TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

THESIS NO: T08/074

Mode Choice Modeling of Graduate Level Engineering Students of Kathmandu Valley

by

Piyush Chataut

A THESIS

SUBMITTED TO THE DEPARTMENT OF CIVIL ENGINEERING

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE IN

TRANSPORTATION ENGINEERING

DEPARTMENT OF CIVIL ENGINEERING

LALITPUR, NEPAL

AUGUST, 2020

COPYRIGHT

TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

DEPARTMENT OF CIVIL ENGINEERING

The undersigned certify that they have read, and recommended to the Institute of Engineering for acceptance, a thesis entitled **“Mode Choice Modeling of Graduate Level Engineering Students of Kathmandu Valley"** submitted by **Piyush Chataut** in partial fulfillment of the requirements for the degree of Master of Science in Transportation Engineering.

____

Supervisor, Dr. Pradeep Kumar Shrestha
Assistant Professor
Department of Civil Engineering

____

External Examiner, Saroj Kumar Pradhan
Joint Secretary
Ministry of Physical Infrastructure and Transport

____

Committee Chairperson, Anil Marsani
Coordinator, M.Sc. in Transportation Engineering
Department of Civil Engineering

____

Date

ABSTRACT

A large part of trips in Kathmandu valley is occupied by educational trips, whose magnitude is bound to increase as the average time spent studying is increasing. The commutes of graduate-level engineering students follow a similar trend and thus will increase the demand for means of transportation, creating load to the suffering transportation system, and therefore proper planning interventions are needed.

This study forms mixed logit models with the help of the ‘R-Studio’ statistical package to analyze these trips. Models consist of revealed preference (RP) models to understand the current mode preference, and stated preference (SP) models to determine the modal shifts when different alternative modes of bus rapid transit (BRT) and metro trains are introduced, these SP choices were formed by the analysis of alternative transportation in different urban metropolises and consist of four options “Bus1”, a luxurious but costly BRT, “Bus2” normal BRT, “Metro1” a comfortable but costly metro, and lastly “Metro2” normal metro all concerning the existing public transportation modes. Data required to form these models were collected by performing RP and SP surveys.

For the RP part the use of personal mode of transport was found to be increased by transport trip distance and vehicle ownership and decreased by number of siblings. Similarly, the use of public transport was found to be increased by trip distance, walking being the base criteria for both cases.

For the SP part, trip distance was found to increase and footpath availability was found to decrease the use of “Bus1”, trip distance was found to increase the use of “Bus2” mode, Journey time and trip distance were found to increase and gender was found to decrease the use of “Metro1” and finally trip distance was found to increase the use of “Metro2”, walking being the base condition in all cases.

The study recommends planning interventions for management of motorized means in long distance trips, improvement of walking condition in short distance trips and development of gender friendly infrastructure for existing transportation modes. For the purposed alternative means of transport metro is seen as the best preferred option for the study group therefore further analysis of this mode is suggested.

Keywords: Educational mode choice, Discrete choice model, Mixed logit, Revealed preference, Stated preference.

ACKNOWLEDGEMENTS

I am grateful to my supervisor, Dr. Pradeep Kumar Shrestha for his valuable support, guidance, and motivation to accomplish this task. I would also like to express gratitude towards MSc Transportation Program Coordinator, Anil Marsani, and other respected faculty members for guiding from time to time for the study. I would like to thank my colleagues and seniors from Department of Roads and other offices of Government of Nepal for providing valuable suggestions on data modeling and analysis. I would also like to thank the faculty members of different engineering colleges for allowing me to collect data and lastly, I would like to thank the students who contributed their valuable time in filling the data forms.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

GON	Government Of Nepal
DOR	Department of Roads
CBS	Central Bureau of Statistics
CANN	Clean Air Network Nepal
MOEST	Ministry of Education Science and Technology
RP	Revealed preference
SP	Stated Preference
VMT	Vehicle Miles Travelled
NPC	National Planning Commission
MOPIT	Ministry of Physical Infrastructure and Transport
MaYA	*Manav Kendrit Yatayat Abhiyan*
JICA	Japan International Cooperation Agency
KSUTP	Kathmandu Sustainable Urban Transport Project
USA	United States of America
BRT	Bus Rapid Transit
AMTS	Ahmedabad Municipal Transport Service
MMRDA	Mumbai Metropolitan Region Development Authority
DMRC	Delhi Metro Rail Corporation(DMRC)
DTB	Delhi Tourism Board
WBTC	West Bengal Transport Corporation
CANN	Clean Air Network Nepal
Bus1	Air-conditioned BRT with higher comfort and fare, taking lesser travel time
Bus2	BRT similar in terms of comfort and fare, taking lesser travel time
Metro1	Metro having higher comfort (lesser crowding, seat availability, etc.) and cost, taking lesser travel time
Metro2	Metro Train similar in terms of and cost providing higher comfort and taking lesser travel time

CHAPTER ONE: INTRODUCTION

1.1 Background

Transportation infrastructure has been found to impact positively on economic development (Edington, 2006) this effect is even more crucial in urban areas as properly managed and designed transportation infrastructure ensures efficient use of urban space, provide efficient and affordable mobility and ensure access to workplace, school/colleges or social/recreational sites (MaYA, 2014). The degree of this impact however depends upon several factors the efficiency of the transportation system is one of them, which in turn relies on proper planning.

Concerning Kathmandu valley which the capital city and one the major economical centers of the country the existing transportation system does not seem to fully deliver the desired outcomes as its efficiency as well as an understanding of the current travel demands is not satisfactory as shown by many studies.

Among the three components of the transport system, public transportation consists of 4 components (Bus Tempo Micro and Taxi) which exists in almost all of the major routes of the city, however, it is suffering from many different problems including.

- Public Vehicles coming to the city center and hampering the traffic flow as the existing bus park's capacity cannot handle all incoming traffic (JICA, 2012).
- About 70% male and 85% female finding the vehicles overcrowded (World Bank, 2013)
- About 16% male and 34% female feeling insecure in the vehicles (World Bank, 2013)
- About 60% of Passengers not satisfied with the time the commute is taking (MAYA, 2014)

Not only the users of public transport system but also those using the private modes are facing problems like higher travel costs, congestions, safety issues, etc. yet are reluctant to shift to the public transportation due to longer travel time, longer waiting time, irregular operation and overcrowding (JICA, 2012)

Similar to the previous two modes of transport the walking conditions in the valley with a walkability index of 557 (CAAN, 2010) are not user friendly as well.

To overcome this suitable planning intervention are required, it includes an understanding of the current travel demands, development of transportation infrastructure and proper management of transportation facilities, by ensuring these things the desired impacts of an efficient system of transport can be ensured

The various trips that need to be analyzed for transportation planning are: work trip i.e. trips to job/workplace, educational i.e. trips to an institution of study and social/recreational trip i.e. trips for shopping, holidays, gathering, etc. For both social and recreational trips a traveler will have the option to choose the destination hence a trip distribution, as well as mode choice analysis, are significant. Work trips and educational trips in contrast require only the mode choice analysis as a choice of work or study location is usually long term decision thus beyond the range of model forecast (Mannering). A proper analysis of these trips and their efficient incorporation into the overall transportation planning is very important for the development of an efficient system

Among the various trips to be analyzed educational trips have great importance in terms of current share in travel demand as well as their future implication, from the example in USA school trips were estimated to cost around $20 billion yearly, occupied 1% of VMT, contributed to 1.5% of the pollution and caused the problem of childhood obesity and to reduce obesity of children the White House task force for child obesity had to be formed which had taken the goal to increase the walking percent of students by 50% in 5 years (Mc Donald, 2009). Similarly, for colleges going students in USA mode choice was found to create a perception and thus influence how a student behaves for the mode choice in future e.g. a student who drives to college now is likely to buy his/her car when they start to earn (Carlos JL Balsas 2003). In China, it was also found that maximum car travel can decrease children's motor functionality and cognitive capacity (Feng and Linn 2017). These studies and many more have concluded that to make children physically fit (McDonald, 2009), to increase/enhance cognitive capacity, to produce adults with a better view towards travel mode (Carlos JL Balsas 2003) and to enhance their safety one needs to know which factors affect the choice between the various modes of travel to education institutions.

In Nepal also the education sector is quite large and comprises of school education wing with 35,601 schools, 1,47,538 teachers, and 73,91,524 students providing children with high Scholl level education. The higher education system on the other hand consists of 15 universities and 1407 colleges and are serving 3,61,077 enrolled students. Having such a huge population the educational trips form a large part of the total travel demand e.g. in Kathmandu they occupy 20% of total trips (JICA, 2012), since both the school-going population and the average years of enrollment are increasing these trips not only occupying a major share presently but are also bound to do so in the future (CBS, MOEST). Among the various fields of higher education the graduate level of engineering compromises of 18,764 Students, 12,981 in province 3 (MOEST) among which around 7000 are in Kathmandu.

The current study analyzes these trips concerning the modal split for the trips made by above-stated graduate-level engineering students of Kathmandu valley and develops a mode choice model using the mixed logit with a revealed preference part for analyzing the existing transportation facilities and stated preference part to understand the choice of the users towards the hypothetical alternative choices.

1.2 Problem Statement

The policy framework of Nepal targets to provide a safe and managed public transportation system to ensure a comfortable, economical and reliable journey for the masses (NPC, 2020) for this to be attained the transport infrastructure and facilities shall be in line with the user demands. The reality however as explained above is quite different for Kathmandu valley. To overcome this proper planning interventions are required, which in turn relies on proper travel demand analysis of the users, this can be achieved by modeling the travel behavior of individuals.

About 19% of the total trips in the Kathmandu valley fall under the educational trip category (JICA, 2012). Since these trips constitute a major part of the generated trips the users of these commutes are subjected to the many problems of the current transportation system, and in turn contribute to congestion, pollution, increase in public transportation use (the level of serviceability of which is not satisfactory) and tends to increase accidents e.g. in the year 2075 about 138 lost lives in a bike accident in

Kathmandu valley (The Himalayan Times, 2019) among whom many are college-going youths who still have much to contribute to the society.

The above-stated phenomena can also be seen in the trips of graduate-level engineering students e.g. many are preferring walking over public transportation in distances as large as 4 Km, the average speeds of bike commuters are as high as 65Kmph (spot speeds will even be higher for such high averages).

These trips have not been analyzed until now, thus despite the proportion being low a detailed study into them can provide new insight into the travel behavior of the selected population. The findings can also be used as a preliminary trial model for models of other educational mode choices thus assisting in the transport planning process.

1.3 Research Questions

The questions that are expected to be answered from this research are:

1) What is the current mode share of engineering students of Kathmandu Valley and what is the expected mode share if the currents trends continue and no new alternative is provided?

2) What type of alternative is preferred by the users if metro rail or (and) BRT are to be introduced and what will be the mode share under that scenario?

1.4 Research Objective

The main objective of this study was to analyze the mode choice behavior of graduate-level engineering students of Kathmandu valley. The specific objectives are

- To study the current mode share and predict modal share if present trends persist
- To study modal shifts and trends under hypothetical modes of transportation

1.5 Scope of Study

This study focuses on the educational trips of Graduate level engineering students of Kathmandu Valley. The study provides detailed insights into the current mode choice and the factors making it so as well as the preferred alternative transport for the group under study and well as shade some light on the nature of factors affecting the overall educational trips.

1.6 Limitation

This study is done in a resource-constrained scenario. Hence limitation exists in the size of the sample and design of the questionnaire. The population of the study group is small therefore despite the study giving a detailed perspective into the travel behavior of the study group and also some hints regarding the nature of educational trips further studies are needed before concluding the overall pattern for educational trips. Lastly, with regards to alternative modes, this study gives the general liking of masses and affecting factors for alternative modes of transportation but cannot be solely taken for implementation as liking is only one of the many factors for implementing a new transport mode. Other factors like space availability, demand threshold, linkup availability, resources, and technical feasibility, etc. also play a crucial role, therefore the conclusion while can act as a milestone for alternative transportation planning proper study must be conducted before implementation.

CHAPTER TWO: LITERATURE REVIEW

2.1 Four-Step Transport model

Demand analysis is one of the most important phases of the urban transportation planning process. Proper estimates of the current and the future, the magnitude of travel frequencies, and the mode used for them are crucial for proper transport planning.

Travel demand depends upon various factors whose effects are nonlinear and growth is dynamic thus models have to be developed for accurate prediction. Among the different models used for travel demand analysis "four step transportation model" is the most frequently used method, it was developed in the early 1950s. It uses methodologies to linkup land use and user behavior to travel demand. The order of steps in the four step transportation model is: trip generation, trip distribution, modal split, and traffic assignment. These steps can be described in the table below

Table 2.1: Four step transportation model

Activity	Output
• Trip generation	• Origin of trips • Number of trips • Departure time
• Trip distribution	• Destination of the trips
• Mode choice	• Mode selected for various trips
• Traffic assignment	• Traffic in each mode of each route

Modal split which is the third step helps us to determine which mode of commute will be used by an individual and what factors affect their choice of a particular mode in general.

2.2 Utility-based choice theory

Utility-based choice theory is a theoretical basis for modal split. Utility is the measure of the attractiveness of alternatives and is a function of socio-economic, trip, and infrastructure characteristics. Utility choice theory states that “an individual will select the alternative from their set of choices for which their utility is maximized.”

2.3 Choice Models

The choice of a particular mode of transport depends upon the utility of that mode to an individual which in turn depends upon socioeconomic as well as trip characters of the individual. Individual utility has two parts the systematic part and the random part. The mathematical functions used to determine the utility can estimate the systematic part and only a portion of the random part, the degree of incorporation of the random component depends upon the model used, which will be discussed later in this chapter. The various mathematical models used to relate utility and parameters upon which the utility (or the mode choice) depends are known as choice models, they are of following types

Aggregate choice model: These models are developed based on the share of individuals choosing each alternative and socio-demographic attributes of the group

Discrete choice model: Discrete choice models are based on observed choices made by individuals from a finite set of alternatives. Model is developed based on the individual choice and socio-economic and trip characters of the individual

Based on the nature of formulation and the parameters used discrete choice models are found to be better performing as compared to aggregate choice models. Discrete choice models are found to have better transferability based upon transferability test statistics, transfer index, and goodness-of-fit of fit based comparative studies models developed for two cities with similar characteristics of Iran (Baghestani et. al. , 2014), greater predictability based on models developed for train commuters in Seoul, Korea (Chasung et. al. , 2009) and better sensitivity based on a study of CO_2 emission analysis in USA (Mclean et. al. , 1994)

2.4 Surveying

As stated above the utility model requires socio-economic as well as the trip-related information for their formulation as well as validation. The process of collecting these data is called surveying. Surveying can be done through many different methods like questionnaires, home interviews, direct observation, etc. The method used most popularly for the mode choice modeling is questionnaire survey. The questionnaire contains socio-economic and trip-related information the trip-related part can contain one or both of the following parts

RP (Revealed preference) survey part: This part helps to understand the user's current travel demand and their behavior towards the existing modes of transport. In the RP part of the survey, the users make the choice based on their knowledge of the nature of the mode of commute and are only provided with choices of existing modes of transport.

SP (Stated Preference) survey part: This part helps to understand the user behavior under hypothetical scenarios of modes of transport, it is used to access the user's behavior towards the possible transport interventions, thus acting as an input to the alternative transportation planning process. In SP part the user is provided with the details of the alternative and based upon which they make the choice

2.5 Regression

Regression is a statistical measurement used in finance, investing, and other disciplines that attempt to determine the relationship between one dependent variable and a series of other changing independent variables by fitting the data into a suitable model. The various parameters of choice models/ equations of modal split can be evaluated using regression it can be

Binomial Regression: Where there are two choices e.g. Head/Tail

Multinomial Regression: Where there are more than two choices e.g. mode choice between bus train and flight

Similarly, regressors are parameters of the regression equation and influence the various steps of the four-step transportation planning process they are of two type

Alternative invariant: These characteristics vary only as per the individual and not as per the choice e.g. age, income etc.

Alternative Variant: These characteristics vary over the individuals as well as the choice e.g. price paid for orange juice and mango juice is different for the same person, also the price paid for same juice type is different for the same person if they may enter different shops

2.6 Logit Model

These are discrete choice models that are used to form the systematic part of the utility function and also to determine the probability of choosing an alternative, which can then be used to predict the outcome by observing the alternative with the highest probability.

Many models are available to do so e.g. linear model, Probit model, logit model etc. Logit model which assumes the CDF of the probability is logistic is one of the most commonly used. There are various forms of logit model they are:

Logit: Used for alternative invariant i.e. regressors which vary only between individuals and not between the alternatives its formulation is

$$U_{ij} = V_{ij} + e_{ij} \tag{2.1}$$

$$V_{ij} = \beta_j X_i \tag{2.2}$$

Where

U_{ij} = Utility of the individual i for the j^{th} alternative

V_{ij} = Systematic part of the utility U_{ij}

e_{ij} = Random part of the utility U_{ij}

β_j = Coefficient for the j^{th} alternative

X_i = Alternative invariant character

Conditional Logit: Used for alternative variant and invariant regressors its model formulation is

$$U_{ij} = V_{ij} + e_{ij} \quad (2.3)$$

$$V_{ij} = \beta X_{ij} + \beta_j X_i \quad (2.4)$$

Where

U_{ij} = Utility of the individual i for the j^{th} alternative

V_{ij} = Systematic part of the utility U_{ij}

e_{ij} = Random part of the utility U_{ij}

β = Coefficient for alternative variant regressors

X_{ij} = Alternative variant regressors e.g. Price

β_j = Coefficient for alternative invariant regressors

X_i = Alternative invariant regressors e.g. income, age

Mixed Logit: Similar to conditional logit but the coefficient of the specific variant is modeled as a random parameter. The formulation of this model is

$$U_{ij} = V_{ij} + e_{ij} \quad (2.5)$$

$$V_{ij} = \beta_i X_{ij} + \beta_j X_i \quad (2.6)$$

U_{ij} = Utility of the individual i for the j^{th} alternative

V_{ij} = Systematic part of the utility U_{ij}

e_{ij} = Random part of the utility U_{ij}

β_i = Coefficient for alternative variant regressors modeled as a random parameter with mean β and standard deviation σ_β

X_{ij} = Alternative variant regressors e.g. Price

β_j = Coefficient for alternative invariant regressors

X_i = Alternative invariant regressors e.g. income, age

The probability that an individual chooses a mode i equal the probability that the conditional utility of mode i is the largest of the conditional utility of all other modes.

$$P_j = P\,(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik}) \quad j, k = 1, 2, 3, \ldots J\ , i \neq j \quad (2.7)$$

Since the systematic part of the utility function only is know the mode choice is predicted based on the probability of the particular mode which is given by the CDF of the logit model as follows

$$\boldsymbol{P_{ij}} = \frac{\boldsymbol{e^{V_{ij}}}}{\sum \boldsymbol{e^{V_{ij}}}} \tag{2.8}$$

2.7 Multinomial logit model

Multinomial logit model is one of the most widely used discrete choice models, it has been used in many fields e.g. to find out the effects of uncertainty on marital status in social sciences (Blossfield et. al. 2005), to relate the physical activity level to weight and obesity (Toriola, 2018.), to model the modal choice behavior of commuters and many more. The applications in mode choice modeling are further discussed here

2.7.1 Educational Trips

Many studies have been conducted to describe the educational mode choice by using the multinomial (MNL) logit model

Scheiner et al. (2019) stated that the student mode choice depends on factors like

- Trip Character: Length of trip, weather of commute
- Child character: Age, gender
- Household character: Parental resources, number of siblings, parental attitudes (Education, fear level, gender attitude)
- Transport environment: Traffic, footpath availability, number to times to cross
- Built-up character: Urbanization
- Social character: Attitude of society (behavior towards walking and car ownership), the trust of the individual to the locality, availability of children of the same age which parents can trust

They analyzed primary school mode choice in Germany and out of all different factors they have described they have found that

- With an increase in distance motorized travel is preferred
- German-born parents prefer walking more

- If the father is employed then walking chances are more if mother employed the n motorized travel chance is high
- Among the walking and cycling groups attitude shifts from walking to cycling with an increase in age [for same age boys prefer cycling more]

Zhang et al. (2012) pointed out that mode choices of students in china are different than western countries because

- One can give birth to one child only, therefore, he/she is given too much priority
- Suburban areas have too much population while the schools are in urban areas
- To be admitted in school system one must have a house and vehicle

Based on their research they concluded that

- Car ownership increases chances of motorized travel
- Rush hour departure increases chances of motorized travel
- Increased distance increases chances of motorized travel
- Working parents increases chances of motorized travel

Lu Ma et al. (2019) studied the effect of weather condition on the commute of students and concluded that along with other factors: sky condition, temperature, air quality and humidity effects the mode choice of students

Singh and Vashudevan (2018) analyzed students mode choice in the Indian city of Kanpur and found that

- The walking proportion was too low which can be contributed to negative walking conditions
- Lower-income families prefer bus travel
- In high-income families having cars, boys are more likely to be driven while girls prefer to take rickshaw or auto with friends(*Sahelis*)

2.7.2 Work/Recreational trips

Rai (2012) investigated the mode choice of work trips in Kathmandu and concluded that the factors affecting the choice are waiting time, traveling time, total income, cost of commute, and age. Joshi (2019) investigated the mode choice for intercity travels among Kathmandu and different cities using a revealed preference (RP) survey of Car, Bus, and air travel and an SP survey with different types of trains. The RP part concluded that taking ‘Bus’ as the base category, the odds of using a car for travel are affected by the cost bearer, frequency, vehicle ownership, and family size and the odds of using air travel are affected by cost bearer and frequency only. The SP part concluded that the odds of the different types of railways are affected by travel time, travel cost, and distance with varying magnitude for each railway mode provided as compared to existing modes of transport. Ghimire (2019) investigated work trips of Kathmandu valley concerning three modes of travel Public transport, two-wheeler and four-wheeler and concluded that the odds of using 2 wheeler or 4 wheeler as compared to public transportation is affected by travel cost, travel time, gender, and monthly income.

2.8 Conditional/Mixed Logit

With regards to conditional logit, Many different trips were found to be analyzed, a typical example for educational trips is a study in Vietnam Duy Quy Nugen Phuuc et al. (2018) which applied the conditional logit model to analyze the current factors influencing the choice between dominating modes (walking, bus and motorbike) and also provided hypothetical conditions of efficient and reliable public transport. They concluded that age, gender, income and travel time have strong impact and also students are likely to switch to public transport if effective and reliable transport is developed

With regards to mixed logit, studies have not been done in the field of educational mode choice as the value of alternative variant regressors for all alternatives of each user is difficult to evaluate for each individual. Unlike other fields like medical where these parameters are almost impossible to evaluate for each individual, some approximate methods are available to evaluate the parameters of modes rejected but these have not been used in educational mode choice

Despite not being used in educational mode choice many studies have been conducted using this model for other types of trips.

Bai. et. al. (2016) analyzed the effect of various factors on the mode choice of general consumers with the main task of identifying the effect of cost adjustment. They used Multinomial logit, conditional logit, and Nested logit and concluded that MNL is the best all the methods have the same accuracy (McFadden Rho being the same) and since all other logit models have unnecessary complications. The findings of this study are summarized in Table 2.2

Table 2.2: Factors affecting mode choice of commuters (Bai. et. al. 2016)

Mode	Statistically Significant Factors and Effects (Base category Bus)
Bus	Base category
Walk	Distance (Negative), Gender(Male Positive), Time(Negative) Household size (Positive)
Bike	Distance (Positive), Gender (Male positive), Time (Negative), Household size (Negative)
Car	Distance (Positive), Gender (Male positive), Time (Negative), Household size (Positive), Number of Kids (Positive)

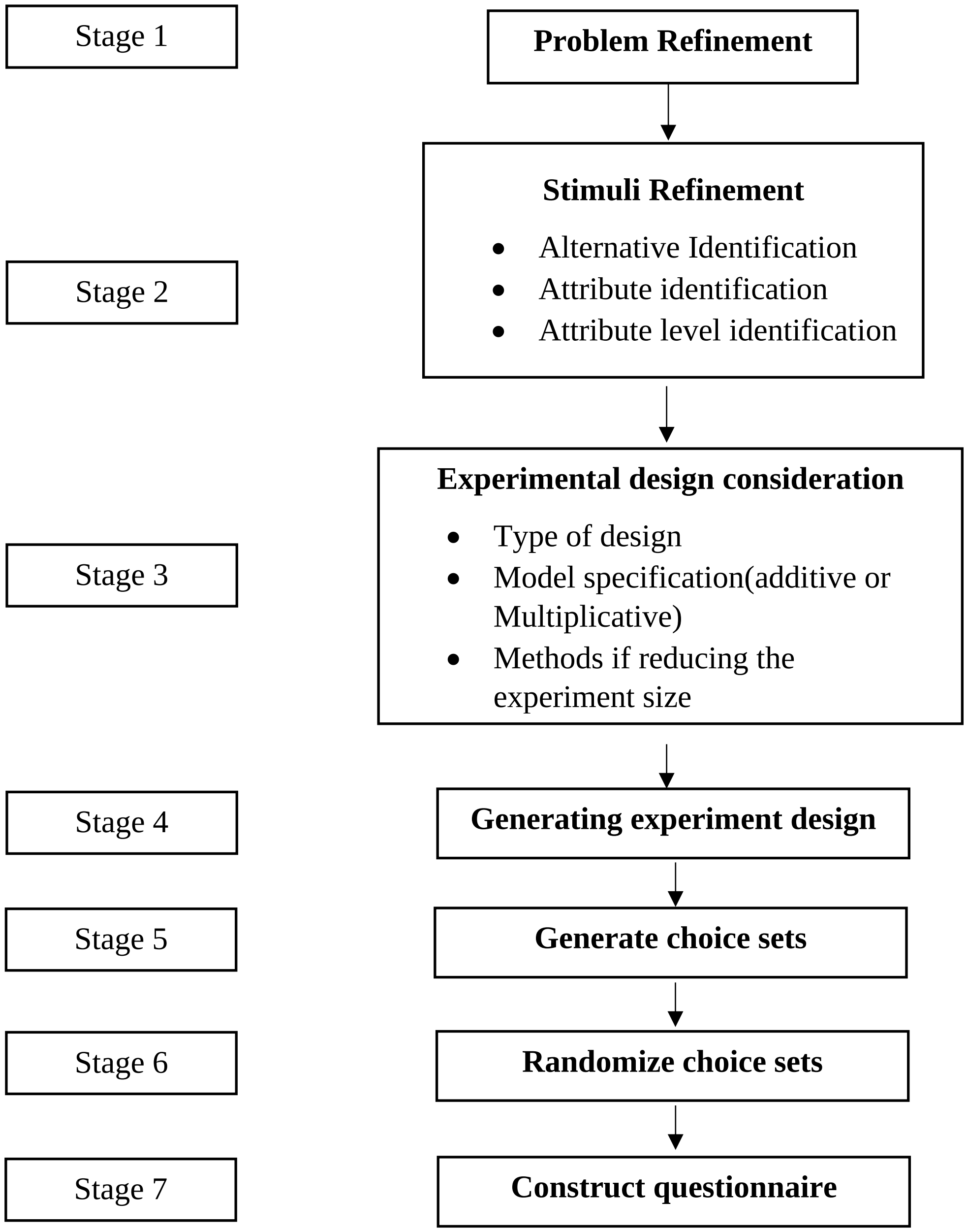
Kye-Lee et al. (2018) investigated the factors which will enhance or degrade the choice between the KTX (A new purposed high-speed rail) and the existing aircraft They selected popular variable like travel time, frequency and travel cost from the previous studies and appointed a team of 8 experts which used AHP to select additional variables of safety and Availability of duty-free shops to be taken in the study. They took KTX as the base category and their findings are summarized in Table 2.3

Table 2.3: Factors affecting mode choice of commuters (Kye-Lee et al. 2018)

Parameter	Business Passenger	Leisure Passenger
Fare	Negative	
Travel time	Negative	
Frequency	Positive(More than Leisure)	Positive
Safety	Positive(More than Leisure)	Positive
Availability of duty-free shop	Positive(More than Leisure)	Positive

2.9 SP experiment formation

Stated preference survey provides the commuters with planned choice sets, since these sets don't exist at presently due care shall be taken to ensure that they are properly designed. The process of formulation of the SP part survey (Hensher, 2005) is shown in figure 3.5.



(Source: Hensher, 2005)

Figure 2.1: The SP design process

The first two stages are associated with the analysis of the aspects of the SP problem. Problem refinement ensures that what the research hopes to evaluate and what are its objectives are understood, the second part ensures that the researcher becomes well aware of the alternatives, their attributes to be considered, and their levels. At the end of the second stage if the analyst is not satisfied the two steps can be repeated

If the analyst is satisfied the next step will be to design the experiment. An experiment is a process where the variation in the dependent variable is observed for the variation in the independent variable, since experimentation is a scientific process the variation of the independent variables shall also be in a scientific manner i.e. the manipulation of the independent variables occurs in a designed manner. The process of designing the manipulation of the independent variables is called "experimental design". There are various methods of experimental design like full factorial design, ratio estimate method, and magical choice probability method. Among these methods, the full factorial design is most preferred as it ensures the orthogonality and is simple to administer (Sanko, 2001). In this method, the number of combinations is given by

$$\text{Experiment Sets} = \Pi\ (\text{Levels})^{\text{Attributes}} \qquad (2.9)$$

Say if we have 2 attributes with 2 levels and 2 attributes with 3 levels then the number of sets will be 2^2 X 3^2 = 36 sets. After the experimental design, the researcher has to determine model type weather the model will be additive i.e. analyze the main effect separately or will be multiplicative i.e. analyze the interaction between the attributes. After defining the model the next step will be to reduce the number of experiment sets. There are many methods to do so, the fractional factorial design method assumes that the effect of 3rd or higher degree interactions is very low hence we can select only one of these interactions thereby reducing the size of the experiments. One of the main benefits of this method is that despite the orthogonality not being preserved in the interaction the orthogonality of the main effect is still preserved while the major drawback is that the trivial games i.e. the choice sets which will surely be rejected or the choice set which will surely be selected are retained this decreases the accuracy of prediction and also make the model nonrealistic as these sets never occur in reality e.g. cheap, fast and comfortable bus, similarly for a large number of attributes and their levels even after the fractional factorial design the number of sets is too high, as a result of which the choice set may occur in only a few questionnaires thereby giving rise to the problem of lack of sufficient data for modeling

Another approach is the removal of trivial games which studies the relation of different choices in the existing markets and thus reduces the number of experiment sets as well as provides the magnitude of the difference in the attributes. A major problem of this method is that the orthogonality of the experimental sets is lost. Orthogonality is used during the experimental design to avoid the collinearity between the attributes. However, the orthogonality does not need to be preserved during the model estimation stage. Hensher (1994) differentiated between the design data orthogonality and estimate data orthogonality and pointed out that non-retention of orthogonality during estimation does not drastically affect the logit model.

After finalizing the experimental design, it is then executed to obtain all the test sets. Since even after all the reduction strategies used in many cases the experiment sets are far too many to be included in a single set therefore the choice sets are divided into subsets so that they can be included in the questionnaire. To overcome possible biases from order effects, the order of appearance of these choice sets are randomized across questionnaires.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Overview

An overview of methodology is shown in figure 3.1 below. It consists of the following parts.

3.1.1 Data collection:

This phase is associated with the formation of tools for the collection of data. The first step to do so was the review of the literature associated with the process of the formation of tools and studies of previously done in this field. After sufficient background data was known the various parts of the tools can be formed based on the nature of commuters and their commutes. Finally, data was collected by conducting a survey using the tool formed previously, the survey was be conducted personally.

3.1.2 Choice modeling

This phase is associated with the analysis of the collected data. Data was first converted into spreadsheets using computer programs, then divided into training data set i.e. data set used to calibrate the model and validation data set i.e. data set used to validate the model and then analysis was performed.

The first part of the analysis was the preliminary data analysis which investigates the nature of data based on the various attributes like age, gender, etc., it was followed by the model formation which involved calibration of the RP and SP models using the training data and examining the nature of fit of model using statistical parameters, finally, the model formed was validated for its accuracy, however, in practice many models were first formed and the one with better statistical parameters and accuracy were selected as final

3.1.3 Drawing conclusion

This phase is associated with the analysis of the model. The various coefficient of the models were assessed to find out their effects on the model choice. RP models were analyzed to evaluate the current mode choice, while SP models were analyzed to evaluate the mode choice under hypothetical conditions. Analysis was in terms of odds ratio and statistical significance. Finally based on the significant results conclusions were drawn regarding the mode choice of the current population and the planning implications of the findings. The Study was concluded by stating the limitations and future research recommendations.

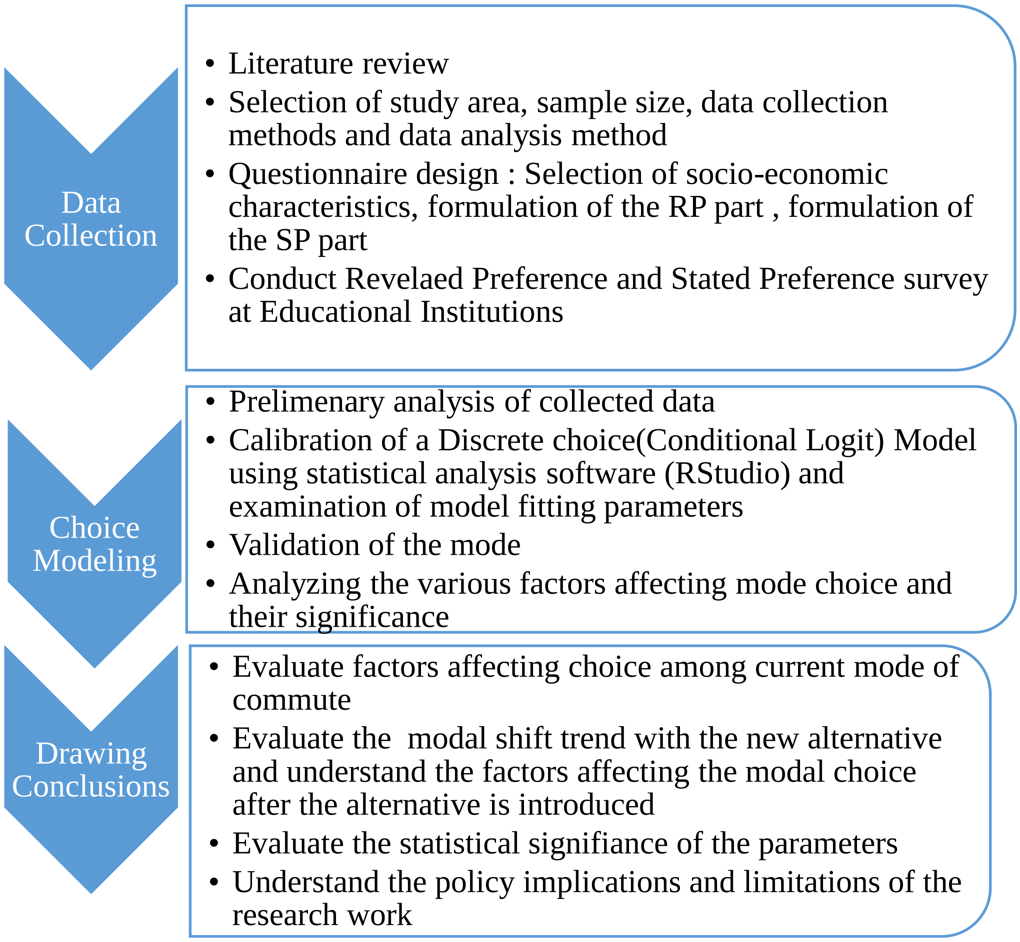


Figure 3.1: Framework of Methodology

3.2 Study Area

3.2.1 Kathmandu valley

The study is based on the educational mode choice of graduate-level engineering students of Kathmandu valley.

Socio-economically, Kathmandu valley is the largest urban center and is the largest contributor in the country's economy. The valley is inhabited by the ethnic community the *Newars* and the migrated people from all over the country.

Demographically the valley contains a population of permanent inhabitants of 11,95,285 persons(CBS, 2017) in 3 Rural municipalities, 16 municipalities, 1 Sub Metropolitan Municipality, and 1 Metropolitan Municipality. The total residents on the other hand are as high as 4 million persons but official data are not available on this front.

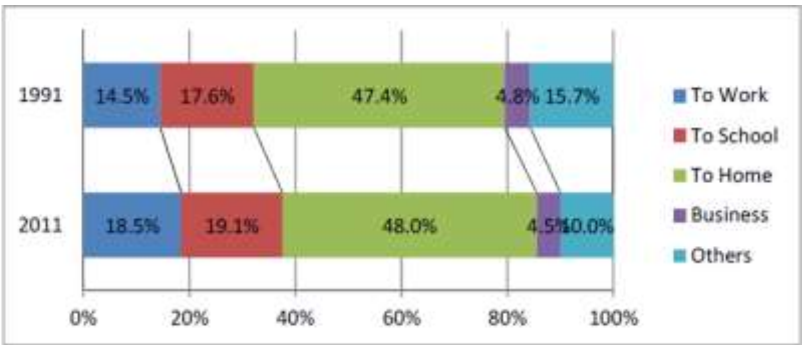
3.2.2 Travel demand and educational trips in Kathmandu valley

In a household survey of 18100 houses by JICA (2012) the trips were estimated to be 3,483,349 in number, the nature of these trips based on the purpose and comparison between 1991 and 2011 is shown in table 3.1 and figure 3.2 below. It can be seen that the maximum number of trips are for educational purposes followed by other types of trips.

Table 3.1: Trip volume based on purpose of Kathmandu

Trip Purpose	**Number of Trips**	**Percentage (%)**
To Work	634461	18.5
To School	657030	19.1
To Home	1649236	48
Business	153469	4.5
Others	344197	10
Total	3438393	100

(Source: JICA 2012)



(Source: JICA 2012)

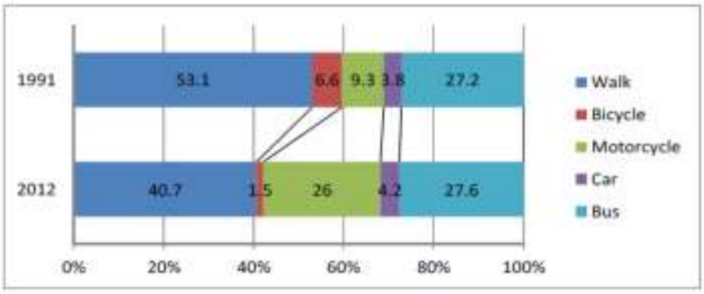
Figure 3.2: Comparison of trips by purpose between 1991 and 2011 for Kathmandu

The distribution of these trips based on mode used and comparison between 1991 and 2012 data is shown in table 3.2 and figure 3.3 below. It can be seen that currently the maximum share is of walking choice but the proportion of people doing so have decreased drastically from the past

Table 3.2: Trip volume based on Mode

Travel Mode	Number of Trips	Percentage
Walk	1398378	40.7
Bicycle	52445	1.5
Motorcycle	893126	26
Car	145980	4.2
Bus	948464	27.6
Total	3438393	100

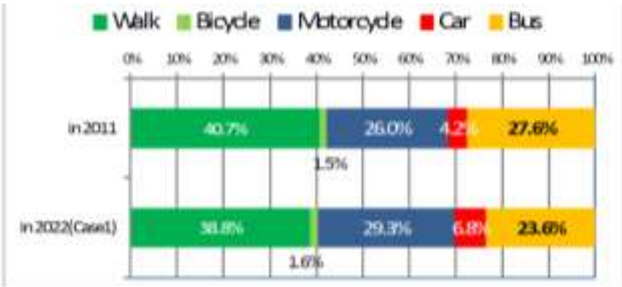
(Source: JICA, 2012)



(Source: JICA, 2012)

Figure 3.3: Comparison of trips by mode between 1991 and 2011

The study has also concluded if the current conditions prevail then there will be 1.59 times growth in the trips and overall vehicle ownership in Kathmandu valley and the mode choice of 2021 will be as shown in figure 3.4.



(Source: JICA, 2012)

Figure 3.4: Modal Split in 2011 vs 2022 for Kathmandu

The study concluded that if the demands grow as shown above without any intervening measures than more than 80% of the roads in the valley will be congested and have recommended an overall transportation plan as a solution.

3.3 Questionnaire design

3.3.1 Selection of socio-economic characters

Based on the study of previous works and the socio-cultural characters of the country the following socio-economical characters are considered for the study: Gender, age group, family income, family education level, size of family, vehicle ownership, and number of siblings.

3.3.2 Formulation of RP part

This part obtains the data regarding the currently available modes of transport. The RP part comprises of the journey (or trip) attributes and mode attributes

Journey attributes: It comprises of data like origin and destination of the trip, availability of companion for the trip, availability of footpath, and distance.

Mode attributes: It comprises of data like mode selected, travel cost and travel time.

3.3.3 Formation of SP Part

The SP part was formed using the procedure by (Hensher, 2005). A study MOPIT (2018) conducted a study to compare the various possible mass transit modes and identify the route for their operation, the study performed a multicriteria analysis of different modes of transport using the analytical hierarchy process and ranked rail 1st and BRT 2nd based on 6 criteria and 22 sub-criteria. Also, an analysis of the metropolises of neighboring countries showed that BRT and Metro rail were the dominating modes of mass transit, therefore the study aimed to consider these two as the alternatives to be examined. Three properties of the mass transit were taken into consideration time, comfort, and cost. While time was assumed to be faster for all the alternatives the other two attributes found to have two levels. For comfort same and comfortable for BRT and Comfortable and Much comfortable for metro rail, similarly for cost same and more for each alternative which gives $2^2 = 4$ sets for each alternative which are shown below

Table 3.3: Full factorial SP experiment design

Set	Cost	Comfort
BRT		
1	Same	Same
2	Same	More
3	More	Same
4	More	More
Metro Rail		
5	Same	Comfortable
6	Same	Much Comfortable
7	More	Comfortable
8	More	Much Comfortable

To understand the relationship between the conventional public transport and the BRT they were compared for the city of Ahmedabad in India where both are operational. Similarly, to quantify the levels of above full factorial design as well as to understand the trivial and constrained games the nature of BRT, Luxury BRT and Metro trains of 3 metropolis of India, Delhi, Mumbai and Kolkata were analyzed the results are presented below.

Ahmedabad: Among the many modes of transport in the city, two public transport modes the BRT (*Janmarg* BRT) and the municipal bus which is almost the same as the public buses currently in operation in the Kathmandu valley were studied for their cost the result obtained is shown in table 3.4 below.

Table 3.4: Comparison of fare prices of BRT and ordinary bus in Ahmedabad

Distance	Price of BRT (₹)	Price of Public Bus (₹ 3 per 2 Km maximum ₹ 25)
Up to 10 Km	10	15
10 – 12 Km	18	18
12-15	25	23
More than 15 Km	27	25

(Source: Ahmedabad BRT, AMTS)

Mumbai: Among the many modes of transport in the city BRT, air-conditioned BRT and Mumbai metro were compared for their prices the result obtained is shown in table 3.5 below.

Table 3.5: Comparison of fare prices of BRT Metro and ordinary bus in Mumbai

Distance	Price of BRT (₹)	Price of air-conditioned BRT (₹)	Price of Mumbai Metro (₹)
Upto 5 Km	5	7	10
5 – 10 Km	10	13	15
10-15 Km	15	19	30
More than 15 Km	20	25	40

(Source: Reliance Mumbai metro, MMRDA, Times of India)

Metro Occupancy:

Total commutes in a day: 380,000 passenger

Per train capacity = 1500

Total Trips = 370 per day

Peak hour = 9 am to 11 am & 4 pm to 6 pm = 4 Hours

Peak hour frequency = 3.5 minute per train

Peak Hour Trips = 240/3.5 ~ 69 trips

Total peak hour capcity = 1500*69 = 103,000 passenger

Maximum peak Hour flow = 50,000 * 2 = 100,000 passenger

Off peak hour commute = 280,000

Off peak Hour trains ~ 301

Off peak hour average occupancy = 280,000/301 ~ 930 passengers

Off peak saturation = 930/1500 = 0.62

Delhi:: Among the many modes of transport in the city BRT, air-conditioned BRT and Delhi metro were compared for their prices the result obtained is follows

Table 3.6: Comparison of fare prices of BRT Metro and ordinary bus in Delhi

Distance	Price of BRT (₹)	Distance	Price of air-conditioned BRT (₹)	Distance	Delhi Metro fare (₹)
Upto 4 Km	5	Upto 4	10	Upto 2 Km	10
4 – 10 Km	10	4 – 10	15	2 – 5 Km	15
More than 10	15	10- 12	20	5 -12 Km	30
		More than 12	25	12 – 21 Km	40

(Source: DMRC, DTB , Economic times, Business Today)

Metro Occupancy:

Total commutes in a day: 2,800,000 passenger

Capacity = 4,050,000

Total Trips = 2700 per day

Peak hour = 9 am to 11 am & 4 pm to 6 pm = 4 Hours

Peak Hour Trips = (26 trains per hour * 2 lines + 23 trains per hour * 4 lines) * 4 Hours

= 576 trips

Total peak hour capacity = (2 lines *56260 passenger per hour per line + 4 line *49300 passenger per hour per line)*4 hours

= 1,238,880 passenger

Maximum peak Hour flow = (2 lines *50186 passenger per hour per line + 4 line *42524 passenger per hour per line)*4 hours

= 1,081,872 passenger

Off peak hour commute ~ 1,718,100

Off peak Hour trains ~ 2124

Off-peak hour average occupancy = 1718100/2124 ~ 810 passengers

Kolkata: Among the many modes of transport in the city BRT, air-conditioned BRT and Kolkata metro were compared for their prices the result obtained is shown in figure 3.7 below

Table 3.7: Comparison of fare prices of BRT Metro and ordinary bus in Kolkata

Distance	Price of Normal Bus (₹)	Price of air-conditioned Bus (₹ 5 per 2 Km)	Distance	Kolkata Metro fare (₹)
Upto 4 Km	7	10	Upto 5 Km	5
4 – 12 Km	9	10-30	5 – 10 Km	10
12 – 16 Km	11	30 – 40	10 -15 Km	15
16 -20	12	40 – 50	15 – 20 Km	20

(Source: WBTC, Kolkata metro, Business-Standards, Times of India)

Metro Occupancy:

Total commutes in a day: 700,000 passenger

Per train Capacity = 2700 passengers

Total Trips = 284 per day

Peak hour = 9 am to 12 am & 4 pm to 7 pm = 6 Hours

Peak hour frequency = 5 min per train

Peak Hour Trips = 360/5 = 72 trips

Total peak hour capcity = 2700*72 = 194,400 passenger

Maximum peak Hour flow = 175000 (peak hour saturation = 0.9)

Off-peak hour commute = 505600

Off-peak Hour trains ~ 212

Off-peak hour average occupancy = 505600/212 ~ 2385 passengers

Off peak saturation = 2385/2700 = 0.88

It was be seen from the scenario of Ahmedabad that the normal bus services and the BRT have almost same prices, also since BRT is a replacement for the normal bus in a given route same prices made sense, therefore, the study assumed the price of the BRT to be same as that of the Current public transport

The comparison of normal and AC BRT for Delhi showed a variation in the price of magnitude around 2, similarly, the prices in Mumbai were also found to be varying by a magnitude of 2 times. Despite Kolkata not having BRT service the price comparison of normal and AC bus provided a variation of about 1.5 to 3 times. Therefore a variation factor of 2 for India was concluded to be logical since the prices of normal and AC Public bus in India was found to vary by 2.6 times only while it was found to vary by around 3.4 times in Nepal not considering any extra factors like higher tax, higher living costs of country, the higher maintenance cost of the vehicle the variation of 2.5 times between the Normal BRT and AC BRT was established

For accessing the metro rail scenario metro rail was directly compared to normal bus in Kolkata while in Delhi and Mumbai BRT was taken as contemporary to the Normal bus as stated above. It was observed that for Delhi and Mumbai the metro cost are quite high with a magnitude of variation ranging from 1.5 to 2 times and that metro was running in an off-peak hour at low saturation, thus a conservative value of 1.5 times for even more comfortable metro rides was established as the current policy framework has adopted the prioritization of metro rail in Kathmandu. It was observed for Kolkata that the price of metro and the normal bus were almost the same while the train runs at almost saturation in the off-peak hours also, therefore a scenario of less comfortable and cheap metro i.e. having price same as that of the normal bus and more comfortable but costly one i.e. having price 1.5 times that of the normal bus was concluded to be feasible. On analyzing the full factorial design scenario 1, 4, 5, and 8 are selected for the SP design.

- The first option was a luxurious BRT which is more comfortable (in terms of seats, spacing, air condition availability etc.) and costlier by 2.5 times as compared to current public transportation, which was abbreviated as “Bus1”
- The second option was a BRT whose comfortability and cost are the same as compared to current public transportation, which was abbreviated as “Bus2”
- The third option is a Metro rail which was more comfortable (in terms of acceleration characteristics, crowding, etc.) and costlier by 1.5 times as compared to/ current public transportation, which was abbreviated as “Metro1”. And,
- The Fourth option was a Metro rail which is comfortable (in terms of acceleration characteristics but with vessels saturated) but has the same cost as compared to current public transportation, which was abbreviated as “Metro2”

Since only two scenarios are selected for each alternative mode, no choice set or randomization is necessary. A sample of the questionnaire is attached in annex 1

3.4 Sample size

Sample size for the study was determined as follows

(i) Based on variance (Kish, 1965).

$$\boldsymbol{n} = \frac{\bar{\boldsymbol{n}}}{\boldsymbol{1} + \frac{\bar{\boldsymbol{n}}}{\boldsymbol{N}}} \quad (3.1)$$

Where,

N is the total population

n is sample size from finite population

$\bar{n}$ is sample size from an infinite population

Sample size for an infinite population is calculated using:

$$\bar{\boldsymbol{n}} = \frac{\boldsymbol{S}^2}{\boldsymbol{se}(\bar{\boldsymbol{x}})^2} \quad (3.2)$$

Where,

S^2 = variance of population

$se(\bar{x})$ is the standard error of sampling population

For a 95% confidence level, the sample size is calculated as:

$$\boldsymbol{se}(\bar{\boldsymbol{x}})^2 = \frac{\boldsymbol{0.1\mu}}{\boldsymbol{1.96}} = \boldsymbol{0.051\mu} \quad (3.3)$$

$$\bar{n} = \frac{S^2}{(0.051\mu\,)^2} = 384cv^2 \quad (3.4)$$

Taking the coefficient of variation as 1 and $\frac{\bar{n}}{N}$ being very small, the minimum sample size was **384**.

(ii) Based on the population proportions the sample size can be calculated as (Cochran)

$$\boldsymbol{se(\bar{x})^2} = \frac{\mathbf{0.1\mu}}{\mathbf{1.96}} = \mathbf{0.051\mu} \qquad (3.5)$$

$$\bar{n} = \frac{S^2}{(0.051\mu)^2} = 384cv^2 \qquad (3.6)$$

Where

S = Sample size for infinite population

S'= Sample size for finite population

p = Population proportion taken as 0.5 for largest possible sample size

E = Margin of error

N = Population size

Taking E = 5% = 0.05, $Z_{\alpha/2}$ = 1.96 for 95% confidence of interval we get, S = 384.16

As sample size increase with increase in population let us take a conservative maximum value of Population size as N = 15000 then, S' = 374.43 ~ 375 samples

(iii) Based on empirical formula

(Green, 1991) Suggested a rule of thumb for determining minimum sample size for multiple correlations is 50+8*m* where, *m* is the number of factors. For the model to be developed in the study, 9 factors are taken at maximum. Hence minimum sample size is 122.

On analysis of all the three cases, a minimum sample size of 384 was adopted for model formulation and half of it was estimated to be required for model validation which made the least number of samples to be collected to be 576.

3.5 Data Collection and processing

Questionnaires were distributed by visiting colleges. The questionnaire contains data for socio-economic characteristics like sex, age, family income, family education status, family size, sibling number, and trip characteristics like Mode selected, route length (same for all modes) and cost/journey time for the mode selected as well as those not selected e.g. a person using 2 wheeler personal mode for educational trip provided with the distance (which is same if they choose any mode) and the cost and journey time for all the modes like personal. Some guidance was required for persons traveling with public transport for what will be the cost of travel by personal modes. The SP part of the questionnaire consists of 4 options as explained in the SP part formulation.

Students were requested to read the SP part clearly and then allowed to make queries if they were confused about the nature of the alternative choices.

The collected data was then separated into calibration data set and validation data set randomly from the whole data set.

3.6 Preliminary Data analysis

Preliminary Analysis of the data was done using Microsoft excel. Although package mlogit of R-Studio has better performance in handling many data types and extensions of logit model, results obtained at this stage were obtained from excel package only as mlogit package needs restructuring of data into "wide" format where each choice has rows equal to the number of alternatives and an additional column indicating the choice. The model development used logit package at the end of preliminary analysis.

3.7 Model formulation

Random utility theory is the theoretical foundation of these models. According to this theory the utility of any mode is expressed as:

$$U_{ij} = V_{ij} + \epsilon_{ij} \qquad (3.7)$$

For conditional logit, the systematic part of the utility function is given as

$$V_{ij} = \beta_j * W_i + \beta * W_{ij} \qquad (3.8)$$

Where

U_{ij} = Utility of i^{th} individual for J^{th} alternative

V_{ij} = systematic part of the utility function

ϵ_{ij} = random part of the utility function

β_j = Regression coefficient for alternative invariant regressors that varies as per the options but not per the individual

W_i = Alternative invariant regressors that do not vary with alternative e.g. age, sex, income, etc.

β = Regression coefficient for alternative variant regressors that do not vary with the options or individual

W_{ij} = Alternative variant regressors whose value varies with alternative as well as individual e.g. price per km, fever etc.

The probability that an individual chooses a mode i equal the probability that the conditional indirect utility of mode i is the largest of the conditional utility of all other modes.

$$P_j = P\ (V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik}) \quad j, k = 1, 2, 3, \ldots J\ , i \neq j \tag{3.9}$$

The CDF for probability is given by

$$\boldsymbol{P_{ij} = \frac{e^{V_{ij}}}{\sum e^{V_{ij}}}} \tag{3.10}$$

3.8 Model Estimation

It involves the process of determining the coefficients of the model. If the model to be estimated has J alternatives it will have (j-1) regression equation for (j-1) alternatives with one alternative as the base criterion. The coefficients and intercept terms of each equation were evaluated by the maximum likelihood method

In the maximum likelihood method, the coefficients are evaluated such that the value of the likelihood function is maximum i.e. the value of parameters are set such that the observed sample is most likely to occur. This is the most common method of evaluating the parameters of the logit model and has two steps, developing the likelihood function and estimating the value of the parameters that maximize the likelihood function

3.9 Modal Estimation Parameters

The strength of the model was determined by analyzing various statistical parameters and tests. They are as follows.

3.9.1 Z test and p-value test

These two tests were used to check the significance or the strength of the coefficients of the logit model. Both Z value test and the p-value test are used for the following hypothesis testing

Null Hypothesis: The coefficient for the given independent variable is zero i.e. there is no effect of the given parameter

Alternative hypothesis: The coefficient is not zero i.e. there is an effect of the parameter on the utility

Z Value test: Calculated Z value ($Z_{calculated}$) is the ratio of the calculated mean (or parameter) and the standard error i.e. the standard deviation of the population parameter. It provides us with the number of standard deviation times the calculated mean (of parameter) deviates from the assumed mean in the null hypothesis. Let α is the significance level i.e. the probability of rejecting a true null hypothesis or probability of false-negative then $Z_{critical}$ is the value which covers 1-α % of sample then if $Z_{calculated}$ is less than $Z_{critical}$ the calculated mean lies in the acceptance zone i.e. there is sufficient data to accept the null hypothesis of the parameter is equal to zero if $Z_{calculated}$ is more than Zcritical the calculated mean lies in the rejection zone i.e. there is not sufficient data to accept the null hypothesis and therefore the alternative hypothesis of the estimated parameter being non zero is valid.

P-value test: p-value is the probability of obtaining a z value as high as the calculated value if the mean assumed in the null hypothesis is true. Let α is the level of significance, if the p-value is larger than α then the probability of getting the z value is more than α, therefore, the null hypothesis is accepted i.e. there is sufficient data to prove that the parameter is zero, similarly, if the p-value is smaller than α then the probability of obtaining the calculated Z value under the mean assumed in the null hypothesis is smaller than α which means that there is not sufficient data to accept the null hypothesis and therefore the alternative hypothesis of the estimated parameter being non zero is valid.

3.9.2 Likelihood ratio test

Likelihood is a statistical function that measures the goodness of fit of the statistical model developed from the sample data, it is made from the CDF of the probability function and its peak provided it exists gives the values of the parameters of the function for which it has been formed. This is the principle behind the likelihood maximization method of parameter estimation which estimates the parameters such that the log-likelihood is maximum.

The likelihood ratio test determines the strength of the model, it compares the likelihood of the restricted model i.e. the model with no parameters or the intercept only model and the unrestricted model i.e. the model with all the parameters considered. It checks whether the difference between the two likelihoods is significant or is due to sampling error only and is the same. The hypothesis tested are

Null Hypothesis: The unrestricted and the restricted models are the same.

Alternative Hypothesis: The unrestricted model is better than the restricted model.

The likelihood ratio is given by

Likelihood ratio (LR) = 2(Log-likelihood of unrestricted model – Log-likelihood of restricted model)

The likelihood ratio thus computed is found to follow chi-square distribution with degree of freedom equal to the number of parameters used in the unrestricted model minus one, provided that the null hypothesis is true. Let α is the level of significance and $\chi^2_{critical}$ is the critical value of chi-square i.e. the value of χ^2 the probability of getting any values higher the which is less 1-α for given degrees of freedom then, if the calculated value of chi-square or the likelihood ratio is less than the critical value of chi-square then the null hypothesis i.e. the restricted and the unrestricted models are the same is accepted since the probability of occurrence of the computed χ^2 value is more than the level of significance, similarly if the calculated value of chi-square is more than the critical value of chi-square then the null hypothesis i.e. the restricted and the unrestricted models are the same is rejected and the alternative hypothesis is accepted since the probability of occurrence of the computed χ^2 value is less than the level of significance

3.9.3 R Squared value

R squared is another method of comparing the intercept only model and the unrestricted model. It represents the magnitude by which the error in the intercept model is reduced by the unrestricted model, higher the value of R squared better is the fit of the model, predictability of the model, and the ability of the model to describe the data.

For linear regression, the R squared is simply given by the ratio of difference of the sum of the square of errors of intercept only model and unrestricted model to the sum of the square of errors intercept only term. However, for logistic regression, there is no single R squared and many version of it exists. The most popular version and also given by R is the Mc Fadden's pseudo R squared, which is equal to the ratio of difference of Log-likelihood of the unrestricted model and the intercept only model to the log-likelihood of the intercept only model. Mathematically

$$Mc\ Fadden's\ pseudo\ R^2 = \frac{LL\ of\ unrestricted\ model - LL\ of\ intercept\ only\ model}{LL\ of\ intercept\ only\ model}$$

The higher the value of the pseudo R^2 the better the model is to predict the outcomes as well as to describe the dataset. However, Mc Fadden (1977) while discussing the various methods of model validation introduced the R^2 value and the Rho squared(Mc Fadden's R^2) value and stated that despite many planners practicing ordinary regression analysis are more familiar with the R^2 index it is not well behaved for maximum likelihood estimates as compared to the Rho squared and warned the researchers who are new to the Rho squared that its value is significantly lower than the R^2 index, therefore, the rules applied for R^2 should not be used to judge the models "good fit" and a Rho squared between 0.2 to 0.4 can be taken as an excellent fit

3.10 Model Validation

Model validation was done by determining the choice prediction of the validation data by comparison of the actual and the predicted choices. A confusion matrix whose rows gives the actual choices while the columns gives the predicted choice is generated and was used for this purpose. Accuracy was calculated as the number of correct predictions which is equal to the sum of diagonals divided by the sum of the matrix higher is the accuracy higher is the predictability of the model.

CHAPTER FOUR: MODEL SPECIFICATION AND ANALYSIS

4.1 Preliminary data analysis

The minimum number of total samples to be collected was 576 while the total number of samples collected for the study was 631, after separation, 440(Around one third) were used for model calibration purpose i.e. as training data set and 191 (Around two-third) were used for model validation purpose i.e. as validation data set.

The training data was separated into private campuses and institutional campuses to see variation in composition and mode selection between the two. The nature of mode choice among the colleges was found to be almost the same which ruled out the requirements for the development of separate model formation for these institutions. The nature of the populations as well as their mode choice based on different parameters is shown as follows.

4.1.1 Descriptive statistics

The nature of data according to different categories for private and institutional colleges is shown in table 4.1.

Both types of institutions show similar trends for mode, gender, age group, family education level, Sibling number, footpath availability, and vehicle ownership. On analyzing these categories in the aggregate population, mode wise maximum number of students (47.73%) have preferred public transportation followed by walking (33.86%) and then by personal modes (18.41%), gender wise the proportion of male (78.86%) to female (21.14) ratio is around which is quite different than the overall ratio of Nepal which shows lesser number of female students in engineering stream, age wise highest frequency is that of age 18-22 years(86.59%), followed by 22-26 years (12.5%) and then by 26-30 years (0.91%), Education wise highest frequency is that of students form families with only one parent having high school or higher education (40.23%) followed those from families with both parents having high school or higher education (36.36%) followed by families with none of the parents having high school or higher education (23.41%), with reference to sibling number the highest frequency is that of students having two siblings (47.05%) followed by those having three siblings the by those having one sibling only and finally by those having more or more siblings, with reference to footpath availability more students have footpath in their route of commute (71.82%) then the students not having so (21.18%), lastly with reference to vehicle availability more number of students have access to personal vehicle (62.73%) then the students not having so (37.27%)

Unlike the categories stated above which show similar trends for both types of institutions, family income, family size, trips distance and friend availability show a different trend in private and institutional campuses

Concerning family income, for private campuses the highest frequency is that of students from families with an income of Rs.45000 a month or above (30.19%), followed by students with family income Rs.25000-35000 (28.25%) then by students with family income Rs.35000-45000 (18.18%), then by students with family income Rs.15000-25000 (16.23%) and finally by students with family income less than Rs.15000 (7.14%), similarly for institutional campuses, the highest frequency is that of students from families with an income of Rs.45000 a month or above (34.09%), followed by students with family income Rs.35000-45000(22.73%), then by students with family income Rs.25000-35000 (20.45%), then by students with family income Rs.15000-25000 (12.88%) and finally by students with family income less than Rs.15000 (9.85%).

Concerning for private campuses the highest frequency is that of students having four or less members (49.68%) followed by students from families having five to eight members (46.1%) and then by students from families with eight to twelve members (2.92%) and finally by students from families having more than twelve members (1.31%). Similarly, for institutional campuses the highest frequency is that of students having five to eight members (49.24%) followed by students from families having four or less members (46.21%) and then by students from families with eight to twelve members (4.55), students from families having more than twelve members are nil.

Concerning friend availability, for private campuses students having a friend sharing the route (54.22%) are more than students not having so (45.78%), similarly for institutional campuses students having a friend sharing the route (47.73%) are less than students not having so (52.27%)

Table 4.1: Descriptive data analysis for different categories

Category	**Private Campuses**		**Institutional Campuses**		**Total**	
	Count	**Percentage**	**Count**	**Percentage**	**Count**	**Percentage**
Mode Choice						
Walk	119	38.64	30	22.73	149	33.86
Public	132	42.86	78	59.09	210	47.73
Personal	57	18.51	24	18.18	81	18.41
Gender						
Male	238	77.27	109	82.58	347	78.86
Female	70	22.73	23	17.42	93	21.14
Age Group						
18-22	252	81.82	129	97.73	381	86.59
22-26	52	16.88	3	2.27	55	12.5
26-30	4	1.3	0	0.0	4	0.91
Family Income Level						
< 15000	22	7.14	13	9.85	35	7.95
15000-25000	50	16.23	17	12.88	67	15.23
25000-35000	87	28.25	27	20.45	114	25.91
35000-45000	56	18.18	30	22.73	86	19.55
> 45000	93	30.19	45	34.09	138	31.36
Family Education level						
Both High School	109	35.39	51	38.64	160	36.36
One only High School	122	39.61	55	41.67	177	40.23
None High School	77	25	26	19.7	103	23.41
Family Size						
Upto 4	153	49.68	61	46.21	214	48.64
5 to 8	142	46.1	65	49.24	207	47.05
9 to 12	9	2.92	6	4.55	15	3.41

Category	Private Campuses		Institutional Campuses		Total	
	Count	**Percentage**	**Count**	**Percentage**	**Count**	**Percentage**
More than 12	4	1.3	0	0.0	4	0.91
Sibling Number						
One	46	14.94	30	22.73	76	17.27
Two	148	48.05	59	44.7	207	47.05
Three	73	23.7	29	21.97	102	23.18
Four or more	41	13.31	13	9.85	54	12.27
Trip Distance						
Upto 2 Km	89	28.9	23	17.42	112	25.45
2 to 4 Km	59	19.16	22	16.67	81	18.41
4 to 6 Km	46	14.94	28	21.21	74	16.82
6 to 8 Km	39	12.66	12	9.09	51	11.59
Greater than 8 Km	75	24.35	47	35.61	122	27.73
Friend Availability						
Available	167	54.22	63	47.73	230	52.27
Not Available	141	45.78	69	52.27	210	47.73
Footpath Availability						
Available	219	71.1	97	73.48	316	71.82
Not Available	89	28.9	35	26.52	124	28.18
Vehicle ownership						
Yes	197	63.96	79	59.85	276	62.73
No	111	36.04	53	40.15	164	37.27

4.1.2 Gender wise mode selection

Gender wise mode choice is shown in figure 4.1, both the male students (100 nos private and 61 nos institutional) and female students (32 nos private and 17 nos institutional) have preferred public transport more for both the institution types. The prominent difference between walking and public transport using students in the institutional campuses as compared to the private ones is mainly because of the travel distances of the two types of campuses being different i.e. the institutional campuses having greater average travel distance then private campuses.

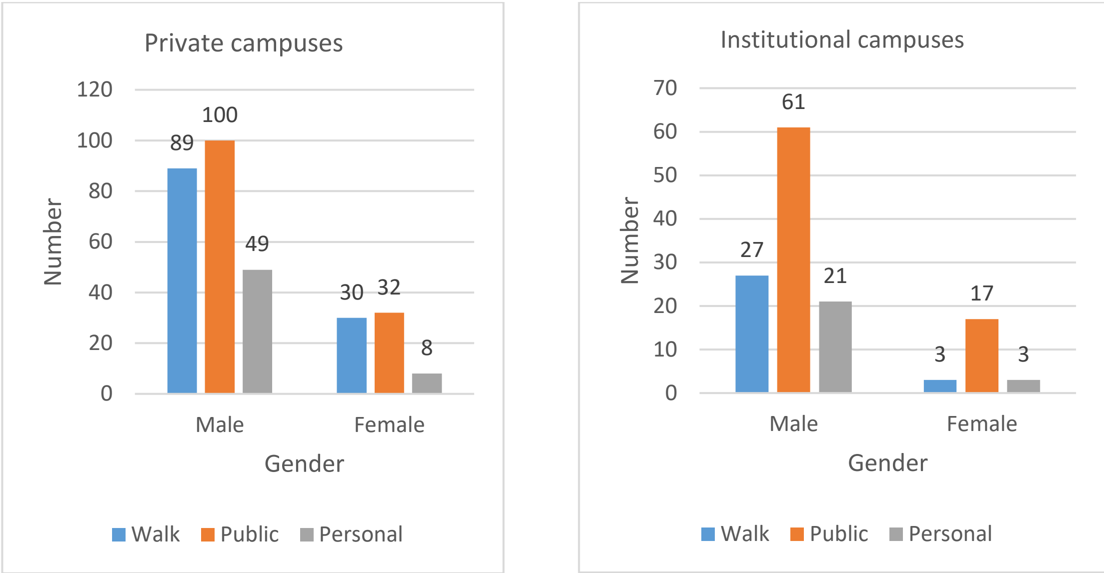
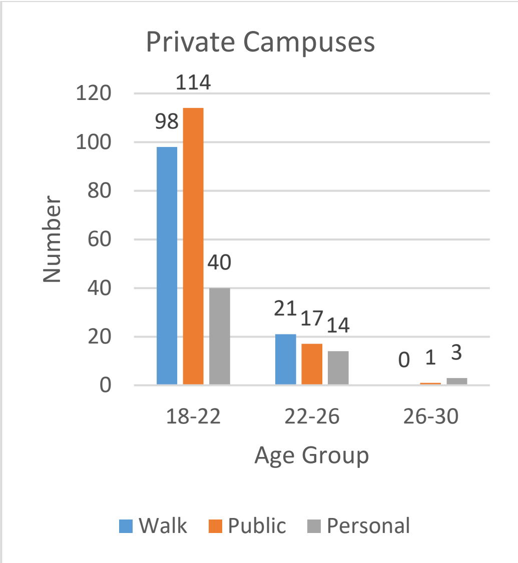


Figure 4.1: Gender wise mode selection

4.1.3 Age-wise mode selection

Age-wise mode choice is shown in figure 4.2, for privates campuses age group 18-22 have preferred public transportation (114 nos), age group 22-26 have slightly preferred the walking mode (21 nos) and despite low in number age group 26-30 years has preferred personal mode of transport (3 nos). Similarly for institutional both age groups 18-22 years (76 nos) and 22-26 years (2 nos) have preferred public transportation while the last age group i.e. 26-30 years is nil. The difference between walking and public transport using students is more prominent in the institutional campuses as compared to the private campuses.



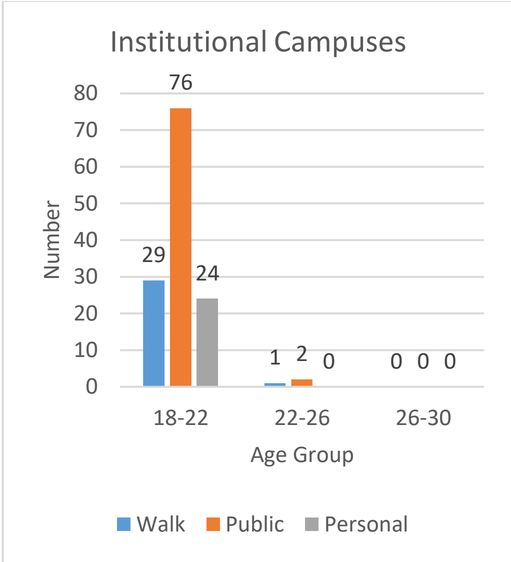
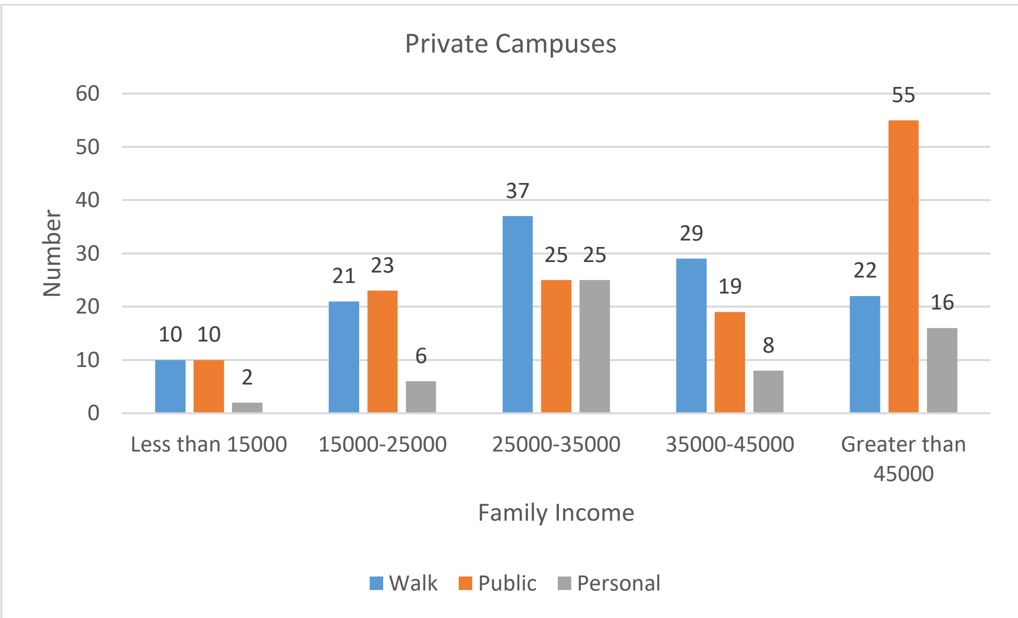


Figure 4.2: Age wise mode selection

4.1.4 Family Income wise mode selection

Family income-wise mode choice is shown in figure 4.3, for private campus students with family income less than Rs 15000 (10 nos) and those with family income more than Rs 45000 (55 nos) have preferred public transportation more, while those with family income between Rs.15000-25000 (21 nos), Rs.25000-35000 (37 nos) and Rs.35000-45000 (29 nos) prefer walking more, whereas all the income groups have preferred public transport in case of institutional colleges, this is due to larger trip distances in these institutional campuses.



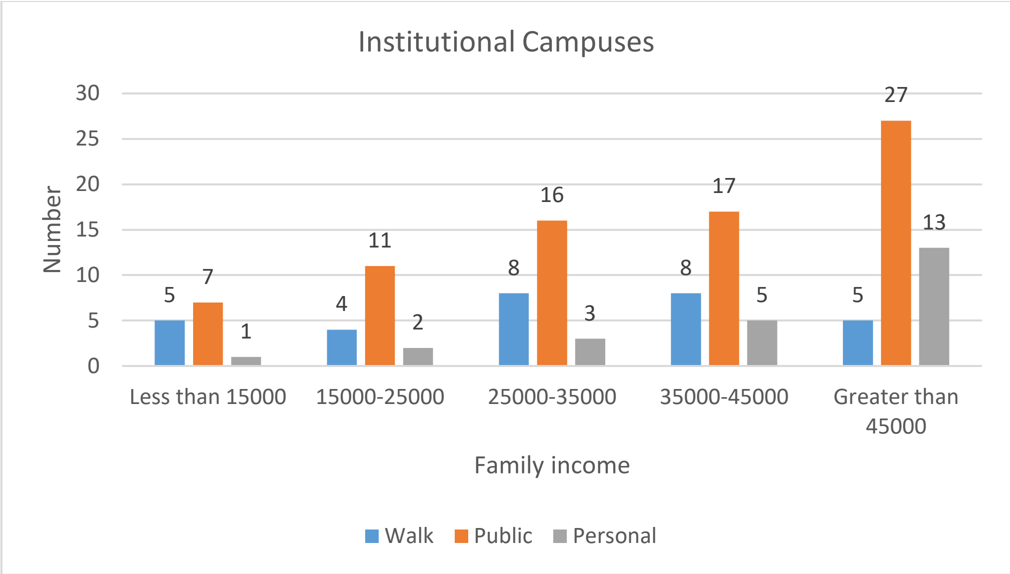
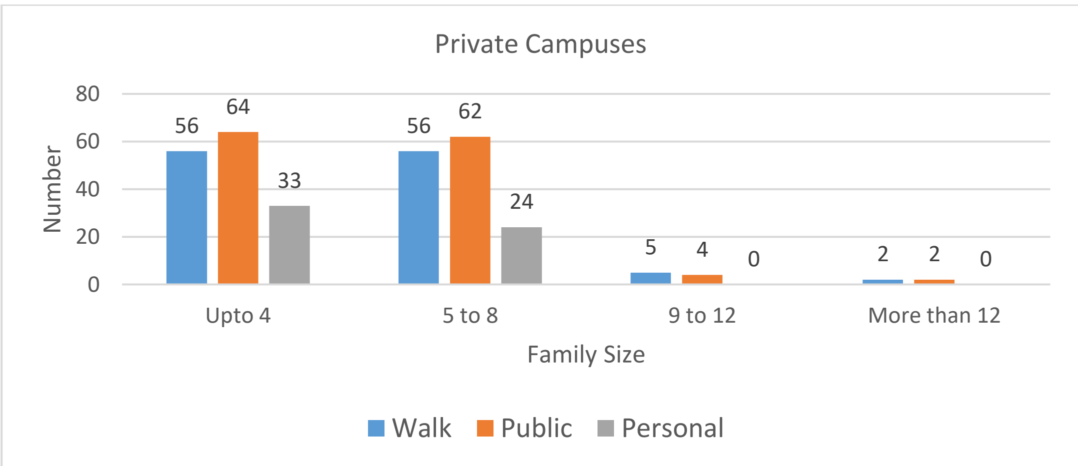


Figure 4.3: Family wise distribution and mode selection

4.1.5 Family size and mode selection

Family size mode choice is shown in figure 4.4, for private campuses students having up to four members in the family (64 nos) and those having five to eight members in the family (62 nos) have preferred public transportation more while students having nine to twelve members (5 nos) and those having twelve or more members in the family have preferred walking more, similarly, for institutional campuses students having up to four members in the family (54 nos), those having five to eight members in the family (24 nos) have preferred public transportation more while students having nine to twelve members prefer walking and private modes equally (2 in each group) and students having twelve or more members in the family are nil.



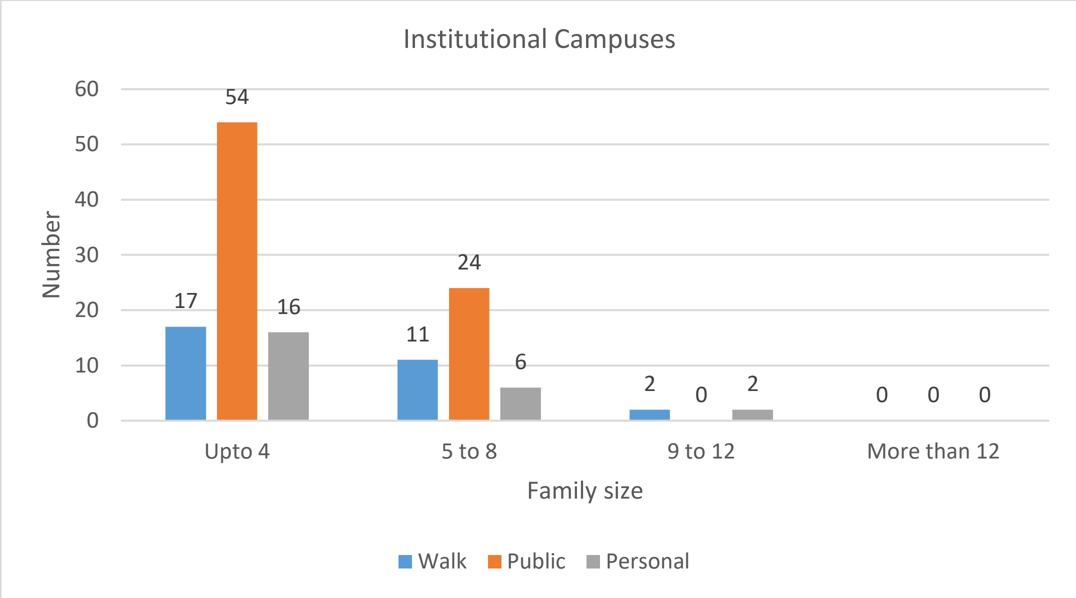


Figure 4.4: Family size and mode selection

4.1.6 Family education and mode selection

Family education-wise mode choice is shown in figure 4.5, Mode wise each family education level group have the highest preference to public mode followed by walking followed by personal mode of travel.

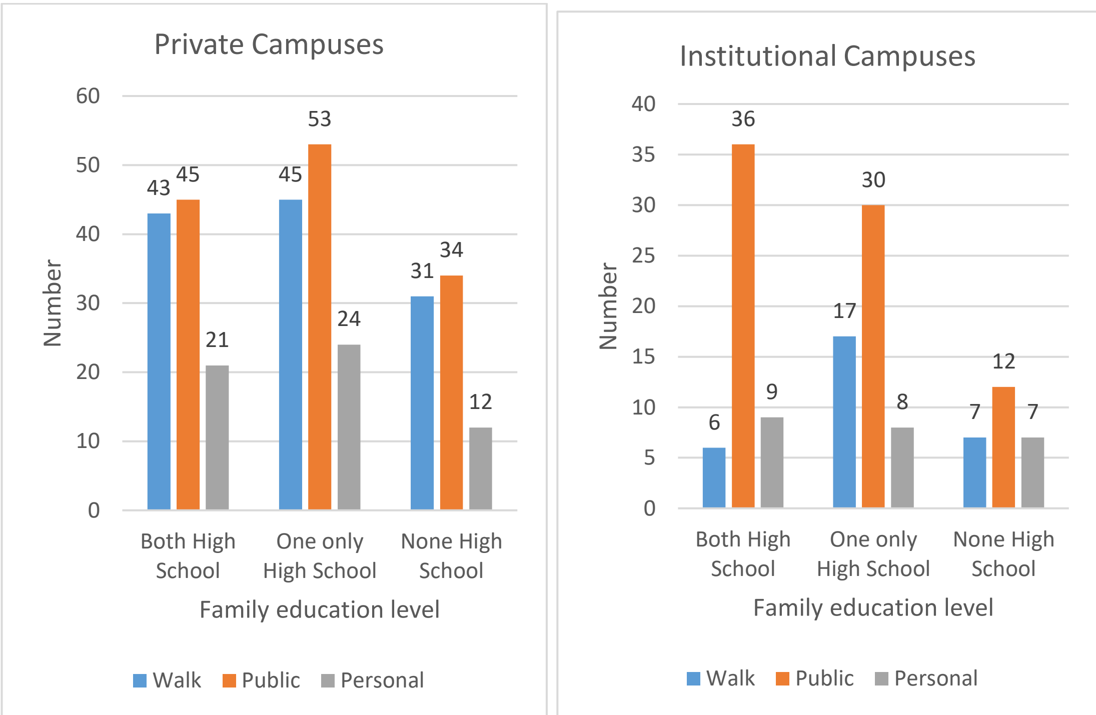


Figure 4.5: Family education and mode selection

4.1.7 Sibling number and mode selection

Mode wise each category based on sibling number has the highest preference for public transport mode, for private campuses, public modes are followed by walking mode followed by personal mode, while for institutional campuses public modes are followed by personal modes for students with one sibling (8 nos) and two siblings (6 nos) while public modes are followed by walking for students having three siblings (4 nos) and four or more siblings (17 nos)

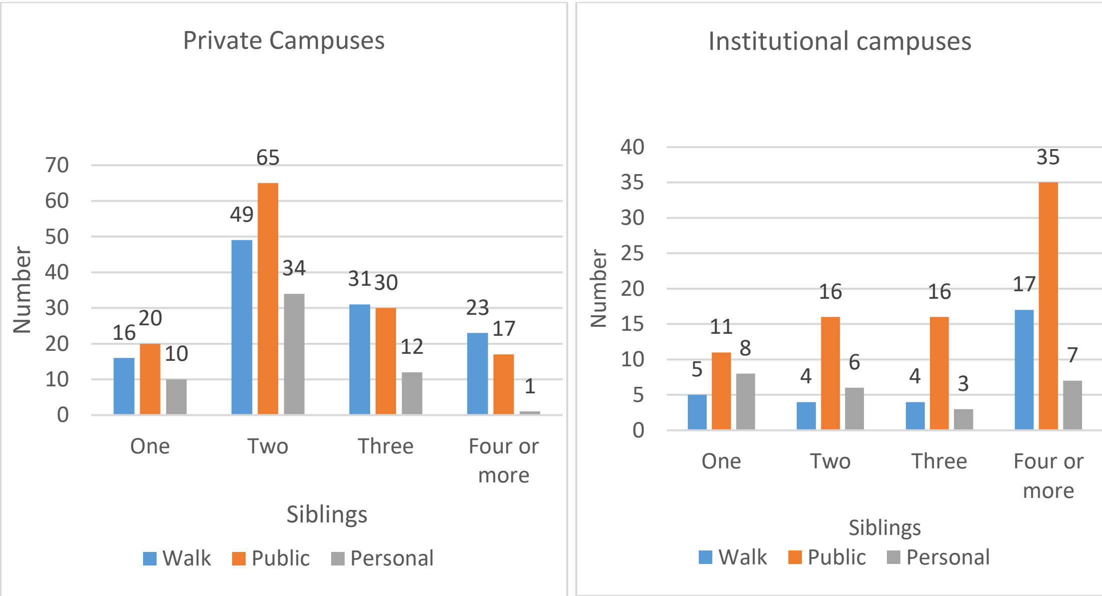
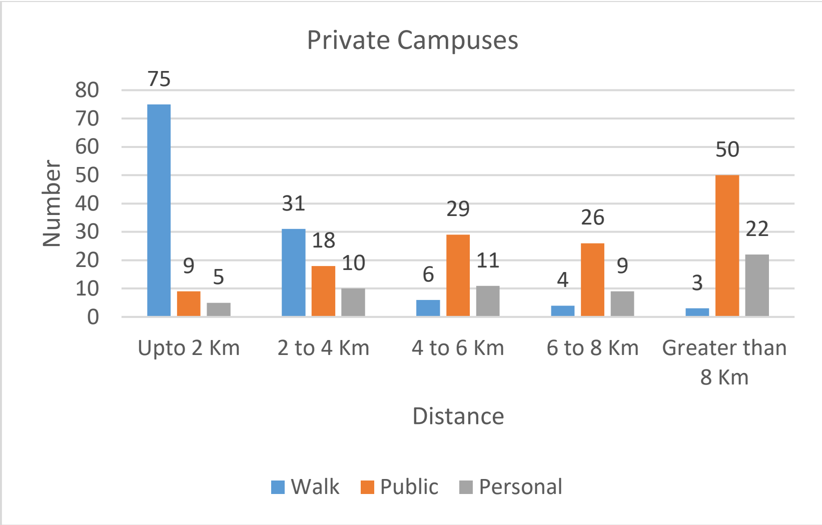


Figure 4.6: Sibling wise mode selection

4.1.8 Distance and mode selection

Distance-wise mode choice is shown in figure 4.7, for private campuses students have preferred walking for lower distances like up to 2 kilometers (75 nos) and 2-4 kilometer (31 nos), while they have preferred public transportation more for greater distances like 4-6 kilometer (29 nos), 6-8 kilometers (26 nos) and more than 8 kilometers (50 nos), similarly for institutional campuses students have preferred walking for lower distances like up to 2 kilometers (18 nos), while they have preferred public transportation more for greater distances like 2-4 kilometers (10 nos), 4-6 kilometer (21 nos), 6-8 kilometers (7 nos) and more than 8 kilometers (35 nos)



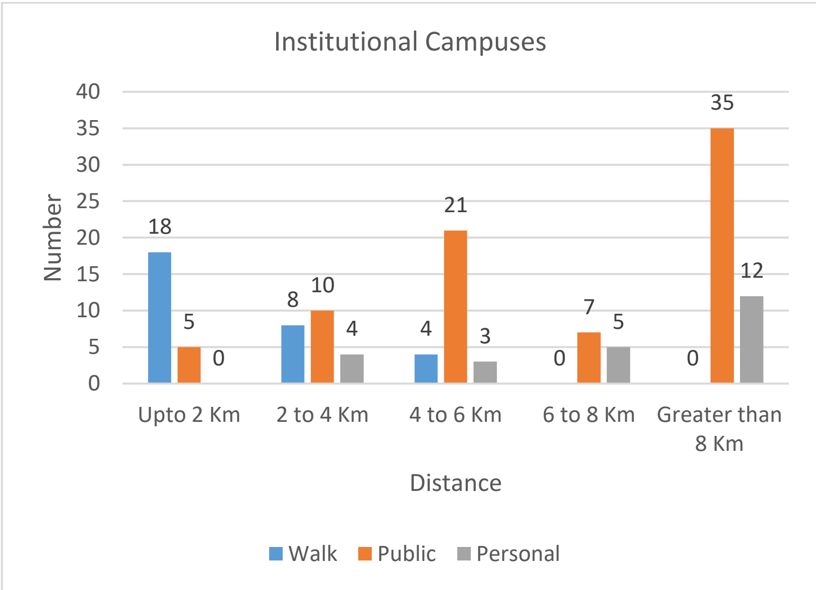


Figure 4.7: Trip distance and mode selection

4.1.9 Friend availability and mode selection

Friend availability wise mode selection is shown in figure 4.8, for private campuses students having friend sharing the route (76 nos) prefer walking more while students not having so (69 nos) prefer public transportation more, similarly for institutional campuses both students having a friend sharing the route (32 nos) and those not having so (46 nos) have preferred public transport mode more

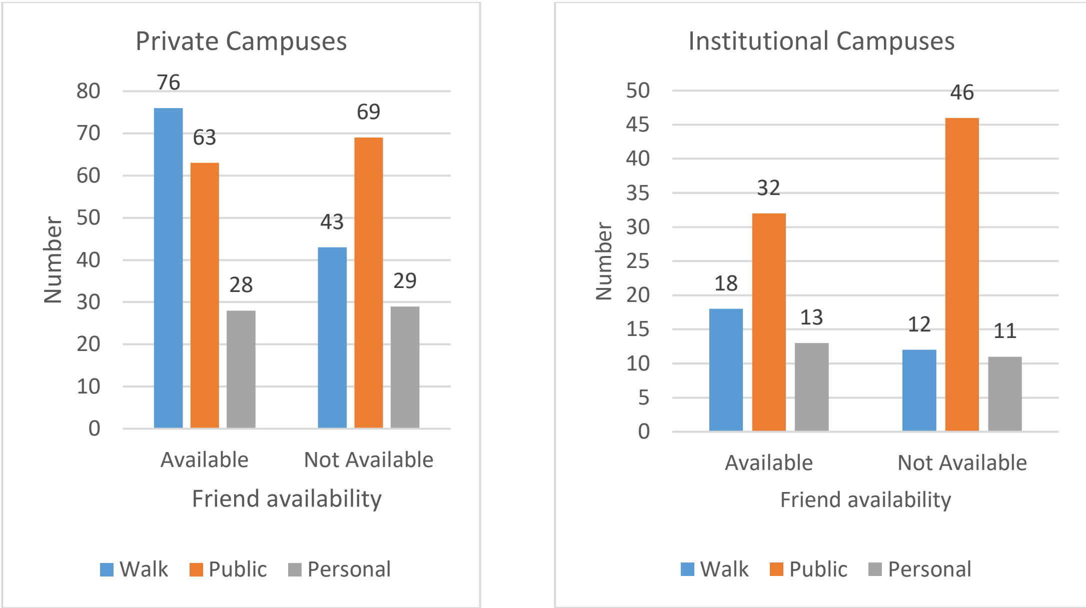


Figure 4.8: Friend availability wise mode selection

4.1.10 Footpath availability and mode selection

Footpath availability wise mode choice is shown in figure 4.9, for private campuses students having footpath in their route commutes (93 nos) have preferred walking more while those not having footpath (41 nos) have preferred public transportation more, similarly for institutional campuses students having footpath (54 nos), as well as students not having so (24 nos), have preferred public transportation

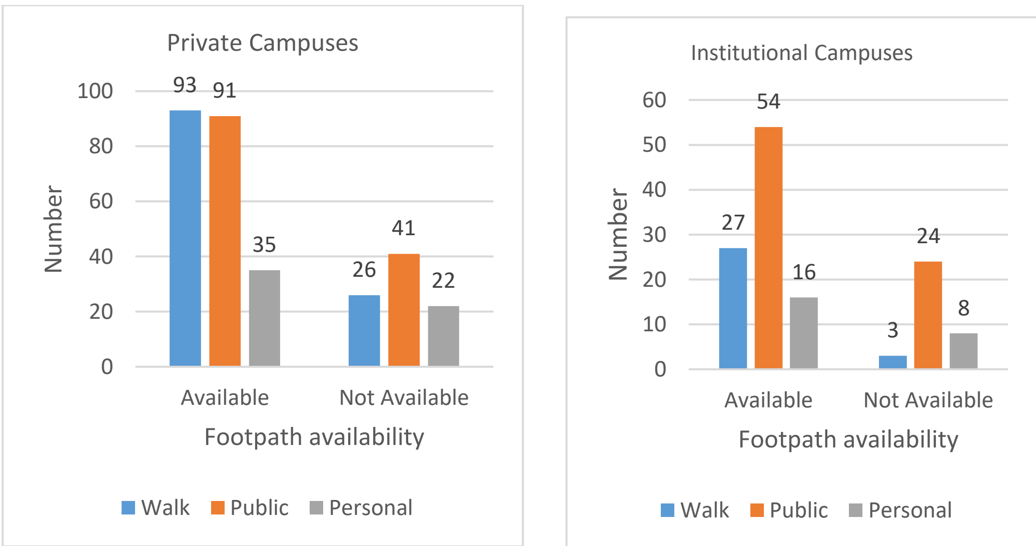
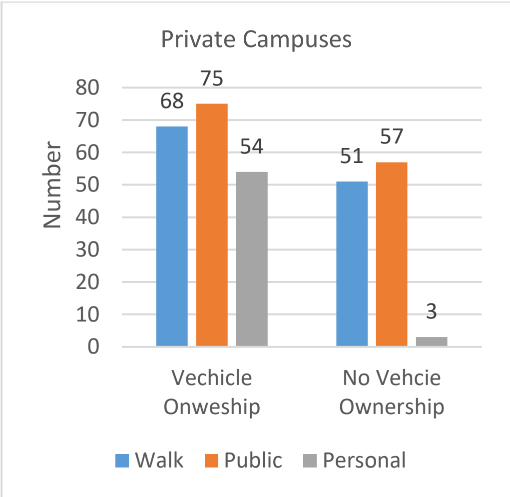


Figure 4.9: Footpath availability and mode choice

4.1.11 Vehicle ownership and mode selection

Vehicle ownership wise mode choice is shown in figure 4.10, for private campuses students owning a vehicle (75 nos) as well as those not doing so (57 nos) have preferred public transport similarly for institutional campuses students owning a vehicle (40 nos) as well as those not doing so (38 nos) have also preferred public transportation



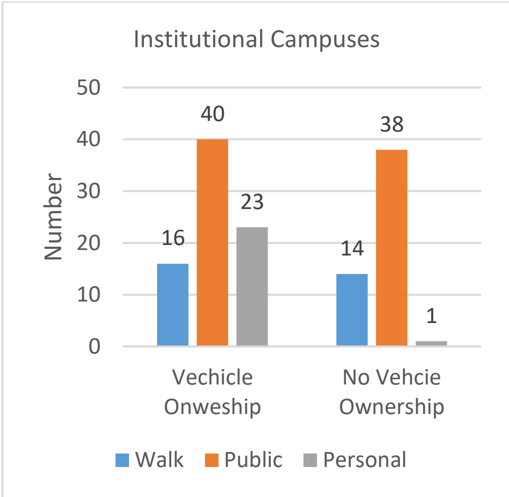


Figure 4.10: Vehicle ownership and mode choice

4.2 Model estimation for revealed preference (RP) models

An initial model was developed using all the collected attributes talking walking as the base mode. The parameters or coefficients of this model for personal modes of transport are shown in table 4.2, while that for public transportation modes is shown in table 4.3 and the model fitting parameters of this model are shown in table 4.4.

Table 4.2: Coefficients and their properties of personal mode for RP full model

Mode: Personal					
Coefficient	**Estimate**	**Std. Error**	**Z Value**	**P_r (>|Z|)**	**Significance**
Intercept	1.914	1.723	1.110	0.267	
Journey Time	0.007	0.004	1.492	0.135	
Age group	0.289	0.471	0.613	0.539	
Gender	-0.818	0.530	-1.542	0.124	
Trip Distance	0.679	0.007	10.355	2.2 X 10^{-6}	0.001
Number of Siblings	-0.353	0.120	-2.941	0.003	0.01
Size of family	-0.165	0.119	-1.386	0.165	

Mode: Personal					
Coefficient	**Estimate**	**Std. Error**	**Z Value**	**P_r (>|Z|)**	**Significance**
Family education level	-0.089	0.278	-0.321	0.748	
Family Income	0.009	0.192	0.0446	0.9644	
Vehicle ownership	-3.259	0.618	-5.27	$1.34X10^{-7}$	0.001
Friend availability	-0.08	0.411	-0.190	0.845	
Footpath Availability	0.743	0.475	1.563	0.117	

Coefficient	Estimate	Std. Error	Z Value	P_r (>|Z|)	Significance
Price	0.0067	0.0065	1.001	0.3125	
Sd. Price	$3.53X10^{-5}$				

Table 4.3: Coefficients and their properties of public transport modes for RP full model

Mode: Public					
Coefficient	**Estimate**	**Std. Error**	**Z Value**	**P_r (>|Z|)**	**Significance**
Intercept	-1.911	1.452	-1.315	0.188	
Journey Time	0.007	0.004	1.492	0.135	
Age group	-0.74	0.48	-1.43	0.133	
Gender	-0.11	0.40	-0.272	0.785	
Trip Distance	0.647	0.005	13.415	$2 X 10^{-6}$	0.001
Number of Siblings	-0.068	0.1120	-0.542	0.587	
Size of family	-0.121	0.081	-1.49	0.135	
Family education level	-0.10	0.224	-0.48	0.631	
Family Income	0.176	0.149	1.174	0.240	
Vehicle ownership	-0.114	0.363	-0.315	0.752	
Friend availability	0.182	0.342	0.537	0.594	
Footpath Availability	0.443	0.419	1.053	0.290	

Coefficient	Estimate	Std. Error	Z Value	P_r (>|Z|)	Significance
Price	0.0067	0.0065	1.001	0.3125	
Sd. Price	$3.53X10^{-5}$				

Table 4.4: Model fitting information for RP full model

Mc Fadden R^2	0.355
Likelihood ratio test	316.04
p-value of likelihood	$2.22X10^{-16}$

The confusion matrix used for the validation of this model is shown in table 4.5, the rows of this matrix denote the actual mode used while the columns denote the mode predicted by the model. The accuracy is the ratio of correct predictions to total sample size i.e. the ratio of the sum of diagonal to the sum of matrix

Table 4.5: Confusion matrix for RP full model

	Public	**Walk**	**Personal**
Public	67	0	0
Walk	86	0	0
Personal	38	0	0
Accuracy			**35.09%**

The initial model showed several parameters being statistically insignificant i.e. $p > 0.05$. The Mc Fadden R^2 is within the limits and the likelihood ratio and its p-value also show that this model is better than the intercept only model. Since statistically significant parameters were few the prediction of the validation data was very poor the model had to be reconfigured. To do this analyzing and grouping terms having a high degree of correlation (e.g. family size and siblings, family education level/family income and vehicle ownership, etc.) was done and only one data that provided a better model among in the group was used. Also attributes having equal distribution of all modes among one another (e.g. each age group almost had the same choice type distribution) were removed to obtain a model with better predictability without losing the Mc Fadden R^2 and the likelihood ratios.

After evaluating many such models which are shown in the codes shown in the appendix, finally, that model was selected after which the removal of statistically insignificant parameters neither added to the improvement of fitting parameters nor the accuracy of the model. The details of the final model is shown below, parameters or coefficients of this model for personal modes of transport are shown in table 4.6, while that for public transportation modes are shown in table 4.7, the model fitting parameters of this model are shown in table 4.8 and the confusion matrix used for the validation of the model is shown in figure 4.9.

Table 4.6: Model processing mode personal for RP final model

Mode: Personal					
Coefficient	**Estimate**	**Std. Error**	**Z Value**	**P_r (>|Z|)**	**Significance**
Intercept	2.349	1.294	1.815	0.069	0.1
Journey Time	0.007	0.004	1.662	0.096	0.1
Gender	-0.901	0.493	-1.827	0.067	0.1
Trip Distance	0.680	0.064	10.474	2.2×10^{-6}	0.001
Number of Siblings	-0.413	0.108	-3.80	$1.43\text{X}10^{-4}$	0.001
Family Income	-0.025	0.177	-0.138	0.89	
Vehicle ownership	-3.127	0.602	-5.18	$2.13\text{X}10^{-7}$	0.001
Friend availability	0.04	0.38	0.114	0.908	

Coefficient	**Estimate**	**Std. Error**	**Z Value**	**P_r (>|Z|)**	**Significance**
Price	0.0066	0.0066	1.00	0.3123	
Sd. Price	$3.53\text{X}10^{-5}$				

Table 4.7: Model processing mode public for RP final model

Mode: Public					
Coefficient	**Estimate**	**Std. Error**	**Z Value**	**P_r (>|Z|)**	**Significance**
Intercept	-3.06	1.052	-2.90	0.004	0.01
Journey Time	0.007	0.004	1.6621	0.096	0.1
Gender	-0.0901	0.390	-0.236	0.813	
Trip Distance	0.656	0.048	13.724	2.2×10^{-6}	0.001
Number of Siblings	-0.094	0.106	-0.895	0.370	
Family Income	0.174	0.142	1.266	0.219	
Vehicle ownership	-0.050	0.335	-0.15	0.881	
Friend availability	0.234	0.326	0.717	0.473	

Coefficient	Estimate	Std. Error	Z Value	P_r (>|Z|)	Significance
Price	0.007	0.007	1.00	0.312	
Sd. Price	$3.53X10^{-5}$				

Table 4.8: Model fitting information RP final model.

Mc Fadden R^2	0.340
Likelihood ratio test	301.62
p-value of likelihood	$2.22X10^{-16}$

The confusion matrix for this model is shown in the table 4.9 below. The accuracy is 74.34 for this model.

Table 4.9: Confusion matrix for RP final model.

	Public	**Walk**	**Personal**
Public	56	11	0
Walk	0	86	0
Personal	28	10	0
Accuracy			**74.34%**

4.3 Modal Explanation of RP models

4.3.1 Personal mode:

With walking as the base criteria Personal mode of transport was found to be affected by trip distance siblings and vehicle ownership only the coefficients of the final model for personal modes are shown in table 4.6. The Coefficients for different attributes can be explained as follows

Journey time: The coefficient of journey time for the personal mode of transport is 0.007 which indicates that with an increase in journey time personal modes of transport are preferred more as compared to the walking mode and the odds of choosing personal travel modes as compared to walking increases by 1.007 times per minute increase in travel time provided other factors remain same. Since more journey time means effort if one walks as a result doing so for too long is not practically feasible for daily commute so the nature of coefficient seems logical. Since p-value of journey time

(0.096) is more than 0.05 therefore the effect is statistically significant at 10% level of significance only

Gender: The coefficient for gender is -0.901 which indicate that females students are less likely to prefer personal mode of transport over walking rather than male students the odds of choosing the personal mode of transport over walking decreases by 0.406 times if the student is female rather than male taking all other factors same. The nature of the coefficient might be due to factors like greater health concern or females not having alternative options of exercise (Clifton et. al. 2004) or absence of any street related harassment etc. but these are not the part of the studies. Since p-value of gender is more than 0.05 therefore the effect statistically significant at 10% level of significance only

Trip distance: The coefficient of journey time for the personal mode of transport is 0.680 which indicates that with an increase in trip distance personal modes of transport are preferred as compared to walking mode, the odds of choosing personal travel modes as compared to walking increases by 1.97 times with each kilometer increase in trip distance provided other factors remain same. Since more trip distance means labor and after a certain logical point, it is not practically feasible to walk on a daily commute so the nature of coefficient seems logical. Since the p-value of journey time (0.000) is less than 0.05 therefore the effect is statistically significant.

Number of siblings: The coefficient for sibling number is -0.413 which indicates that students with more siblings are less likely to prefer personal mode of transport over walking, the odds of choosing the personal mode of transport over walking decreases by 0.661 times for each extra sibling provided all other factors are same. Since personal mode of transport cannot accommodate many people and the fact that walking with siblings is less tiresome as compared to walking alone the nature of the coefficient seems logical. Since the p-value of siblings (0.000) is less than 0.05 therefore the effect is statistically significant.

Family income: The coefficient for family income is -0.025 which indicate that students higher family income is slightly less likely to prefer personal mode of transport over walking the odds of choosing the personal mode of transport over walking decreases by 0.98 times for each Nrs.10000 increase in family income provided all other factors are same. This seems opposite to what one will expect, this effect might be due to greater health consciousness in the high-income family, view towards use of personal modes during student life, etc. but these are not considered in this study. Since the p-value of family income (0.89) is far more than 0.05 the effect is not statistically significant.

Vehicle Ownership: The coefficient for vehicle ownership is -3.12 which indicate that students without vehicle ownership in the family which they can use or can pick/drop them are highly unlikely to prefer personal mode of transport over walking, although the trip is possible by sharing a mode with a friend but this case is not seen in the database, the odds of choosing a personal mode of transport over walking decreases by 0.044 times if there is no vehicle ownership in the family provided all other factors are same. Since personal vehicle use without its ownership is highly unlikely the nature of this coefficient needs no further explanation. Since the p-value of vehicle ownership (0.000000213) less than 0.05 therefore the effect is statistically significant.

Friend availability: The coefficient for friend availability is 0.044 which indicate that students without friends sharing the same route are more likely to prefer personal mode of transport over walking, the odds of choosing the personal mode of transport over walking increases by 1.04 times if there is lack of friend sharing the same route provided all other factors are same. Since walking alone is cumbersome especially with footpaths missing in many cases and the existing ones not being commuter-friendly the nature of this coefficient seems logical. Since the p-value of friend availability (0.908) is far more than 0.05 therefore the effect is not statistically significant.

Price: The coefficient of price is 0.006 with a standard deviation of 0.0004 which indicate that with an increase in price the students are more likely to use the personal mode of transport than walking, the odds of choosing the personal mode of transport over walking increases by 1.006 times if there is increase of Nrs.1 in the travel cost provided all other factors are same. Since using a personal mode is more comfortable as compared to walking more preference to it when there is price increase seems logical. Since the p-value of price (0.312) is far more than 0.05 therefore the effect is not statistically significant.

4.3.2 Public mode:

With walking as the base criteria public mode of transport are found to be affected by trip distance siblings and vehicle ownership only, the coefficients of the final model for public transportation modes are shown in table 4.7. The Coefficients for different attributes can be explained as follows

Journey time: The coefficient of journey time for the personal mode of transport is 0.007 which indicates that with an increase in journey time public mode of transport is preferred as compared to walking mode, the odds of choosing personal travel modes as compared to walking increases by 1.007 times with each minute increase in travel time provided other factors remain same. Since more journey time means more labor for walking doing so for daily commute is not practical therefore the nature of coefficient seems logical. Since the p-value of journey time (0.091) is more than 0.05 therefore the effect is statistically significant at a 10% level of significance only.

Gender: The coefficient for gender is - 0.0901 which indicate that females students are slightly less likely to prefer public mode of transport over walking rather than male students, the odds of choosing the public mode of transport over walking decreases by 0.913 times if the student is female rather than male taking all other factors same. The nature of this coefficient despite being of same the value is far less as compared to personal modes of transport (- 0.901) indicating that females are more likely to choose public modes as compared to personal modes for non-walking commutes. The nature of the coefficient might be due to lack of comfort, feeling of insecurity, and harassment in public transportation as reported by many female commuters (World Bank, 2012), but these are not the part of the studies. Since the p-value of gender (0.813) is far more than 0.05 the effect is statistically insignificant.

Trip distance: The coefficient of journey time for the personal mode of transport is 0.656 which indicates that with an increase in trip distance public modes of transport are preferred more as compared to the walking mode and the odds of choosing public travel modes as compared to walking increases by 1.927 times per kilometer increase in trip distance provided other factors remain same. Since longer trip distance means more labor and after a certain logical point, it is not practically feasible to walk on a daily commute thus the nature of coefficient seems logical. Since the p-value of trips distance (0.000) is less than 0.05 therefore the effect is statistically significant.

Number of siblings: The coefficient for sibling number is -0.094 which indicates that students with more siblings are less likely to prefer public mode of transport over walking the odds of choosing the public mode of transport over walking decreases by 0.910 times for each extra sibling provided all other factors are same. The nature of this coefficient despite being of same the value is far less as compared to personal modes of transport indicating that with a greater number of siblings students are more likely to choose public modes as compared to personal mode for non-walking commutes. Since public transportation in Kathmandu valley is crowded and unsafe especially for smaller children (World Bank, 2012) and also since walking with siblings is less tiresome as compared to walking alone the nature of the coefficient seems logical. Since the P-value of siblings (0.37) is more than 0.05 therefore the effect is statistically insignificant.

Family income: The coefficient for family income is 0.174 which indicate that students with higher family income are more likely to prefer public mode of transport over walking the odds of choosing the public mode of transport over walking increases by 1.19 times for each Nrs.10000 increase in family income provided all other factors are same. With greater income it is easier to afford public transportation therefore the nature of the coefficient seems logical. The P-value of family income (0.219) is more than 0.05 therefore the effect is statistically insignificant.

Vehicle Ownership: The coefficient for vehicle ownership is -0.05 which indicates that students without vehicle ownership in the family which they can use or which is available to pick/drop them are less likely to prefer public mode of transport over walking. The odds of choosing the public mode of transport over walking decrease by 0.951 times if there is no vehicle ownership in the family provided all other factors are the same. Without vehicle ownership, there is a lack of vehicle dependent psychology in the commuters which decreases the interests towards public transportation also therefore the nature of the coefficient seems logical. Since the p-value of vehicle ownership (0.881) is far more than 0.05 therefore the effect is statistically insignificant.

Friend availability: The coefficient for friend availability is 0.234 which indicate that students without friends sharing the same route are more likely to prefer public mode of transport over walking, the odds of choosing the public mode of transport over walking increases by 1.263 times if there is lack of friend sharing the same route provided all other factors are same. Since walking alone is cumbersome especially with footpaths missing the nature of this coefficient seems logical. Since the p-value of friend availability (0.473) is far more than 0.05 therefore the effect is statistically insignificant.

Price: The coefficient of price is 0.007 with a standard deviation of 0.00000356 which indicate that with an increase in price the students are more likely to use the public mode of transport than walking, the odds of choosing the public mode of transport over walking increases by 1.007 times if there is increase of Nrs.1 in the travel cost provided all other factors are same. Since using a public mode is more comfortable as compared to walking more preference to it when there is price increase seems logical. The P-value of price (0.3123) is more than 0.05 therefore the effect is statistically insignificant.

4.4 Sensitivity analysis of Revealed preference mode choice

A sensitivity analysis describes the effects of the variation of one parameter on the nature of the function or system. In mode choice sensitivity analysis is done by creating probability plots, these plots present the variation of the probability of various choices concerning a change in one parameter keeping all others the same. For the revealed modes the choice depends on three factors trip distance, number of siblings and vehicle ownership to perform the sensitivity of these attributes, two sets of probability plots one showing the variation in the probability concerning distance and the other with several siblings respectively with different curves for presence and absence of vehicle ownerships are generated which are shown below

4.4.1 Trip Distance

The probability plots for trip distance for vehicle ownership and non-ownership are shown in figure 4.12. It can be seen that the probability of walking is high at lower distances irrespective of vehicle ownerships, similarly, for larger distances, the probability of using public modes increases with increase in distance, it being lower for cases where there is vehicle ownership and the probability of using personal modes increases with increase in distance, the probability being higher for cases where there is vehicle ownership.

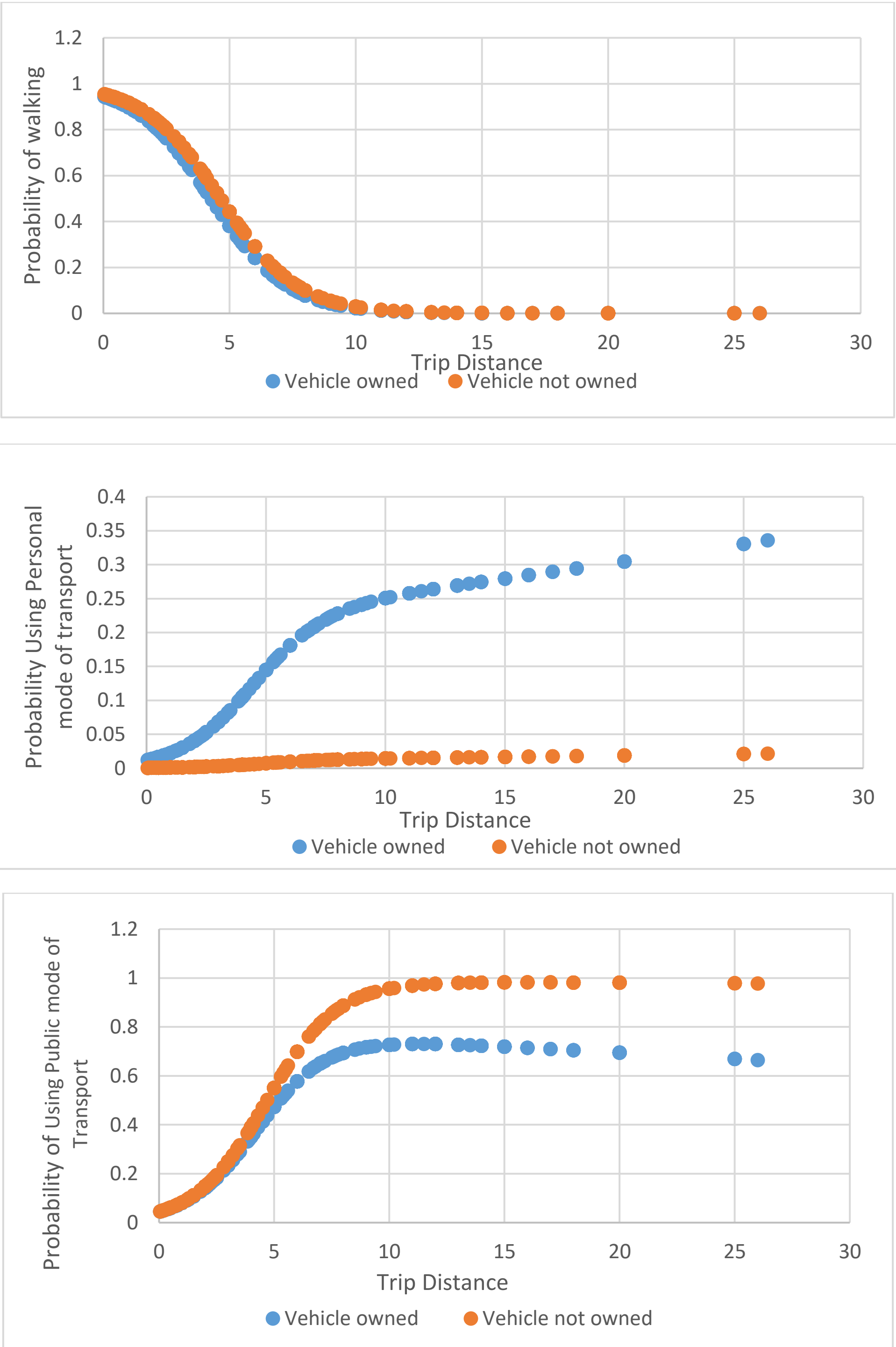
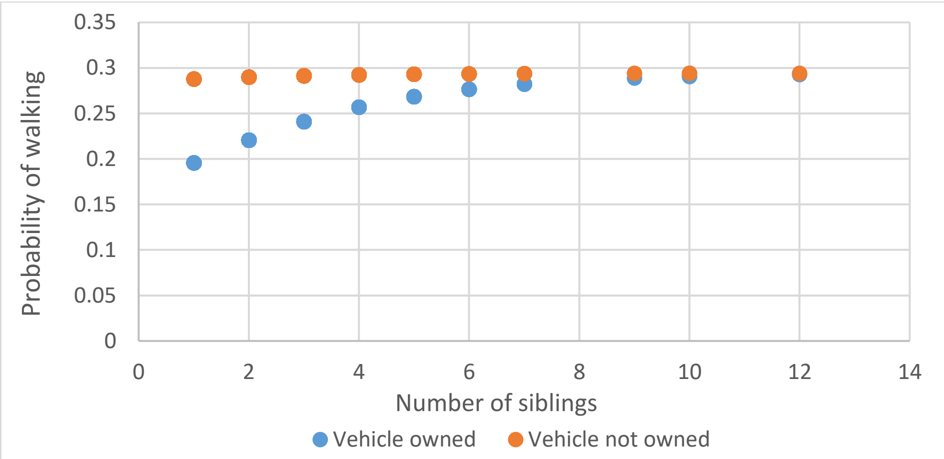
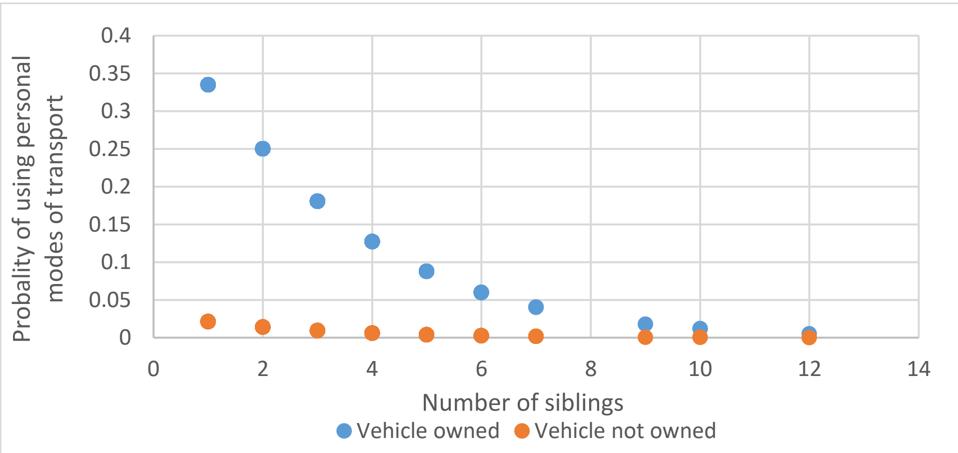


Figure 4.12: Probability plots for sensitivity analysis of trips distance

4.4.2 Sibling Number

The probability plots for sibling number for vehicle ownership and non-ownership are shown in figure 4.13. It can be seen that the probability of walking increases as siblings are more, the probability being less if there is vehicle ownership for lesser number of siblings whereas for a higher number of siblings it is same irrespective of vehicle ownership, similarly, the probability of using public modes increases with increase in distance, the probability being less if there is vehicle ownership for lesser number of siblings whereas for a higher number of siblings it is same irrespective of vehicle ownership and finally the probability of using personal modes decreases with increase in sibling number for cases with vehicle ownerships whereas it is negligible for cases with lack of vehicle ownership.





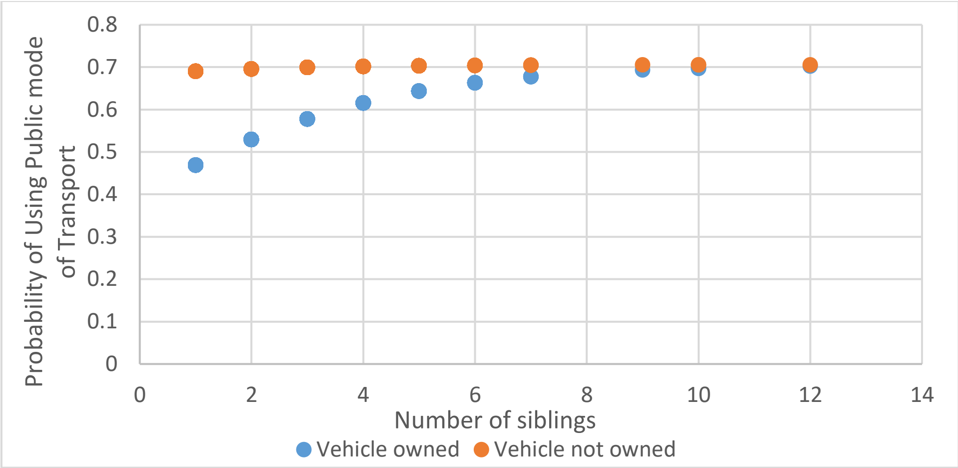


Figure 4.13: Probability plots for sensitivity analysis of Sibling number

4.5 Estimation of results of stated preference (SP) models

Respondents were asked whether they would like to switch to any one of the following four alternative transportation models from the current modes. The answers obtained are shown in figure 4.14. forty-three Students did not want to shift their current travel mode, on analyzing the choice of those who wished to shift it can be seen the highest number students wanted to shift to a metro rail having the same cost and almost same crowding as currently existing public transportation modes, followed by a comparatively costlier metro rail with lesser crowding, followed by a BRT having the same cost and almost same crowding as their currently existing public transportation modes, followed by a costly yet comfortable and lesser crowded bus option

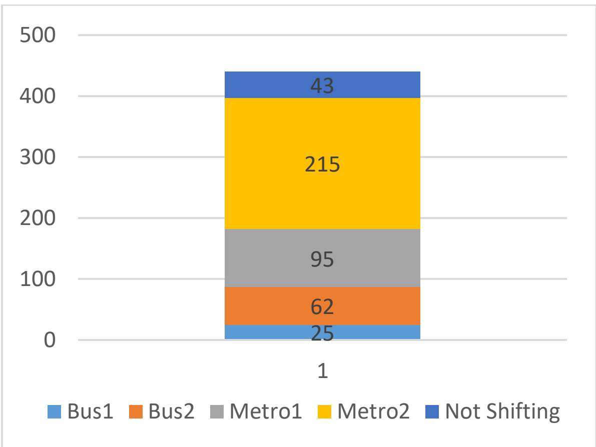


Figure 4.14: Modal shift towards alternative modes

After replacing the choice options with the alternative choice for each of the 4 alternatives the modified date sets were obtained their nature for "Bus1" and "Bus2" mode are shown in figure 4.15 and for "Metro1" and "Metro2" mode are shown in figure 4.16

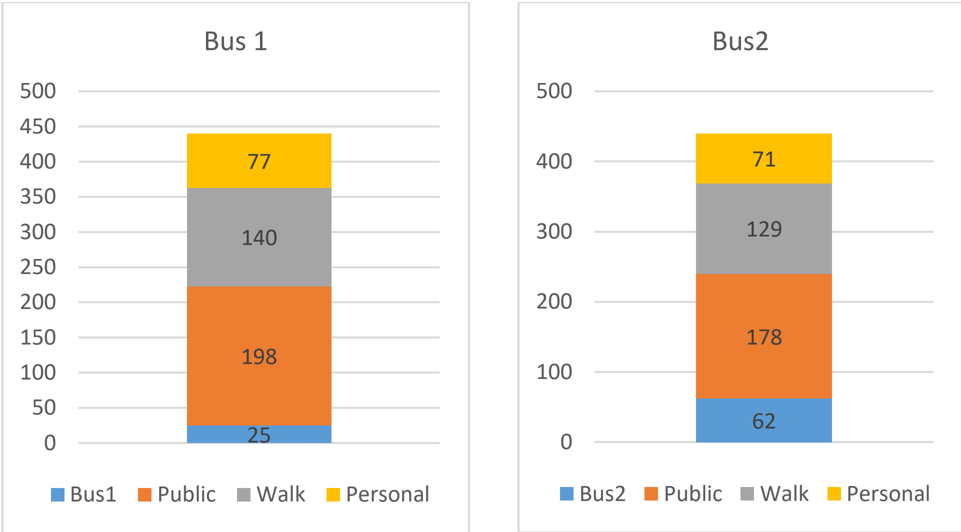


Figure 4.15: Modal choice sets obtained after modification for Bus options

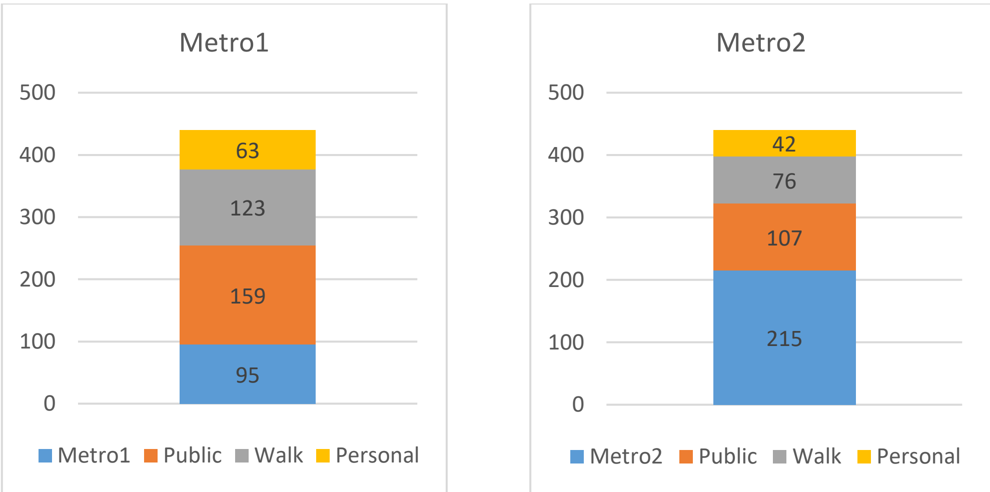


Figure 4.16: Modal choice sets obtained after modification for Metro options

Initially, full models were developed using all the attributes collected taking walking as the base criteria for each of the above data set shown it showed several parameters being statistically insignificant i.e. $p > 0.05$, few more models were examined, in which the nature of coefficients were the same with varying magnitudes with almost same R^2 value the below-shown model was selected finally for the datasets. The coefficients and their properties of the model for "Bus1" mode are shown in table 4.10 and model fitting information for that model are shown in table 4.11, The coefficients and their properties of the model for “Bus2” mode are shown in table 4.12 and model fitting information for that model are shown in table 4.13, The coefficients and their properties of the model for “Metro1” mode are shown in table 4.14 and model fitting information for that model are shown in table 4.15, The coefficients and their properties of the model for “Metro2” mode are shown in table 4.16 and model fitting information for that model are shown in table 4.17.

Table 4.10: Model processing SP modal Choice Bus 1

Mode: Bus1 (Base criteria walking)					
Coefficient	**Estimate**	**Std. Error**	**Z Value**	**Pr (>|Z|)**	**Significance**
Intercept	-4.38	1.622	-2.70	0.006	0.01
Journey time	0.008	0.004	1.838	0.061	0.1
Gender	0.62	0.58	1.07	0.281	
Trip Distance	0.498	0.076	6.54	$5.94X10^{-11}$	0.001
Sibling Number	-0.17	0.2653	-0.65	0.51	
Family Income	-0.205	0.225	-0.91	0.362	
Vehicle Ownership	-0.477	0.528	-0.904	0.388	
Footpath Availability	1.360	0.561	2.429	0.015	0.05
Price	0.006	0.0061	0.94	0.346	
SD of Price	0.003				

Table 4.11: Model fitting information SP modal Choice Bus 1

Mc Faden R^2	0.284
Likelihood ratio test	293.06
p-value of likelihood	$2.22X10^{-16}$

Table 4.12: Model processing SP modal Choice Bus 2

Mode: Bus2 (Base criteria walking)					
Coefficient	**Estimate**	**Std. Error**	**Z Value**	**Pr (>|Z|)**	**Significance**
Intercept	--3.89	1.10	-3.52	0.0004	0.001
Journey time	0.007	0.004	1.63	0.102	

Mode: Bus2 (Base criteria walking)					
Coefficient	**Estimate**	**Std. Error**	**Z Value**	**Pr (>|Z|)**	**Significance**
Gender	0.064	0.436	0.142	0.883	
Trip Distance	0.528	0.051	10.208	$2.2X10^{-16}$	0.001
Sibling Number	0.066	0.117	0.558	0.576	
Family Income	0.082	0.159	0.513	0.608	
Vehicle Ownership	-0.126	0.372	-0.339	0.734	
Footpath Availability	0.702	0.405	1.735	0.083	0.1
Price	0.004	0.007	0.604	0.548	
SD of Price	0.001				

Table 4.13: Model fitting information SP modal Choice Bus 1

Mc Faden R^2	0.2351
Likelihood ratio test	263.7
p-value of likelihood	$2.24X10^{-16}$

Table 4.14: Model processing SP modal Choice Metro 1

Mode: Metro1 (Base criteria walking)					
Coefficient	**Estimate**	**Std. Error**	**Z Value**	**Pr (>|Z|)**	**Significance**
Intercept	-1.065	1.11	-0.95	0.338	
Journey time	0.0112	0.004	2.47	0.013	0.05
Gender	-0.916	0.427	-2.14	0.032	0.05
Trip Distance	0.547	0.052	10.415	$2.2X10^{-16}$	0.001
Sibling Number	-0.193	0.128	-1.50	0.131	
Family Income	0.221	0.154	1.43	0.151	
Vehicle Ownership	-0.655	0.354	-1.853	0.064	0.1
Footpath Availability	0.381	0.392	0.97	0.330	
Price	0.004	0.007	0.604	0.548	
SD of Price	0.0057				

Table 4.15: Model fitting information SP modal Choice Metro 1

Mc Faden R^2	0.201
Likelihood ratio test	233.34
p-value of likelihood	$2.24X10^{-16}$

Table 4.16: Model processing SP modal Choice Metro 2

Mode: Metro2 (Base criteria walking)					
Coefficient	**Estimate**	**Std. Error**	**Z Value**	**Pr (>|Z|)**	**Significance**
Intercept	-0.566	1.03	-0.54	0.58	
Journey time	0.004	0.005	0.8	0.43	
Gender	0.216	0.392	0.552	0.580	
Trip Distance	0.567	0.102	5.54	$2.94X10^{-8}$	0.001
Sibling Number	-0.194	0.10	-1.94	0.075	0.1
Family Income	0.125	0.141	0.883	0.376	
Vehicle Ownership	-0.067	0.30	-0.215	0.83	
Footpath Availability	-0.241	0.388	-0.62	0.534	
Price	0.003	0.011	0.537	0.720	
SD of Price	-0.011				

Table 4.17: Model fitting information SP modal Choice Metro 2

Mc Faden R^2	0.19
Likelihood ratio test	198.85
p-value of likelihood	$2.22X10^{-16}$

4.6 Modal Explanation of SP models

The coefficients obtained for the different models can be described as follows

4.6.1 Mode Bus 1:

The coefficients for the model of mode “Bus1” are shown in table 4.10. The probability of choosing Bus1 over walking is influenced at 5% level of significance by trip distance with coefficient of 0.498 and p-value of 0.001 and footpath availability with a coefficient of 1.360 and p-value of 0.015, similarly it is influenced at 10% level of significance by journey time with a coefficient of 0.008 and p-value of 0.061. The effect of attributes other than these are found to be statistically insignificant. The positive nature of the coefficients for all three attributes indicates that an increase in trip distance, increase in journey time and lack of footpath facility increases the odds of choosing Bus1 over walking.

4.6.2 Mode Bus 2:

The coefficients for the model of mode "Bus2" are shown in table 4.12. The probability of choosing Bus2 over walking is influenced 5% level of significance by trip distance with a coefficient of 0.528 and a p-value of 0.001 and at a 10% level of significance by footpath availability with a coefficient of 0.702 and p-value of 0.083. The effect of attributes other than these are found to be statistically insignificant. The positive nature of the coefficients for these attributes indicate that increase in trip distance and lack of footpath facility increases the odds of choosing Bus1 over walking.

4.6.3 Mode Metro 1:

The coefficients for the model of mode "Metro1" are shown in table 4.14. The probability of choosing Metro1 over walking is influenced at 5% level of significance by journey time with a coefficient of 0.011 and p-value of 0.013, gender with coefficient of -0.916 and p-value of 0.032 and trip distance with a coefficient of 0.547 and p-value of 0.001, similarly it is influenced at 10% level of significance by vehicle ownership with a coefficient of -0.655 and p-value of 0.064. The effect of attributes other than these are found to be statistically insignificant. The positive nature of the coefficients for journey time and trip distance indicates that an increase in trip distance and journey time will increase the odds of choosing Metro1 over walking. Similarly, the negative nature of the coefficient of gender and vehicle ownership indicates that the student's gender being female rather than male and lack of vehicle ownership will decrease the odds of choosing metro1 against walking

4.6.4 Mode Metro 1:

The coefficients for the model of mode “Metro2” are shown in table 4.16. The probability of choosing Metro2 over walking is influenced at a 5% level of significance by trip distance with a coefficient of 0.567 and a p-value of 0.001, similarly, it is influenced at a 10% level of significance by the number of siblings with a coefficient of -0.194 and p-value of 0.074. The effect of attributes other than these are found to be statistically insignificant. The positive nature of the coefficients for trip distance indicates that an increase in trip distance will increase the odds of choosing Metro2 over walking. Similarly, the negative nature of the coefficient number of siblings indicates that more number of siblings will decrease the odds of choosing metro1 against walking.

CHAPTER FIVE: CONCLUSION AND RECOMMENDATION

5.1 Conclusion

This study provides an insight into what factors affect the mode of the study group. It also provides insights on the type of vehicular mode of transport preferred for motorized commutes. Despite the share of these trips being small, the findings can guide the policymakers, planners, and infrastructure-related stakeholders about the conditions that demand motorized or non-motorized commutes for the trips of this group so that the preferred mode of choice can be developed ensuring the optimum satisfaction of the users.

Among the existing modes of transport, for personal modes of transport taking walking as base criteria, trip distance with a coefficient of 0.680, is found to increase the use of personal modes if distance increase, sibling number with a coefficient of - 0.413 and vehicle ownership with a coefficient of -3.127 is found to decrease the use of personal modes if sibling number increase or there exists lack of vehicle ownership. On the other hand, for the public mode of transport taking walking as the base criterion trip distance with a coefficient of 0.656 is found to increase the use of public modes of distance increase

Since motorized commutes are more preferred in longer distance and walking commutes in shorter ones, it is recommended to develop proper mass transit facilities between the residential areas and the educational hubs while developing proper walking conditions within the educational hub for the commutes done by this group.

Another important finding in the study was that females prefer walking more than motorized for this group, based on which this is recommended that more gender-friendly walking conditions be promoted this can be done by providing proper lighting, proper visibility in roads, escape paths, a lower degree of isolation, proper footpath facilities etc. this not only providing better services to currently walking female commuters can also further aid to increase the part of walking women, another finding for females is that the probability choosing public transport is more as compared to private modes in longer distance, based on this it is recommended to make the public transport more women-friendly, this not only provides better facilities to existing female users but can also aid to increase the part of public transportation using females.

Lastly, vehicle ownership is found to decrease the odds of walking or public transportation being used therefore planning and policy interventions are recommended in this field.

For the stated preference parts taking walking as base criteria, for mode “Bus1” trip distance and footpath availability with coefficients 0.498 and 1.36 were found to increase the use of “Bus1” if trip distance increases or footpath is absent, similarly for mode “Bus2” trip distance with a coefficient of 0.528 was found to increase the use of “Bus2” mode if distance increases, similarly for mode “Metro1” Journey time gender and trip distance with a coefficient of 0.0112, -0.916 and 0.547 respectively were found to increase the use of “Metro1” if journey time or trip distance increase or the user is male and finally for mode “Metro2” trip distance with a coefficient of 0.567 was found to increase the use of “Metro2” if trip distance increase.

Since trip distance is found to increase the use of these alternatives and they are preferred only for longer distances only. The purposed alternatives are recommended to be used considering the long-distance commutes only in context trips of this group in Kathmandu valley.

It was also observed that the coefficient of “Metro2” option is highest and modal shift for it is maximum therefore it can be concluded a general liking in the study group is seen for it and further studies on this mode seems justified and a metro rail option seems most feasible as an alternative transport for the current population.

5.2 Recommendation for further research

The study has used the most widely used model for mode choice analysis, other discrete choice models like nested logit, MLP neural network, etc. can be applied to provide additional insights or comparison to the evaluated models. A combined model is formed for revealed preference analysis of private and institutional colleges as the mode choice was almost the same. This study recommends the use of above stated tools with separate or combined models of private and institutional organizations. Since feeder availability is a major requirement for metro operation which was a general liking in this group it is recommended that the analysis of feeder to metro network in Kathmandu be performed.

REFERENCES

Ahmedabad BRT fare (2019): http://www.ahmedabadbrts.org/

Ahmedabad Bus service fare (2019): http://www.amts.co.in/

Baghestani, A., Mamdoohi, A.R., Abrishami, S., (2014), "A Comparison of Aggregate and Disaggregate Mode Choice Models Transferability." , International Journal of Transportation Engineering, Vol.2/No.2, pp 146 – 154.

Bai T., Lia, X., Sunc, Z., (2017), "Effects of cost adjustment on travel mode choice: analysis and comparison of different logit models." , Transportation Research Procedia 25C pp 2653–2663.

Balsas, C.J.L., (2003), "Sustainable transportation planning on college campuses", Transport policy pp 35-39.

Blossfeld, H.P., Klijzing. E., Mills, M., Kurz, K., (January 2006), "Globalization Uncertainty and Youth in Society", First edition, Routledge, London.

CANN., (2010), "CANN Fact sheet no 1: Walkability in Kathmandu", Clean Air Network Nepal, Kathmandu

Central Bureau of Statistics (CBS)., (2017), "Population of 753 Local Unit", Kathmandu.

Clifton, K.J., and Livi, A.D., (2004), "Gender difference in walking behavior, attitudes about walking behavior and perceptions of the environment in three Maryland communities", Research on women's issues in transportation- conference report, Volume 2 technical papers, pp 79-88

Cochran W.G., (1977), "Sampling techniques", Second edition, John Wiley and Sons, New York, USA.

Delhi bus price DTB (2019):
http://delhitourism.gov.in/delhitourism/transport/city_bus.jsp

Delhi metro flow (2019): https://en.wikipedia.org/wiki/Delhi_Metro

Delhi metro flow data Business today (2019): https://www.businesstoday.in/current/economy-politics/delhi-metro-daily-ridership-sees-a-drop-amid-hike-in-ticket-prices/story/330857.html

Delhi metro flow data Economic times (2019): https://economictimes.indiatimes.com/industry/transportation/railways/crowded-delhi-metro-in-peak-hours-to-continue-no-new-coaches-as-lines-run-below-capacity/articleshow/47496979.cms?from=mdr

Delhi metro rail fares (2019): http://www.delhimetrorail.com/

Eddington, R. (2006), The Eddington transportation study, Vol 1 - Transport's role in sustaining the UK's productivity and competitiveness, UK Government

Feng, J.T., Lin, J.J., (2017), "School travel modes and children's spatial cognition", Urban Study. 54 (7), pp 1578–1600.

Ghimire, A., (2019), "Mode choice modeling for work trips in Kathmandu Valley", M.Sc Thesis, Department of Civil Engineering, Tribhuwan University – Institute of Engineering, Pulchowk, Lalitpur, Nepal.

Green, S.B., 1991, "How Many Subjects Does It Take To Do A Regression Analysis.", Multivariate Behavioral Research 26(3): pp 499–510.

Hensher, D.A., (1994), "Stated preference analysis of travel choices: the state of practice.", Transportation 21, pp. 107 – 133

Hensher, D.A., Rose, J.M., Greene, W.H., (2005), "Applied Choice Analysis: A Primer." , Cambridge university press, New York.

JICA. (2012)., "Data collection on traffic improvement in Kathmandu valley", JICA, Kathmandu

Joshi, M., (2019), "Mode choice modeling for Intercity travel in Nepal", M.Sc Thesis, Department of Civil Engineering, Tribhuwan University – Institute of Engineering, Pulchowk, Lalitpur, Nepal.

Kee-Le, J., Han-Song, K.E.Y.K., (2016), "A study on travelers' transport mode choice behavior using the mixed logit model: A case study of the Seoul-Jeju route." , Journal of Air Transport Management pp 1-7.

Kim, C., Choi, C.G., Cho S., & Kim, D., (2009), "A comparative study of aggregate and disaggregate gravity models using Seoul metropolitan subway trip data." , Transportation Planning and Technology, 32:1, pp 59-70.

Kish, L., (1965), "Survey Sampling", John Wiley and Sons, New York, USA.

Kolkata AC bus fares Times of India (2019): https://timesofindia.indiatimes.com/city/kolkata/kolkata-ac-bus-fare-revised-by-rs-5-per-2km/articleshow/64969171.cms

Kolkata bus fares Business standards (2019): https://www.business-standard.com/article/pti-stories/new-bus-taxi-fare-to-be-effective-from-june-11-118060801294_1.html

Kolkata bus fares West Bengal transportation corporation (2019): https://wbtc.co.in/

Kolkata metro flow Wikipedia (2019): https://en.wikipedia.org/wiki/Kolkata_Metro

Kolkata Metro rail fare (2019): http://www.kolkatametrorail.com/

Maa, L., Xiongb, H., Wangc, Z., Xieb, K., (March 2019), "Impact of weather conditions on middle school students' commute mode choices: Empirical findings from Beijing, China." , Transportation Research Part D: Transport and Environment, Volume 68, pp 39-51.

Manav Kendrit Yatayat Abhiyan (MaYA)., (2014 March), MaYA Fact sheet Number 4 – "Public transport in Kathmandu valley restructuring and reforming", Clean Air Network Nepal (CAAN) and Clean Energy Nepal (CEN), Kathmandu

Mannering, F.L., Washburn, S.S., (September 2016), "Principles of Highway Engineering and Traffic Analysis", Sixth Edition, John Wiley & Sons, Inc., New York, USA.

McDonald, N.C., Brown, A.L., Marchetti, L.M., Pedroso, M.S., (2009), "US school travel", American Journal of Preventive Medicine Volume 41, Issue 2, August 2011, Pages 146-151.

McFadden, D., Talvitie, A.P., and associates, (1997), "Urban Travel Demand Forecasting Project Phase 1 Final Report", Volume V, The institute of transportation studies, University of California, Berkley, California, USA.

Mclean, R.M., Shelby, M., Lula, C., Sagan, M., Train, K.E., (1992), "Comparison of forecasts from aggregate and disaggregate models for personal vehicle energy consumption and CO2 emissions" , Energy , Volume 17, Issue 4, April 1992, Pages 321-329.

MOEST. (2017)., "Education in a glance", Government of Nepal, Kathmandu

MOPIT. (2018), "Mass transit options and prioritization study for Kathmandu", MOPIT/ADB, Kathmandu

Mumbai AC bus fares Times of India (2019): https://timesofindia.indiatimes.com/city/mumbai/rs-5-minimum-best-bus-fare-in-a-few-days-with-final-mmrta-nod/articleshow/70067283.cms

Mumbai metro flow Times of India (2019): https://timesofindia.indiatimes.com/city/mumbai/mumbais-metro-ridership-touches-3-8-lakh-commuters-a-day-mark/articleshow/59654251.cms

Mumbai Metropolitan Region Development Authority Bus fare (2019): https://mmrda.maharashtra.gov.in/

NPC, (2020), "15th National Plan", National Planning commission. Kathmandu.

Phuoc, D.Q.N., Gyimahb, R.A., Trana, A.T.P., Phana, C.T., (2018), "Mode choice among university students to school in Danang, Vietnam." , Travel Behaviour and Society 13 pp 1–10.

Rai, B., (2012), "Estimation of work trip mode choice model for public transit Riders in Kahtmandu Valley" , M.Sc Thesis , Department of Civil Engineering , Tribhuwan University – Institute of Engineering , Pulchowk , Lalitpur , Nepal.

Reliance Mumbai metro fare Bus service fare (2019): https://www.reliancemumbaimetro.com/fares

Sanko, N., (2001), "Guideline for Stated Preference Experiment Design" , MBA Thesis , Ecole Nationale des Ponts et Chaussées , Champs-sur-Marne, France

Scheiner, J., Hubera, O., Lohmüllerb, S., (2019), "Children's mode choice for trips to primary school: a case study in German suburbia." , Travel Behavior and Society 15 pp15–27.

Singha, N., Vasudevanb, V., (2018), “Understanding school trip mode choice – The case of Kanpur (India).” , Journal of Transport Geography 66 pp 283–290.

The World Bank. (2013)., “Gender and Public transport”, World Bank, Kathmandu

Toriola, O. (2018), “A Multinomial Logistic Regression Analysis of the Association between Physical Activity and Body Weight Status of University Women in Riyadh, Saudi Arabia.” , Asian Journal of Scientific Research, 11: pp 145-150.

Train K.E., (1998), “Recreation demand models with taste differences over people”, Land Economics, 74(2) pp 230-239.

Zhang, R., Yao, E., Liu, Z., (2017), “School travel mode choice in Beijing, China.”, Journal of Transport Geography 62 pp 98–110.

Appendix A: Questionnaire

MODE CHOICE MODELLING FOR STUDY TRIPS OF GRADUATE LEVEL ENGINEERING STUDENTS OF KATHMANDU VALLEY

SOCIOECONOMIC CHARACTERISTICS

लिङ्ग : पुरुष ( ) महिला ( )

उमेर : १८ वर्षदेखि २२ वर्षसम्म ( ) २२ वर्षदेखि २६ वर्षसम्म ( ) २६ वर्षदेखि ३० वर्षसम्म( )

परिवारको मासिक आम्दानी (ने.रु.मा):(कृपया कुनै एकमा ☑ चिन्ह लगाउनुहोस् ।)

१५००० भन्दा कम ( ) १५००० - २५०००( ) २५००० - ३५०००( ) ३५००० - ४५००० ( ) ४५००० भन्दा बढी ( )

पारिवारिक शैक्षिक अवस्था

माता र पिना दुवै जना उच्च - माध्यामिक तह वा सो भन्दा उच्च अध्यन गर्नु भएको

माता वा पिता मध्ये एक जना उच्च - माध्यामिक तह वा सो भन्दा उच्च अध्यन गर्नु भएको :

माता र पिना दुवै जना उच्च - माध्यामिक तह सम्म अध्यन नगर्नुभएको :

परिवारको सदस्य संख्या(संयुक्त परिवार छ भने सबै सदस्यको संख्या लेख्नु होला):

दाजु/भाई-दिदि/वहिनीको संख्या (आफु सहित) :

TRIP CHARACTERISTICS

(कृपया तपाई अहिले बसाइ गरेको ठाउँबाट शिक्षण संस्था जाँदा लाग्ने तलका विवरणहरु भर्नुहोस् ।)

बसाईको ठेगाना : .... ***शिक्षण संस्थाको ठेगाना :*** ....

रुटको लम्वाई : .... कि.मि.

यात्राको किसिम : *सार्वजनिक सवारी साधन* [बस ( ) माइक्रोबस ( ) टेम्पो ( )]

निजी सवारी साधन [दुई पाङ्ग्रे ( ) चार पाङ्ग्रे ( )]

हिड्नु हुन्छ ()

*घर/डेरा देखि कलेज सम्मको यात्रा गर्नु हुने रुटमा निम्न तथ्यको मान राख्नुहोस्होस (**आफुले प्रयोग गर्ने माध्यम को मात्र नभई अन्य मध्यम को निम्ति पनि अनुमानित मान राख्नु होला**)*

निजी सवारी साधन प्रयोग गर्नेका लागि (एकतर्फि यात्राका लागि मात्र)
जम्मा यात्रा लागत: ने रु..............
जम्मा यात्रा समय: .... मिनेट

सार्वजनिक सवारी साधन प्रयोग गर्नेका लागि (एकतर्फि यात्राका लागि मात्र) ***[यदि हाल सार्वजनिक सवारी साधन चल्दैन भने यदि चल्यो भने कस्तो होला तेस्को वर्णन गर्नुस]***	
जम्मा यात्रा लागत:	*ने.रु..............*
हिड्ने समय:	*.............मिनेट (घर बाट सार्वजनिक सवारी साधनको स्टेशनसम्म)*
पर्खने समय:	*..........मिनेट(सार्वजनिक सवारी साधनको स्टेशनमा)*
सवारी साधनमा लाग्ने समय:	*...........मिनेट*
हिड्ने समय:	*.............मिनेट (सार्वजनिक सवारी साधनको स्टेशनबाट कलेजसम्म)*
जम्मा यात्रा समय:	*..........मिनेट*

पैदल यात्राका निम्ति
समय : .... मिनेर्ट

यदि हिड्नु हुन्छ वा हिड्नु पर्यो भने

१. हिड्नका निम्ति साथी को उपलब्धता : () भएको () नभएको
२. फुटपाथको उपलब्धता : () भएको () नभएको

परिवारमा गाडी(Vehicle ownership) : ()दुइ पाङ्ग्रे ()चार पाङ्ग्रे ()नभएको

यस प्रकारको सवारी साधन पाउनु भयो भने तपाइको रोजाई फेरिन्छ कि?? फेरिन्छ भने कुनै एक मा tick लागाउनुस

बस २.५ गुणा भाडा आरामदायक र कम समय लाग्ने 	बस त्यति नै भाडा र आरामदायक तर कम समय लाग्ने 
मेट्रो रेल १.५ गुणा भाडा धेरै बढी आरामदायक कम समय लाग्ने 	मेट्रो रेल त्यति नै भाडा र आरामदायक तर कम समय लाग्ने 

Appendix B: R Coding

```
mydataRP<-read.csv("C:/Users/DELL/Desktop/BookRP.csv")
mydataBus1<-read.csv("C:/Users/DELL/Desktop/BookBus1.csv")
mydataBus2<-read.csv("C:/Users/DELL/Desktop/BookBus2.csv")
mydataMetro1<-read.csv("C:/Users/DELL/Desktop/BookMetro1.csv")
mydataMetro2<-read.csv("C:/Users/DELL/Desktop/BookMetro2.csv")
attach(mydataRP)
attach(mydataBus1)
attach(mydataBus2)
attach(mydataMetro1)
attach(mydataMetro2)
library(mlogit)
library(ggplot2)

correl=cor(mydataRP[,-1])
symnum(correl)
#RP Models


#Base model with all parameters
mldataRP<-mlogit.data(mydataRP, varying=9:14, choice="Choice", shape="wide")
clogit.modelRP <- mlogit(Choice ~
Price+JT|AgeGroup+Gender+Trip.Distance+Siblings+Sizefamily+Family.education
.level+Family.Income+Veh.Owner+Avail.Friend+Avail.Foothpath,rpar=c("Price"='
n'), data = mldataRP, reflevel="Walk")
summary(clogit.modelRP)


#Restricted model with no parameters
mldataRP<-mlogit.data(mydataRP, varying=9:14, choice="Choice", shape="wide")
clogit.modelRP <- mlogit(Choice ~ 1, data = mldataRP, reflevel="Walk")
summary(clogit.modelRP)
```

#Modified model First Try

mldataRP<-mlogit.data(mydataRP, varying=9:14, choice="Choice", shape="wide")

clogit.modelRP <- mlogit(Choice ~ Price+JT|Gender+Trip.Distance+Siblings+Veh.Owner+Family.education.level+Avail.Foothpath,rpar=c("Price"='n'), data = mldataRP, reflevel="Walk")

summary(clogit.modelRP)

#Modified model Final

mldataRP<-mlogit.data(mydataRP, varying=9:14, choice="Choice", shape="wide")

clogit.modelRP <- mlogit(Choice ~ Price+JT|Gender+Trip.Distance+Siblings+Family.Income+Veh.Owner+Avail.Friend,rpar=c("Price"='n'), data = mldataRP, reflevel="Walk")

summary(clogit.modelRP)

#SP models

#Bus1 Full Model

mldataBus1<-mlogit.data(mydataBus1, varying=9:16, choice="Choice", shape="wide")

clogit.modelBus1 <- mlogit(Choice ~ Price+JT|AgeGroup+Gender+Trip.Distance+Siblings+Sizefamily+Family.education.level+Family.Income+Veh.Owner+Avail.Friend+Avail.Foothpath,rpar=c("Price"='n'), data = mldataBus1, reflevel="Walk")

summary(clogit.modelBus1)

#Bus1 Final

mldataBus1<-mlogit.data(mydataBus1, varying=9:16, choice="Choice", shape="wide")

clogit.modelBus1 <- mlogit(Choice ~ Price+JT|Gender+Trip.Distance+Siblings+Family.Income+Veh.Owner+Avail.Foothpath,rpar=c("Price"='n'), data = mldataBus1, reflevel="Walk")

summary(clogit.modelBus1)

```
# Bus2 Final

mldataBus2<-mlogit.data(mydataBus2, varying=9:16, choice="Choice", shape="wide")

clogit.modelBus2 <- mlogit(Choice ~ Price+JT|Gender+Trip.Distance+Siblings+Family.Income+Veh.Owner+Avail.Foothpath,rpar=c("Price"='n'), data = mldataBus2, reflevel="Walk")

summary(clogit.modelBus2)


# Metro1 Final

mldataMetro1<-mlogit.data(mydataMetro1, varying=9:16, choice="Choice", shape="wide")

clogit.modelMetro1 <- mlogit(Choice ~ Price+JT|Gender+Trip.Distance+Siblings+Family.Income+Veh.Owner+Avail.Foothpath,rpar=c("Price"='n'), data = mldataMetro1, reflevel="Walk")

summary(clogit.modelMetro1)


# Metro2 Final

mldataMetro2<-mlogit.data(mydataMetro2, varying=9:16, choice="Choice", shape="wide")

clogit.modelMetro2 <- mlogit(Choice ~ Price+JT|Gender+Trip.Distance+Siblings+Family.Income+Veh.Owner+Avail.Foothpath,rpar=c("Price"='n'), data = mldataMetro2, reflevel="Walk")

summary(clogit.modelMetro2)
```

Appendix C: Sample Survey Data

Training Data

Choice	Gender	Age Group	Family Income	Family education level	Size of family	Siblings	Trip Distance	Journey Time Public	Journey Time Walk	Journey Time Personal	Price.Public	Price.Walk	Price.Personal	Availability. Friend	Avail.Foothpath	Vehicle.Ownership	Alternative	Institution Type
Walk	1	1	4	2	5	3	0.4	6	5	2	10	0	5	1	1	1	Metro2	Private
Walk	1	1	1	3	14	3	0.42	6	5	2	10	0	5	1	1	1	Metro2	Private
Personal	1	3	3	2	4	2	2	8	8	3	10	0	7	2	1	1	NA	Private
Personal	1	2	3	2	4	2	4	30	45	20	15	0	20	2	2	1	NA	Private
Personal	1	1	3	3	4	2	11	59	100	30	20	0	40	1	1	1	Metro1	Private
Public	1	1	5	2	4	2	8.5	50	70	25	15	0	40	1	1	1	Metro2	Private
Public	1	1	5	1	4	2	3	40	35	15	10	0	15	2	1	1	Metro1	Private
Public	1	1	3	2	4	2	2	26	20	10	10	0	15	1	1	2	NA	Private
Personal	2	1	3	2	4	2	4	42	45	15	10	0	20	1	2	1	Metro1	Private
Public	1	1	3	1	6	4	8	80	90	30	15	0	50	1	1	1	Metro1	Private
Public	1	1	3	2	4	2	8	55	95	26	15	0	45	1	1	2	Metro1	Private
Personal	1	1	4	1	4	2	7	50	60	15	20	0	22	2	2	1	Metro2	Private
Public	1	1	5	3	5	3	6	40	85	20	10	0	15	2	2	1	Metro1	Private
Public	2	1	5	2	5	3	5.6	51	63	25	10	0	30	1	1	1	Metro2	Private
Walk	2	1	4	1	6	4	1.8	17	25	5	10	0	10	1	1	1	Metro2	Private
Public	1	1	5	2	4	2	4.5	45	50	20	10	0	30	1	1	2	NA	Private
Personal	1	2	5	2	4	2	10	115	150	20	30	0	60	2	1	1	Metro1	Private
Walk	1	2	3	2	4	2	1.2	20	18	5	10	0	20	2	1	1	Bus2	Private
Personal	1	1	5	3	4	2	4	25	35	5	10	0	15	2	1	2	Metro2	Private
Personal	1	2	2	3	5	3	2	20	28	7	10	0	10	1	1	1	Bus2	Private
Public	1	1	3	2	4	2	8	20	45	50	20	0	35	1	1	2	Metro2	Private
Walk	1	2	4	1	4	2	1	16	15	5	10	0	7	1	1	1	Metro1	Private
Public	1	1	2	2	4	2	2	20	20	5	10	0	10	1	2	2	Metro2	Private
Walk	1	1	2	1	4	2	2.5	35	25	10	10	0	15	1	1	1	Bus2	Private
Personal	1	1	2	1	5	3	7	55	60	20	15	0	20	1	1	1	Metro1	Private
Personal	2	1	3	2	5	3	10	62	110	45	20	0	20	2	2	1	Metro2	Private
Walk	1	1	3	2	5	3	7	60	65	30	20	0	40	2	1	1	Metro2	Private
Walk	1	1	3	3	4	1	3	50	35	20	10	0	10	2	1	2	Metro2	Private
Walk	1	1	2	3	5	4	5	25	100	30	15	0	20	2	2	1	Metro2	Private
Walk	1	1	2	2	5	3	5	90	35	30	20	0	20	2	1	1	Metro2	Private

Validation Data																		
Choice	Gender	Age Group	Family Income	Family education level	Size of family	Siblings	Trip Distance	Journey Time Public	Journey Time Walk	Journey Time Personal	Price.Public	Price.Walk	Price.Personal	Availability.Friend	Avail.Foothpath	Vehicle.Ownership	Alternative	Institution Type
Public	2	1	4	2	4	2	8	29	100	15	10	0	40	1	1	2	Metro2	Institutional
Public	2	1	5	2	6	4	2.8	17	25	6	10	0	22.5	1	1	1	Metro2	Institutional
Public	1	1	4	3	4	2	8	50	65	30	30	0	45	2	1	1	Metro2	Institutional
Public	2	1	4	1	4	1	12	110	100	40	30	0	80	1	1	2	Metro2	Institutional
Public	1	1	3	3	6	4	8	55	75	24	20	0	35	2	2	1	Metro2	Institutional
Walk	1	2	4	1	5	3	2	15	30	5	10	0	10	2	1	1	Metro2	Institutional
Walk	1	1	5	1	3	1	1	5	15	5	10	0	10	2	1	2	Metro2	Institutional
Walk	2	1	5	2	5	3	2	19	25	25	10	0	10	1	1	2	Metro2	Institutional
Walk	1	2	3	3	8	2	1	5	15	2	10	0	5	1	2	1	Bus1	Institutional
Walk	1	1	2	3	5	2	3	31	35	15	10	0	10	2	2	2	Metro2	Institutional
Personal	1	2	5	2	5	3	16	48	100	30	15	0	80	1	1	1	Metro2	Institutional
Public	2	1	5	1	6	3	5	35	45	10	10	0	30	2	1	1	Metro2	Institutional
Walk	2	1	4	1	4	3	3	55	90	40	20	0	40	2	1	1	Bus1	Institutional
Personal	1	1	5	1	4	2	7.5	32	50	20	30	0	40	1	1	1	Metro2	Institutional
Public	1	1	5	1	6	4	16	165	180	30	25	0	80	1	1	2	Metro2	Institutional
Personal	1	1	5	1	5	3	4	35	40	5	15	0	15	1	1	1	Metro2	Institutional
Walk	2	1	4	1	5	3	1.5	9	20	5	10	0	8	2	1	1	Metro2	Institutional
Walk	1	1	4	1	6	4	1	15	15	15	10	0	2	1	1	1	Metro2	Institutional
Personal	1	1	3	3	7	3	5	56	40	8	10	0	5	2	1	1	Metro2	Institutional
Walk	1	1	3	1	8	5	1	6	5	3	10	0	7	1	1	1	Metro1	Institutional
Public	1	1	2	2	4	2	15	82	150	40	50	0	70	1	1	1	Metro2	Institutional
Personal	1	1	5	2	5	3	4	95	60	10	60	0	5	1	2	1	Metro2	Private
Public	1	1	3	3	5	2	5	21	30	5	10	0	10	2	1	1	Metro1	Private
Personal	1	1	5	1	7	5	12	52	100	30	20	0	15	2	1	1	Metro2	Private
Walk	1	1	5	3	6	4	1	7	5	5	10	0	10	2	1	1	Metro2	Private
Walk	2	1	2	1	5	3	0.2	17	5	3	10	0	5	1	1	1	Metro2	Private
Walk	1	1	4	1	4	2	0.2	12	5	2	10	0	5	1	1	1	Bus1	Private
Public	2	1	2	1	4	2	11	70	110	20	15	0	40	2	1	1	Metro2	Private
Walk	2	1	4	2	5	3	3	35	35	20	10	0	15	2	1	2	Metro1	Private
Public	2	1	5	2	5	3	3	35	30	10	15	0	20	1	1	2	Metro1	Private