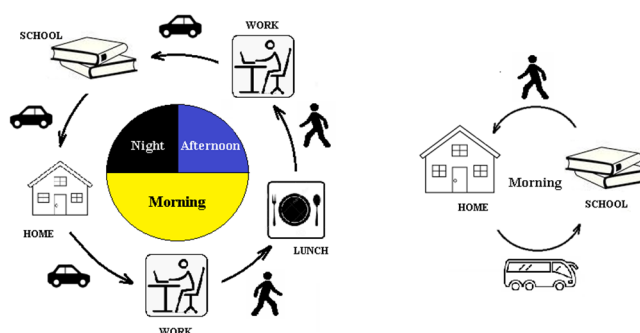

Figure 2. Examples of trip chaining patterns

Thus, over the years, the authors have characterised trip chaining based on the activity diary (Ben-Akiva and Bowman, 1995; Bifulco *et al.*, 2010), representing different aspects of trip chaining choices in a single day (Pitombo *et al.*, 2011; Scheiner and Holz-Rau, 2017) or on multiple days (Assirati and Pitombo, 2019).

2.2. Using *Logit* models in the activity-based approach

Logit models are the most widely used to model activity-based travel choices. Regarding the *Logit Multinomial* model, there are important assumptions that characterise it, such as the independence of the random terms of the utilities functions, the homogeneity of response between individuals and the homogeneity of the error variance (Bhat and Koppelman, 1999). These assumptions gave this model the IIA property, which disregards possible correlations among the alternatives (Luce and Suppes, 1965).

Thus, some *Logit* model variants were developed and applied for activity-based analysis, such as the *Probit*, *Nested Logit* and *duration models* (Bhat and Koppelman, 1999). These approaches are widely indicated for the modelling case for each of the trips in sequence (in this case, the nested models would be the most appropriate considering the dependence between trips made in sequence by the same individual).

The success in forecasting small set choice, as in the case of choosing the travel mode, led to applying the *Logit Multinomial* model in situations with a large number of alternatives (Bordley, 2013). However, the computational tools available are unable to calibrate the large number of parameters needed in disaggregated analyses that have many alternatives and/or many explanatory variables, or, when they do, they require a long time interval and robust computational configurations (Lemp and Kockelman , 2012).

In this context, many studies address this issue, therefore some strategies are adopted in order to alleviate it. McFadden (1978), for example, suggested that in the *Logit* models the estimates should be made from a set of the total of alternatives, generated randomly. Another strategy used is the use of sampling by importance to generate choice sets (Ben-Akiva and Lerman, 1985; Frejinger *et al.*, 2009). Although many researchers adopt the proposed methods mentioned above, there are caveats regarding random draws and the size of the sets of alternatives (Chen et al., 2005; Nerella and Bhat, 2004), and regarding the impossibility of choosing the alternatives in non-intuitive cases (Lemp and Kockelman, 2012), respectively. In many cases, however, there are variables that, although included in the model, have little or no influence on the quality of the model as a whole, causing little effect on its withdrawal with respect to the ability to represent a certain behaviour, but that, computationally, would generate a significant operational gain (Lemp and Kockelman, 2012).

3. MATERIALS AND METHODS
3.1. Summary of method and data used

This paper investigates trip-chaining behaviour for a sample of workers through activity participation, socioeconomic characteristics and aggregated modal characteristics. The latter obtained through the CART algorithm and OD Survey (RP).

Although trip-chaining is being analysed, there is no modelling for each of the trips in sequence. In this paper, all trips, made by an individual throughout the day, are represented by a single block (alphanumerical representation - Subsection 3.4). After alphanumeric representation of the trip-chaining patterns, performed by individuals, they are clustered using the TwoStep Cluster algorithm (Subsection 3.4). This algorithm identifies, through the Likelihood Log, individual clusters who perform similar trip-chaining patterns, considering the number of trips, trip purposes and travel modes.

The 15 clusters obtained through the TwoStep Cluster algorithm are the set of alternatives (totalling 15). Thus, the model estimates the probability of an individual belonging to a cluster (out of 15 possible ones) that performs a certain type of trip-chaining pattern (Subsection 3.6). Figure 3 shows a simplified illustration of the methodological procedures linked to the modelling stage, through *Logit Multinomial*.
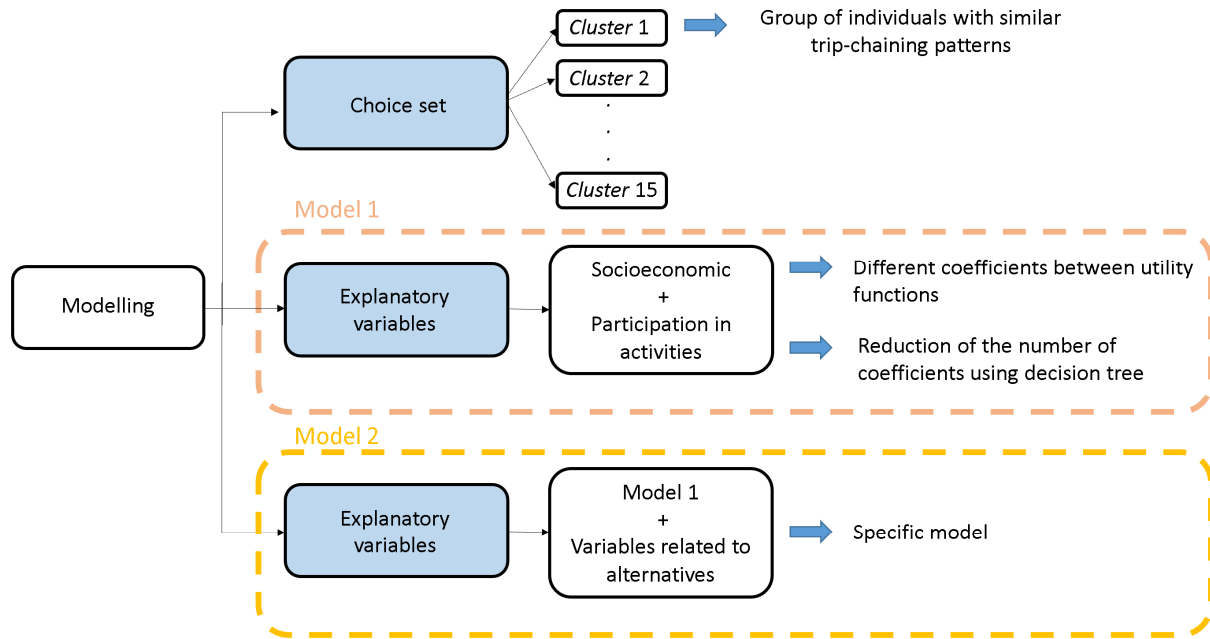

Figure 3. Methodological steps associated with parametric modelling
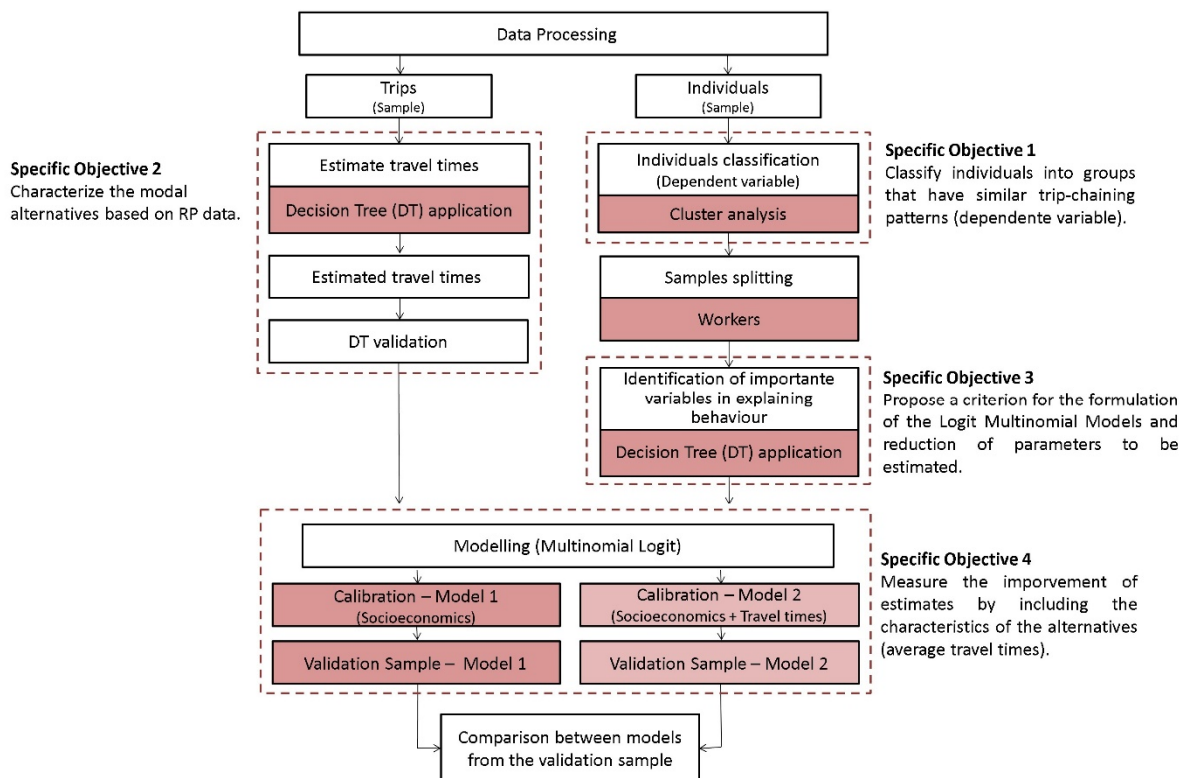

Figure 4. Methodological Procedure

Data from the OD Survey, conducted in the SPMA in 2007, were used. The SPMA consists of 39 municipalities and the region was divided into 460 Traffic Analysis Zones (TAZ's), 320 of which represent the city of São Paulo, which was the focus of this study. During the 2007 OD Survey, information was collected from 30 thousand households, chosen at random, of which approximately 120 thousand people were interviewed. This study consists of four databases: aggregated by TAZ's, disaggregated by trips, disaggregated by households and disaggregated by individuals. Trips database was used to the characterisation of modal alternatives. The disaggregated data, relating to workers, comprising individual, household and trips characteristics. This second database was used in the parametric modelling step (Figure 3).

The entire proposed method comprises six steps, as follows. Figure 4 illustrates the complete sequence of the proposed methodological procedure and the subsections describe the illustrated steps.

3.2. Step 1 – Data Processing

In this step of the method, data were processed to obtain two final disaggregated samples, which were: (A) trips sample: to obtain travel times for all modal alternatives; (B) individuals sample: which is a composition of the other databases, and comprises individuals.

Initially, only data related to the city of São Paulo were selected and incomplete records were deleted, that is, without travel data. Afterwards, the variables that would be used during the analyses were selected. The variables used, associated with the methodological step, were: **(Step 2)** Trip purpose; Departure time; Time walking at the origin and destination; Trip distance; Travel time; Main travel mode; **(Steps 2 and 4)** Number of cars; Family income; Age; Gender; Education level; **(Step 3)** Trip purpose; Main travel mode; **(Step 4)** Level of Instruction; Family Situation; Study; Number of residents in the household; Number of motorcycles; Number of bicycles; Formally hired; Informally hired; Public servant; Self-employed; Employer; Liberal professional; Family business owner; Informal worker, they have another job.

3.3. Step 2 – Determination of aggregated travel times

The objective of this step is to estimate the travel times for all modal alternatives available in the studied area and to make the aggregated characterisation of travel times in the unused travel modes (specific objective 2). To do this, the CART (Classification And Regression Tree) algorithm was used. This technique was used as the main tool to achieve two specific objectives: to characterise the alternatives based on RP data (specific objective 2) and to propose an auxiliary criterion to formulate the *Logit Multinomial* (specific objective 3). For specific objective 2, the CART algorithm was applied with a numerical dependent variable (which was an estimation problem), while for specific objective 3, the dependent variable was categorical, characterising a classification problem.

CART consists of an algorithm that classifies, estimates and represents the relationships existing in a data set. The model is adjusted by successive binary divisions of the data, based on statements such as "If ... then ..." in order to obtain increasingly more homogeneous subsets in relation to the dependent variable (Breiman et al., 1984). It is a non-parametric algorithm and has a structure that resembles a tree. The total data set (root node) is separated by sequential divisions (child nodes), which occur up to the terminal nodes (or leaves), when the formation of any subcluster is no longer possible due to the determined stop rules (minimum number

of observations at the terminal node, maximum tree depth or maximum desired heterogeneity at the node, depending on the dependent variable). Using the data from the OD Survey, this step tests this algorithm in the proposal to estimate the travel time of the other travel modes available in the study area.

The main travel modes were clustered into five categories: private motorised mode (1), bus (2), metro or train (3), bicycle (4) and on foot (5). This procedure was previously proposed in the study conducted by Cerveira *et al.* (2018) and, through filters in the terminal nodes of the CART algorithm, average times were obtained, by travel modes, for different clusters of trips. Figure 5 illustrates an example of the procedure performed. Using the trips database and application of the CART algorithm and travel duration as dependent variable, the data is sequentially divided up to the terminal nodes (Node 4, Node 5, Node 6 and Node 7). In Terminal Node 4, for example, the filter of 500 trips by travel mode can be observed. Finally, the average travel times are calculated for each modal option for trips that comprise Terminal Node 4. The same procedure is repeated for the other terminal nodes obtained.
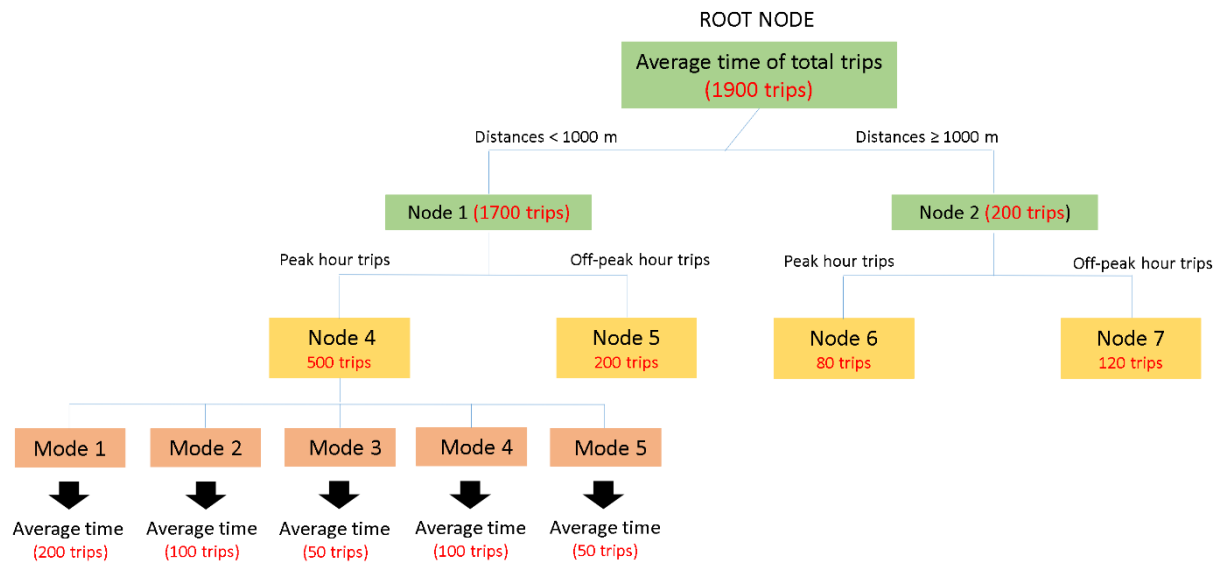

Figure 5. Illustration of the procedure adopted to obtain average travel times by travel mode

It is worth mentioning that the estimated travel times data, for the five travel modes, is represented by an individual in the database. Individuals who make similar trips, identified using the CART algorithm (contained in terminal node 4, for example), will have the same estimated average travel times. The term "aggregated", in this case, refers to the aggregation of trips contained in the same terminal node and does not mean aggregation of individuals.

The validation of this stage of the method was carried out considering only the travel times made by the travel modes effectively chosen by the individuals. Thus, the estimated travel times were compared with the values of the observed travel durations. Both travel times, used for validation purposes, were for the travel modes declared in the RP survey. The sample was randomly divided into training (70%) and testing (30%) and then the measurement errors and correlation with the test sample were calculated: Mean Square Error, Root Mean Square Error, Pearson's Correlation and Absolute Mean Error.

3.4. Stage 3 – Classification of individuals and division of the sample

This step aims to bring individuals into clusters that have similar trip-chaining patterns (specific objective 1). To do this, the database of individuals was used, and the clustering was performed using the Two-Step Cluster algorithm, contained in the IBM SPSS 24 software. The data used in this classification were related to the sequences of trip purposes and travel modes, with a maximum limit of four daily trips made by the same individual. The classification of individuals, according to their linked travel patterns, was the dependent variable used in the parametric modelling stage.

This step consisted of three substeps: (1) defining the trip-chaining patterns; (2) using the TwoStep Cluster algorithm and obtaining the main clusters; (3) determining the nomenclature and/or classification of individuals.

Trip-chaining pattern definition

The first step in determining the dependent variable or classification of individuals was to define the trip-chaining patterns. In this research, travel patterns, which characterise individual behaviour regarding urban travel, were defined by an alphanumeric sequence that represents the trip purposes and travel modes, according to the procedure proposed by Ichikawa *et al.* (2002) and Pitombo *et al.* (2011).

The characteristics of the trips, considered for the composition of the trip-chaining patterns, were adapted and clustered from the original data, from the OD Survey of 2007. In this study, the following nomenclature was adopted: Work (W), Study (S), Household (H), Other activities (A), Motorised private (1), Public (2) and Non-motorised (3).

The letters and numbers indicate the characteristics of the trip sequence taken by the individual on the day before the interview and represent their trip-chaining pattern. From the sample of 4,952 individuals, 284 patterns (combinations) were generated.

"SHWH2211" is an example of a pattern for an individual who made four trips. The first four digits represent the trip purpose sequence ("SHWH"). This sequence indicates that the individual went to school, then returned home, then went to work and, finally, returned home. The second numerical sequence indicates the order of the travel modes used. Thus, "2211" indicates that the first two trips were made by public mode and the last two by private motorised mode. It is worth mentioning that the first and last trips are based on the household.

Application of the TwoStep Cluster algorithm and obtaining the main clusters

After determining the trip-chaining patterns encoded in the previous step, clusters were generated from 2 categorical variables: (1) Sequence of trip purpose (example SH - first trip to school and second trip back home); (2) Sequence of travel modes (example 33 - two trips made by non-motorised mode). The idea of this methodological substep is to explore the similarities of travel patterns using a clustering technique.

The Two-step Cluster algorithm, available in the IBM SPSS 24 package, is a scalable cluster analysis algorithm, designed to handle large data sets. In this study, we opted to apply this clustering method because it clusters categorical variables, using the Likelihood Logarithm as a measure of similarity. In addition, it requires only one data processing, and has two steps: (1) pre-cluster the cases (or records) into many small sub-clusters and (2) cluster the sub-clusters resulting from the pre-cluster to the desired level of clusters. The number of clusters can also be selected automatically (IBM SPSS, 2012).

As it is an exploratory technique, several attempts are foreseen based on the variation in the number of clusters obtained. In this article, several attempts were made to identify the best clusters according to trip-chaining patterns. After analysing the different clusters obtained in the various attempts, the clustering that was able to identify characteristics specific to each cluster was chosen. Finally, 15 clusters were chosen.

Determining nomenclature and/or classification of individuals

In this step, was possible to classify individuals according to similar characteristics inherent to each cluster.

3.5. Stage 4 – Identification of important variables from the CART algorithm

Once the dependent variable was determined (15 clusters obtained in the previous step), the *Logit Multinomial* model was formulated, considering 15 utility functions. The high number of utility functions and parameters to be estimated can be a computational constraint in the calibration step of the discrete choice model. For this reason, this paper proposes to analyse and select potentially non-relevant variables to be excluded in each utility function. The aim is to reduce the number of parameters to be estimated later. This stage comprises the third specific objective of the present study, in which we propose to apply a simplified criterion for the formulation of the *Logit Multinomial* model with a reduction in the number of parameters to be estimated.

In this step of the method, the CART algorithm is also used. In this specific case, where the dependent variable corresponds to the 15 clusters previously obtained (categorical dependent variable), the problem is one of classification and the data is segmented in order to make the child nodes homogeneous according to the dependent variable categories. Having said that, this study proposed the methodological procedure comprising the following steps:

- Initially, the dependent variable and the independent variables are determined. In this study, the dependent variable was the clustering of individuals according to trip-chaining patterns (15 clusters of individuals) and the initial independent variables were socioeconomic characteristics and participation in workers´ activities (19 variables).

- Next, 19 bivariate trees are generated, with only one division of the total data set (dependent variable versus independent variable 1, dependent variable versus independent variable 2, …, dependent variable versus independent variable 19), as shown in Figure 6.

- For each tree generated (in this study, a total of 19), the variation of each one of the categories of the dependent variable can be observed between the two terminal nodes from the proposed division. Using the example in Figure 6, it can be seen that the independent variable 1, when it assumes values less than or equal to 0.5, positively influences the presence in cluster 2 (12% of the observations of Node 1 refer to it) . This can be verified by observing that when the same variable assumes values greater than 0.5, the presence of cluster 2 between trips classified in node 2 is very small (0.5%). Thus, there is a high proportional variation (95.83%) in category 2 (cluster 2), by dividing the parent node from the independent variable 1. For the case of cluster 15, however, it can be observed that the presence of this cluster is similar between child nodes 1 and 2 (1.3% and 1.3%, respectively), indicating that independent variable 1 does not have a significant

influence on the characterisation of cluster 15 (variation of 0%), which can then be excluded from its utility function.

- For each of the independent variables, the alternatives/categories are identified in which their proportional variation between the child nodes differs by up to 10%. The objective of this method is to identify, for each independent variable, relatively high variations in the categories of the dependent variable. These cases signal an influence and/or relevance of a given explanatory variable on the utility of a specific alternative.

It is worth mentioning that the proposed procedure is simplified and that future validation is necessary in order to assess whether the variables taken from certain utility functions would actually be associated with the estimation of non-significant parameters, for a 95% confidence level.
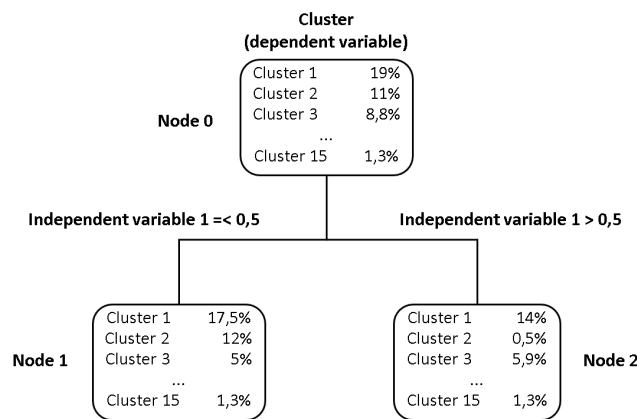
Cluster
(dependent variable)

Node 0	Cluster 1	19%
	Cluster 2	11%
	Cluster 3	8,8%
	...	
	Cluster 15	1,3%

Independent variable 1 =< 0,5   Independent variable 1 > 0,5

Node 1	Cluster 1	17,5%
	Cluster 2	12%
	Cluster 3	5%
	...	
	Cluster 15	1,3%

Cluster 1	14%	Node 2
Cluster 2	0,5%	
Cluster 3	5,9%	
...		
Cluster 15	1,3%	

Figure 6. Example of the procedure adopted with CART to reduce the number of parameters

3.6. Step 5 –*Logit Multinomial* Modelling

In this step, referring to specific objective 4, the Biogeme software (Bierlaire, 2018) was used and the utility functions of the *Logit Multinomial* models were defined, based on socioeconomic variables, participation in activities and the previously proposed criterion, which defined the important variables to be considered in each utility function. The sample was randomly divided into calibration (70%) and validation (30%). Thus, two models were generated. The first contained only socioeconomic variables and participation in activities, while the second model incorporated the aggregated travel time variables. The description of the models is presented below.

Model 1: In this model, the dependent variable was the classification of individuals, generated from Step 2 of the method. Equation 1 corresponds to the example of the Utility 1 function. In this article, 15 utility functions were calibrated, in which Utility 15 was the reference.

$$V_1 = ASC1 + B1\_Household*no\_resid + B1\_No\_CAR*no\_car + B1\_FAM\_INC*fam\_inc \quad (1)$$

ASC1 = Constant related to the Utility 1 function; Bx_Household parameter to be estimated for the variable "number of residents in the household" of the utility function x, where x can vary from 1 to 14; Bx_No_CAR parameter to be estimated for variable "number of cars at home", of utility function x, where x can vary from 1 to 14; Bx_FAM_INC: the parameter to be estimated for the "family income" variable, of the utility function x, where x can vary from 1 to 14; no_resid: number of residents in the household; no_car: number of cars at home; Fam_Inc: family income.

Model 2: In Model 2, in addition to the socioeconomic variables and activity participation, listed above, the estimated aggregated travel times were inserted, with specific parameters.

3.7. Comparison between models based on the validation sample

To compare the models, part of the sample (30%) was used, which was randomly selected, and the quality of the fit of models 1 and 2 was measured using the adjusted R-squared, the Likelihood value, the Log-likelihood and the Akaike Information Criterion.

4. RESULTS AND DISCUSSION
4.1. Aggregated characterisation of alternatives: Estimated travel times (Specific objective 2)

In the study carried out by Cerveira *et al.* (2018), the aggregated characterisation of the alternatives is illustrated, by estimating, by clustering trips (terminal nodes), the average travel times by modal alternative. In the CART algorithm, for the binary data partition, the minimum stop observations at the end node equal to 30 observations were used as the stop criterion. As a result, a total of 57 nodes were obtained, 29 of which were terminal nodes, and a depth equal to 5. Thus, the trip database was clustered into 29 similar travel clusters, according to duration (in minutes). The estimated times correspond to the average travel times, by travel mode, of the 29 terminal nodes obtained from the CART algorithm. Examples of terminal nodes (nodes 35, 51 and 52) associated with the cutting conditions of the independent variables and the average time related to each travel are shown as follows.

- **(Terminal node 35)** Travel distance> 921 m and Trip Purpose = School, Leisure, Personal Purpose or shopping - Car (13 min); Bus (25 min); Metro or Train (25 min); Bicycle (17 min); On foot (22 min).
- **(Terminal node 51)** 7,700 m <Distance <= 11,200, Trip Purpose = Home, School, Leisure, Personal Purpose or Shopping and Departure time (6am to 8pm) - Car (45min); Bus (69 min); Metro or Train (58 min); Bike (-); On foot (-).
- **(Terminal node 52)** 7,700 m <Distance <= 11,200, Trip Purpose = Home, School, Leisure, Personal Purpose or Purchases and Departure time (After 8 pm) - Car (33 min); Bus (65 min); Metro or Train (57 min); Bike (-); On foot (-).
- (The algorithm identified the following variables as being important for segmenting the trips database: "D: distance", "D_cl: departure time clusters", "PO: Purpose of origin" and "PD: Purpose of destination".

The estimated travel times were associated with the terminal nodes, found in the fourth and fifth levels of the tree. From the data generated by the tree, filters were made and identified, in each terminal node, the travel times for the five travel modes (1: Private Motorised; 2: Bus; 3: Metro and Train; 4: Bicycle; 5: On foot).

Considering the observed and estimated values of travel times for the travel modes actually used from the test sample, the following measures were obtained to validate this step: 378,677 for Mean Square Error, 19.46 for Root Error Quadratic Mean, -0.065 for Mean Absolute Error and Pearson's correlation was 0.638. The results indicate a good estimate of travel times and the procedure for the aggregated characterisation of alternatives was tested based on the approach previously proposed by Souza et al. (2017).

4.2. Classification of individuals (Specific objective 1)

The classification of individuals corresponds to the identification and clustering of individuals who show similar displacement behaviours (similar trip-chaining patterns). The clustering was obtained from 2 categorical variables: (1) sequence of trip purpose and (2) sequence of travel modes.

Several attempts were made to identify the best clustering. In this study, an adequate number of 15 clusters was considered to characterise the displacements of the sample comprising workers.

Thus, 4 clusters corresponding to individuals who made two trips, 4 clusters corresponding to individuals who made three trips and 7 clusters associated to individuals who made four trips were obtained. Table 1 describes in detail the clusters of the sample of workers.

Table 1 – Description of the classification of individuals according to trip-chaining patterns

Cluster	Description	Nº of trips	Pattern	Nº of registers	%
1	Go to work and come home. Use public mode.	2	WH22*	939	19%
2	Go to work and come home. Predominantly use non-motorised mode	2	WH23 WH13 WH33	706	14%
3	Other activities or school. Predominantly uses the same travel mode to make both trips	2	AH11 AH33 SH22	437	9%
4	Other activities linked with work. Predominantly uses the same travel mode to make all trips	3	WAH111 AWH222 AWH333	203	4%
5	He/She has two jobs. Predominantly uses the motorised mode to make trips.	3	WWH222 WWH111 WWH221	127	3%
6	Go to work and come home. Uses private motorised mode.	2	WH11*	1.297	26%
7	Other activities, work and school chaining. Predominantly uses the same mode to make all trips.	3	SWH111 WSH222 SAH111	172	3%
8	Other activities. Predominantly uses motorised mode	3	AAH111 AAH222 AAH333	62	1%
9	Work interspersed with household. He/She uses the private motorised mode to make the four trips.	4	WHWH1111*	83	2%
10	Work interspersed with residence. Uses non-motorised mode to make the four trips.	4	WHWH3333*	89	2%
11	Other activities, school and work always interspersed with home. He/She uses the same motorised mode to make the four trips.	4	WHAH1111 AHWH2222 WHSH1111	211	4%
12	Other activities, school and work always interspersed with home. Uses non-motorised mode to make the four trips.	4	SHSH3333 SHWH3333 WHAH3333	65	1%
13	Other activities, school and work always interspersed with home. Predominantly uses the motorised mode to make the four trips	4	WHAH2211 WHWH1122 WHAH3111	198	4%
14	Other chaining activity, school and work. He/she uses the same mode to make the four trips.	4	WAWH3333 WWAH1111 WAWH2222	203	5%
15	Other chaining activity, school and work. Predominantly uses the non-motorised mode.	4	WAWH1331 AAAH3331 WAAH2333	160	3%
				4.952	100%

*only one trip-chaining pattern

Among individuals who make two daily trips (clusters 1, 2, 3 and 6), the predominance of the Home - Work - Home -WH pattern (clusters 1, 2, and 6) can be observed, where there is a variation between the travel modes used, and the private motorised mode was the most frequent (26% of the individuals perform the "WH11" Pattern). Despite being a sample of workers, cluster 3 presents individuals with trip purposes associated with school or other activities (AH or SH), predominantly using the private motorised mode for making such trips.

Among individuals who make three trips (clusters 4, 5, 7 and 8), those who make trips for work and other activities (WA or AW) and do not change the travel mode (cluster 4) were identified. Cluster 5 features individuals who make two consecutive trips to work, whereby 27% of individuals in this cluster have two different jobs. Cluster 7 probably represents workers who also carry out study activities, since it is formed by individuals who make trips from work to school or from school to work (WSH or SWH). Finally, Cluster 8, despite comprising workers, represents individuals who make consecutive trips to perform activities other than work (AAH). This cluster is made up of 40% self-employed.

Clusters 9, 10, 11, 12, 13, 14 and 15 comprise people who make 4 trips. Clusters 9 and 10 represent the individuals who perform the trip-chaining pattern associated with work and home, interspersing both activities (WHWH). Possibly, these individuals return to the household for lunch. There is only variation between the travel mode used between these two clusters. Cluster 10 is formed by those who use non-motorised modes and must live close to the workplace. Clusters 11, 12 and 13 comprise individuals who do school and work activities interspersed with the household purpose (WHAH, SHWH, AHWH, for example). The greatest variation between the three clusters mentioned is in the travel mode sequence. In these clusters, 17% of individuals study and 39% are hired formally. Finally, clusters 14 and 15, are made up of people who make 4 daily trips with purposes "other activities" and "school", finally returning to the household. Cluster 14 is formed by individuals who use the same travel mode to make the 4 trips. The individuals that characterise cluster 15, in turn, predominantly use (in different combinations) the non-motorised mode.

4.3. Identifying important variables for the composition of utility functions from the CART algorithm (Specific objective 3)

This methodological step is directly related to the specific objective concerning proposal of a criterion to formulate the *Logit Multinomial* Model with a reduction in the number of parameters to be estimated. Thus, by performing the steps described in Subsection 3.5, using the CART algorithm, variables potentially relevant or not for the composition of utility functions are identified. Table 2 represents which variables should be included in each of the utility functions from the application of the proposed procedure. In this study, 15 utility functions were developed, since the set of alternatives comprises the 15 clusters of individuals obtained in the previous stage. Red represents the variables that must be excluded from the specific utility functions. Variables to be included in the parametric modelling stage are shown in green. Utility function 15 was the reference function in this study.

Number of residents (1-14); Family situation (1 - responsible; 0 - other); Number of motorcycles (0 - 9), Number of cars (0 - 6); Number of bicycles (0 - 9); Family Income (0 - ≤ R$2,502.00; 1 -> R$2,502.00); Age (1 - up to 11 years; 2 - from 12 to 20 years; 3 - from 21 to 59 years; 4 - ≥ 60 years); Gender (1 - Female); Level of Instruction (1 - illiterate - Incomplete primary; 2 - Complete primary - Incomplete primary; 3 - Complete elementary - Incomplete high

school; 4 - Complete high school - Incomplete higher education; 5 - Complete higher education); Variables associated with activity participation - All binaries with a value of 1 corresponding to "Yes" (Study; Formally hired; Informally hired; Public servant; Self-employed; Employer; Liberal Professional Family Business Owner; Informal worker; They have another job).

Table 2 – Composition of the Utility Functions

Variables	Utility Functions														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Number of residents					X			X	X	X	X	X	X	X	
Family situation						X		X		X		X			
Number of motorcycles							X	X		X		X	X	X	
Number of cars			X	X	X		X	X				X			
Number of bicycles				X			X	X	X			X	X	X	
Study								X							
Family Income				X			X	X							
Age					X			X		X			X	X	
Gender					X		X	X	X	X	X	X		X	
Level of Instruction								X					X		
Formally hired				X						X	X	X			
Informally hired			X	X				X	X	X		X	X	X	
Public Servant					X	X		X	X		X	X	X	X	
Self-employed				X					X	X	X	X	X	X	
Employer				X	X			X			X	X			
Liberal Professional				X	X		X			X		X			
Family Business Owner		X					X	X		X	X	X	X	X	
Informal worker				X											
They have another job				X				X		X		X			

Variables included in the Utility Function

X Variables excluded from the Utility Function

4.4. Logit Multinomial Modelling (Specific Objective 4)
4.4.1. Model 1

Modelling 1 resulted in: adjusted R-squared equal to 0.25 and Akaike information criterion of 14,109.94. Considering the validation sample, the following measures were calculated: Hit rate of 46.54%, Likelihood value L = 9.97 x 10^{-16} and log (L) = - 34.54. Equations 2 to 15 are utility functions calibrated for the fifteen clusters of workers, with the estimated parameters that were significant for a 95% confidence level.

$V_1 = 2.14 + 0.20*no\_resid – 0.72*no\_car – 0.60*fam\_inc$ (2)

$V_2 = 2.93 + 0.16*no\_resid – 0.70*no\_car – 0.71*fam\_inc – 0.21*level\_ins – 0.75*public\_servant$ (3)
$– 0.76*liberal$

$V_3 = 1.47*study– 0.79*fam\_inc + 0.58*age – 0.15*level\_ins + 0.63*self-employed$ (4)

$V_4 = – 1.20 + 1.23*fam\_sit + 0.52*no\_mot + 0.49*sex$ (5)

$V_5 = – 1.86 + 1.26*fam\_sit + 0.54*no\_moto + 0.24*no\_bicycle + 2.53*have\_another\_job$ (6)

$V_6 = 1.62 – 0.07*no\_resid + 0.34*no\_moto + 0.63*no\_car – 0.52*sex$ (7)

$V_7 = – 1.34 + 0.22*no\_resid + 3.64*study$ (8)

$V_8 = – 0.50 – 1.32*formally\_hired$ (9)

$V_9 = – 6.22 + 0.57*no\_moto + 0.88*no\_car+ 1.21*age+ 0.78*employer + 1.55*have\_another\_job$ (10)

$V_{10} = 1.78 – 0.80*no\_car – 0.96*fam\_inc – 0.32*level\_ins$ (11)

$V_{11} = – 1.17 + 0.87*fam\_sit + 0.49*no\_moto + 0.47*no\_car + 1.50*study$ (12)

$V_{12} = 5.96 - 1.03*no\_car + 2.05*study – 1.18*age – 0.68*level\_ins$ (13)

$$V_{13} = 1.91*study + 0.66*employer + 0.68*liberal \qquad (14)$$

$$V_{14} = -1.55 + 1.27*fam\_sit + 0.42*no\_car + 1.55*study + 1.30*have\_another\_job \qquad (15)$$

In the analysis of modelling 1, it can be observed that:

1. In **Utility V_1 (associated with cluster 1 - WH22)**, there is a positive influence of the variable "number of residents", indicating that people belonging to larger family would be within that cluster and would be more likely to make the trips identified in this cluster, which corresponds to individuals who perform the Home - Work - Home by public transport (WH22) pattern. In addition, there is a negative influence of the variables "number of cars at home" and "family income". These relationships are strictly linked to modal choice. Traditionally, in the literature, such relationships have been corroborated (Hartgen, 1974; Barff *et al.*, 1982; Pitombo *et al.*, 2011; Ichikawa *et al.*, 2002).

2. **Utility V_2 (associated with cluster 2 - WH23; WH13 and WH33)** has similar relationships to those found in Utility 1, associated with cluster 1, including the negative influence of the variables "level of instruction", "civil servant" and "liberal professional".

3. **Utility V_3 (cluster 3 - AH11 patterns; AH33 and SH22)** is associated with individuals who study, who are older, have a lower income, with less education and greater propensity for people working as freelancers. The relationships found in this utility function through the estimated parameters, especially with regard to participation in study activities and the self-employed workers class, are consistent with the travel patterns prevalent in this cluster of individuals.

4. **Utility V_4** represents the cluster of individuals who make three trips, linking the trip purpose "other activities" to "work" (WAH; AWH; AWH). Thus, there are important relationships between the family situation, the number of motorcycles and the gender. It can be said, through these results, that female headed households are more likely to carry out the travel patterns that make up cluster 4. Similar relationships are found in the literature related to sex, family situation and trip chaining (Mcguckin and Murakami, 1999; Gordon *et al*, 1989).

5. **Utility V_5 (predominance of patterns WH222; WWH111; WWH221)** has similar relationships to those found in V_4, with the exception of the variable "gender", which is not indicated in this utility, and the positive presence of the variables "number of bicycles at home" and "they have another job", indicating the presence of individuals who have more than one job. The coefficient of the variable "they have another job" is positive and strongly related to the greater propensity of belonging to this cluster of travellers, who make consecutive trips for work and to different locations.

6. In **Utility V_6 (associated with cluster 6 - WH11),** there is a negative influence of the variable "number of residents", indicating that people belonging to smaller family clusters would be within that cluster. In addition, there is a positive influence for men, as well as for people with a greater number of cars and motorcycles in the household. The relationships found here are essentially associated with the choice of car, sex, the size of the household and car ownership (Hartgen, 1974; Train and McFadden, 1978; Barff *et al.*, 1982).

7. **Utility V_7** points out that cluster 7 comprises individuals belonging to larger family clusters and who are studying. These relationships explain travel patterns linked between work and school activities (SWH111; WSH222, SAH111).

8. The parameters estimated in the **Utility 8** function are quite consistent with what is expected in relation to travel behaviour. The modelling shows that formally hired people are less likely to make the trips characterised in Cluster 8 (these trips have non-work trip purpose AAH111; AAH222; AAH333).

9. **Utility V_9** calibration highlights patterns with two trips to work, interspersed with the household (WHWH). Because this cluster comprises the sequence of car use, a positive relationship is observed for the variables "number of cars", for example.

10. **Utility V_{10}** comprises the same sequence of trip purpose as before, with the sequence of travel mode varying from non-motorised modes. Traditional relationships, found in the literature, related to modal choice are recognised. Thus, there is a relationship between the non-motorised mode and low income, less education and fewer cars at home (all parameters associated with these variables are negative).

11. **Utility V_{11}** characterises individuals who make four trips, interspersing "work" with "other activities" or "study" (WHAH1111; AHWH2222 and WHSH1111). In this cluster, it is observed that individuals responsible for the household are more likely to carry out such trip-chaining patterns. There is also an influence of vehicle ownership.

12. The modelling shows, through the calibration of Utility V_{12}, that people who study, who are younger, who have a lower level of instruction and who have few cars at home are more likely to make the trips described in cluster 12 (SHSH3333; SHWH333; WHAH333).

13. The estimated coefficients in **Utility V_{13}** indicate a greater propensity for people who make four linked trips to other activities, school and work, always interspersed at home and using the motorised mode (WHAH2211; WHWH1122; WHAH3111), who are people who study and work as employers or liberal professionals.

14. Finally, the **V_{14}** Utility calibration indicates a greater propensity for people who make four linked trips to other activities, school and work, always using the same travel mode (WAWH3333; WWAH1111; WAWH2222), who are heads of family, who study, who have more cars at home and who work in at least two different jobs.

In summary, most of the relationships found in the first stage of modelling, associated with participation in activities, socioeconomic characteristics and clusters of individual trip-chaining behaviours (Modelling 1), are corroborated in several works in the literature. Some authors have achieved similar results over the decades. According to a study developed by Blumenberg and Pierce (2014), for example, low-income Americans are less multimodal than those with higher income (V_1, V_2, V_3 e V_9). Hartgen (1974), Train and McFadden (1978) and Barff *et al*. (1982) have already proved the relationship between socioeconomic variables and modal choice, such as a greater propensity to use a car (V_6) for men, as well as for older people (V_3, V_6, V_9), a higher level of instruction and more cars at home (V_6, V_{11}). Additionally, Strathman *et al*. (1994) showed that the family structure affects the trip-chaining behaviour (V_4, V_5, V_{11}, V_{14}).

4.4.2. Model 2

In Model 2, the travel times of the five modal alternatives were incorporated into the models: (1) motorised private, (2) bus, (3) metro and train, (4) bicycle and (5) on foot. Specific coefficients associated with the travel times of the modes used by each of the clusters were used.

For a 95% confidence level, the significant parameters for model 2 of the sample of workers are indicated in the equations below (between Equation 16 and Equation 29). For this sample, there was an adjusted R-squared equal to 0.298 and Akaike information criterion of 13,171.62. With the validation sample, the following measures were calculated: a hit rate of 56%, likelihood value L = 1.78 x 10^{-14} and log (L) = - 31.66.

$$V_1 = 0.22*no\_resid - 0.76*no\_car - 0.57*fam\_inc + 0.04*t2\_bus \tag{16}$$

$$V_2 = 5.54 + 0.17*no\_resid - 0.73*no\_car - 0.73*fam\_inc - 0.14*level\_ins - 0.87*liberal - 0.13*t5\_walk \tag{17}$$

$$V_3 = 1.47*study - 0.80*fam\_inc + 0.53*age - 0.13*level\_ins + 0.62*self\text{-}employed \tag{18}$$

$$V_4 = -1.25 + 1.16*fam\_sit + 0.56*no\_moto + 0.51*sex \tag{19}$$

$$V_5 = -1.88 + 1.16*fam\_sit + 0.56*no\_moto + 0.26*no\_bicycle + 2.59*have\_another\_job \tag{20}$$

$$V_6 = 0.44*no\_moto + 0.66*no\_car - 0.41*sex + 0.04*t1\_privmot \tag{21}$$

$$V_7 = -1.50 + 0.25*no\_resid + 3.67*study \tag{22}$$

$$V_8 = -0.60 - 1.31*formally\_hired \tag{23}$$

$$V_9 = -6.19 + 0.62*no\_moto + 0.91*no\_car + 1.16*age + 0.70*employer + 1.60*have\_another\_job \tag{24}$$

$$V_{10} = 5.39 - 0.93*no\_car - 1.19*fam\_inc - 0.24*t5\_walk \tag{25}$$

$$V_{11} = -1.24 + 0.80*fam\_sit + 0.53*no\_moto + 0.49*no\_car + 1.52*study \tag{26}$$

$$V_{12} = 9.54 - 0.99*no\_car + 2.19*study - 0.95*age - 0.58*level\_ins - 0.25*t4\_bicycle \tag{27}$$

$$V_{13} = 1.86*study + 0.55*liberal \tag{28}$$

$$V_{14} = -1.55 + 0.74*fam\_sit + 0.39*no\_car + 1ar,51*study + 1.34*have\_another\_job - 0.04*t5\_walk \tag{29}$$

There was an overall improvement in the model after including aggregated travel times. The expectation was that the increase in travel time for a given modal alternative would negatively contribute to its utility, which occurred in cluster 2 (walking time), in cluster 10 (walking time), in cluster 12 (travel time by bicycle) and in cluster 14 (travel time on foot). However, it was found that in the Utility V_1 function, relative to cluster 1 (WH22) - pattern associated with public transport -, there was a low value and positive coefficient associated with the duration of the bus time. In the Utility V_6 function, relative to cluster 6, a low and positive coefficient is also observed, in this case associated with the time of the private motorised mode.

Table 3 – Summary of the relationships found between trip-chaining, activity participation, socioeconomic characteristics and travel times

Travel Mode Sequence	Number of cars at home Family income Level of instruction Gender
Variation of modal choice in the trip-chaining pattern	Public Servant Liberal Professional
Sequence of trips including purposes other than "Work"	Students Self-employed Professionals Formally hired professionals Liberal professionals
More complex trip chaining, intercalating other activities with work	Women Heads of families
Trip chaining with consecutive trips with "Work" purpose	Individual has two different jobs
Highest number of trips	Family income
Sequence of travel modes	Travel times by travel mode

Making a comparative analysis between Models 1 and 2, there was an increase in accuracy from the inclusion of characteristics of the alternatives. In the case of this study, it was the

aggregated travel times, resulting from a Revealed Preference survey. Table 3 summarises the main relationships found in Models 1 and 2.

5. CONCLUSIONS

The present study aimed to investigate the trip-chaining behaviour for a sample of workers through activity participation, socioeconomic variables and proposing the inclusion of aggregated modal characteristics, obtained through the CART algorithm and the OD Survey carried out in the São Paulo Metropolitan Area in 2007. The major contributions of this article are associated with the methodological procedure, taking into account the following factors:

- Although the OD Survey is used in various countries, the fact that it does not present information related to the alternatives that were not actually used by the individuals is a limiting factor to studies involving modeling. The aggregated travel times of all modal alternatives makes it possible to improve the modeling results;
- To obtain the aggregated characteristic of the alternative, a set of variables associated with travel was used and a criterion was proposed, based on a non-parametric algorithm (CART), to obtain average values of variables that characterise alternatives, which are more robust compared to the empirical data criteria previously used in the literature;
- The proposed procedure, also with the CART algorithm, in order to help the formulation of the *Logit Multinomial* model and reduce the number of parameters to be estimated is an important contribution to any research problem with a large set of alternatives, since there are constraints important operational due to the high number of parameters to be estimated in these cases;
- In the parametric modelling stage, the data indicated an improvement in the model from the inclusion of the variable that characterises the alternative (average travel times);
- Using the sequential use of Multivariate Data Analysis techniques, this study investigated individual travel behaviour, considering the activity-based approach.

REFERENCES

Assirati, L. and Pitombo, C.S. (2019) Classificação de indivíduos segundo comportamento individual relativo a viagens a partir de dados em painel obtidos por smartphones. *Transportes (Rio de Janeiro)*, v. 27, p. 165-181. DOI: 10.14295/transportes.v27i2.1679.

Bajwa, S.; Bekhor, S.; Kuwahara, M. and Chung, E. (2008) Discrete choice modeling of combined mode and departure time. *11th International Conference on Travel Behaviour Research.* August, Kyoto. DOI: 10.1080/18128600808685681.

Barff, R.; Mackay, D. and Olshavsky, R. W. (1982) A selective review of travel-mode choice models. *Journal of Consumer Research*, v. 8, p. 370–380. DOI: 10.1086/208877.

Ben-Akiva, M. and Bierlaire, M. (2003) Discrete choice models with applications to departure time and route choice. *The Handbook of Transportation Science*, v. 2, p. 7-38. DOI: 10.1007/0-306-48058-1_2.

Ben-Akiva, M.E. and J.L. Bowman (1995) Activity-based disaggregate travel demand model system with daily activity schedules, *EIRASS Conference on Activity-Based Approaches: Activity Scheduling and the Analysis of Activity patterns*, Eindhoven, The Netherlands.

Ben-Akiva, M. and Lerman, S. R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge, MA: MIT Press.

Bhat, C. (1995) A heteroskedastic extreme value model of intercity travel mode choice. *Transportation Research B*, v. 29, p. 471-483. DOI: 10.1016/0191-2615(95)00015-6.

Bhat, C. R. and Koppelman, F. S. (1991) A Conceptual Framework of Individual Activity Program Generation. *Transportation Research, Part A*, v. 27, n.6, p. 433-446. DOI: 10.1016/0965-8564(93)90050-U.

Bhat, C.R. and Koppelman, F.S. (1999) Activity-Based Modeling of Travel Demand. *Handbook of Transportaton Science*, vol 23. Springer, Boston, MA. DOI: 10.1007/978-1-4615-5203-1_3.

Bierlaire, M. (2018) *Pandas Biogeme: a short introduction. technical report.* TRANSP-OR 181219. Transport and Mobility Laboratory, ENAC, EPFL.

Bifulco, G. N.; Cartenì, A. and Papola, A. (2010) An activity-based approach for complex travel behavior modeling. *European Transport Research Review*, v. 2 (4), p. 209–221.

Blumenberg, E. and Pierce, G. (2014) Multimodal travel and the poor: Evidence from the 2009 National Household Travel Survey, *Transportation Letters*, v. 6, p. 36-45. DOI: 10.1179/1942787513Y.0000000009.

Bordley, R. (2013) Discrete choice with large choice sets. *Economics Letters,* v. 118, p. 13-15. DOI: 10.1016/j.econlet.2012.05.010.

Bradley, M.; Bowman, J. L., and Griesenbeck, B. (2010) SACSIM: Na applied activity-based model system with fine-level spatial and temporal resolution*, Journal of Choice Modelling*, v. 3, n. 1, p. 5–31. DOI: 10.1016/S1755-5345(13)70027-7.

Breiman, L.; Friedman, J. H; Olshen, R.A. and Stone, C. J. (1984) *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA.

Caldas, M. U. C.; Favero, R. and Pitombo, C. S. (2019) Uso de redes neurais artificiais para abordagem desagregada de distribuição de viagens urbanas com dados de preferência revelada. *In: 33o Congresso da Associação Nacional de Pesquisa e Ensino em Transportes*, Balneário Camboriú. Anais do 33o Congresso da Anpet, p. 2748-2749.

Cerveira, C. F.; Gomes, V. A. and Pitombo, C.S. (2018) Proposta de critério para caracterização agregada das alternativas modais a partir de dados de preferência revelada. *Anais do XXXII Congresso da ANPET,* Gramado, RS, p. 2449-2510

Chapin, F.S. (1971) Free-time activities and the quality of urban life, *Journal of the American Institute of Planners*, 37, p. 411–417. DOI: 10.1080/01944367108977391.

Chen, L. and Duann, W. Hu. (2005) The estimation of discrete choice models with large choice-set. *Journal of the Eastern Asia Society for Transportation Studies*, v. 6, p. 1724-1739

Cullen, I. and V. Godson (1975) Urban networks: the structure of activity patterns, *Progress in Planning*, v. 4, p. 1–96. DOI: 10.1016/0305-9006(75)90006-9.

Ettema, D. (1996) *Activity-based travel demand modeling*. Eindhoven. 280p. Tese (Doutorado em Planejamento Urbano), Universidade de Eindhoven, Holanda.

Fezzi, C.; S. Ferrini, and I. J. Bateman, (2014) Using revealed preferences to estimate the value of travel time to recreation sites*. Journal of Environmental Economics and Management*, v. 67, n. 1, p. 58–70. DOI: 10.1016/j.jeem.2013.10.003.

Frejinger, E.; M. Bierlaire and M. Ben-Akiva (2009) Sampling of Alternatives for Route Choice Modeling. *Transportation Research Part B*, 43, p. 984-994. DOI: 10.1016/j.trb.2009.03.001.

Gordon, P.; Kumar, A. and Richardson, H. (1989) Gender differences in metropolitan travel behavior*, Regional Studies* v. 23, p. 499–510. DOI: 10.1080/00343408912331345672.

Hägerstrand, T. (1970) What about people in regional science? *Papers of the Regional Science Association*, 24(1), p. 7–21. DOI: 10.1111/j.1435-5597.1970.tb01464.x.

Hao, J.; Hatzopoulou, M. and Miller, E. (2010) Integrating an activity-based travel demand model with dynamic traffic assignment and emission models*. Transportation Research Record: Journal of the Transportation Research Board*, v. 2176, n. 1, p. 1–13. DOI: 10.3141/2176-01.

Hartgen, D. T. (1974) Attitudinal and situational variables influencing mode choice: Some empirical findings. *Transportation,* v.4, p. 277-292. DOI: 10.1007/BF00167967.

IBM SPSS Inc. (2012) *IBM SPSS Statistics 21 Algorithms.* Chicago: Author

Ichikawa, S. M.; Pitombo, C. S. and Kawamoto, E. (2002) Aplicação de minerador de dados na obtenção de relações entre padrões de viagens encadeadas e características socioeconômicas. *Panorama Nacional de Pesquisa em Transportes. XVI Anpet.* v.2, p.175-186.

Jones, P. M. (1983) *The Practical Application of Activity-based Approaches in Transport Planning: An Assessment. Recent Advanced Travel Demand Analysis.* Aldershot. England. Gower.

Jones. P.; Koppelman. F. and Orfueil. J. P. (1990) Activity analysis: State-of-the-art and future directions. Developmentes in Dynamic and Activity-Based Approaches to Travel Analysis, Gower Publishing, Brookfield, Vermont.

Kato, H.; T. Oda and Sakashita, A. (2013) Valuation of travel time saving with revealed preference data in Japan: Further Analysis. *In 13th WCT. CPAPER*, Rio de Janeiro, Brasil.

Kim, S. and Ulfarsson, G. (2004) Travel mode choice of the elderly – effects of personal, household, neighborhood, and trip characteristics. *Transportation Research Record* ,1894, p. 117–126. DOI: 10.3141/1894-13.

Kitamura, R. (1984) Incorporating Trip Chaining into Analysis of Destination Choice. *Transportation Research B*18, p. 67-81. DOI: 10.1016/0191-2615(84)90007-9.

Lemp, J. D. and Kockelman, K. M. (2012) Strategic sampling for large choice sets in estimation and application, *Transportation Research Part A: Policy and Practice*, p. 602-613. DOI: 10.1016/j.tra.2011.11.004.

Luce, R. D. and Suppes, P. (1965) Preference, utility, and subjective probability. In R. D. Lute, R. R. Bush, & E. Galanter (Fds.), *Handbook of mathematical psychology*, v. 111, p. 249-410.

Madan, D.B. and Groenhout, R. (1987) Modelling travel mode choices for the Sydney work trip, *Journal of Transport Economics and Policy*, 21, p. 135-149.

McFadden, D. (1973) Conditional Logit Analysis of Qualitative Choice Behavior, *in Frontiers of Econometrics*, ed. by P. Zarembka. New York: Academic Press.

McFadden, D. (1978) Modelling the choice of residential location. *Transportation Research Record*, v. 673, p. 72-77.

McGuckin, N. and Murakami, E. (1999) Examining trip-chaining behavior: Comparison of travel by men and women. *Transportation Research Record: Journal of the Transportation Research Board*, v.1693, p. 79–85. DOI: 10.3141/1693-12.

Nerella, S. and C. R. Bhat (2004) Numerical Analysis of Effect of Sampling of Alternatives in Discrete Choice Models. *Transportation Research Record*, v. 1894, p. 11-19. DOI: 10.3141/1894-02.

Pinjari, A. R. and Bhat, C. R. (2011) Activity-based travel demand analysis, *A Handbook Transport Economics*, 1 ed., p. 1–36.

Pitombo, C. S.; E. Kawamoto and A. J. Soúsa (2011) An exploratory analysis of relationships between socioeconomic, land use, activity participation variables and travel patterns. *Transport Policy*, v.18, n.2, p. 347-357. DOI: 10.1016/j.tranpol.2010.10.010.

Primerano, F., M.A.P Taylor, L. Pitaksringkarn and P. Tisato (2008) Defining and understanding trip chaining behaviour. *Transportation* v. 35, p. 55–72. DOI: 10.1007/s11116-007-9134-8

Scheiner, J., Holz-Rau, C. (2017) Women's complex daily lives: a gendered look at trip chaining and activity pattern entropy in Germany, *Transportation*, v. 44, p. 117–138. DOI: 10.1007/s11116-015-9627-9.

Shrewsbury, J. S. (2012) *Calibration of trip distribution by generalised linear models*. New Zealand Transport Agency Research Report 473. New Zealand.

Souza, H.H.H.; F.F.L.M. Sousa; F.M. Oliveira Neto; R.M.C. Freire and C.F.G. Loureiro (2017) Estimação do valor do tempo com base em pesquisas domiciliares de origem e destino: desafios teóricos e dificuldades práticas. *Anais do XXXI Congresso da ANPET* – Recife, PE.

Strathman J. G.; Dueker K. J. and Davis J. S. (1994) Effects of household structure and selected characteristics on trip chaining. *Transportation* v.21, p. 23-45. DOI: 10.1007/BF01119633.

Train, K. and McFadden, D. (1978) The Goods/Leisure Tradeoff and Disaggregate Work Trip Mode Choice Models, *Transportation Research*, v.12, p. 349-353.