Modeling Transport by GIS:
A Case Study of Famagusta

Sina Darban Khales

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the Degree of

Master of Science
in
Civil Engineering

Eastern Mediterranean University
July 2013
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Elvan Yılmaz
Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Master of Science in Civil Engineering.

Asst. Prof. Dr. Mürüde Çelikağ
Chair, Department of Civil Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Civil Engineering.

Asst. Prof. Dr. Mehmet Metin Kunt
Supervisor

Examining Committee

1. Asst. Prof. Dr. Alireza Rezaei
2. Asst. Prof. Dr. Mehmet Metin Kunt
3. Asst. Prof. Dr. Mustafa Ergil

ABSTRACT

One of the most important issues in travel demand management is assessing different ways of transport and balance travel demands in urban areas. Suppose that five different modes of transportation are available for travellers in a city: Bus, taxi, bike, private car and walking. Motivating people to use non-motorized vehicles for their short-distance travels and public transport for the long-distance ones is one of the biggest concerns of transport planners.

Due to the key role of public transportation in policy making, Travel modal choice is one of the most important stages in transportation planning. It is obvious that using public transportation contributes to more efficient road space usage and fewer accident production and emission compared to using private cars. Besides, if more drivers would be attracted to use public transportation others can benefit from higher level of service.

Finding the most efficient way of using different transit modes is also as much important as modal choice itself. Optimal travel routes, mostly based on travel time are always more preferable to be used.

Most of the previous researches in transit modal choice are based on estimated or constant values for parameters of the model to decrease the complexity. In this study, we first try to find the best route through a multimodal network of five different modes of transport, supported by a geographic information system (GIS) for the city of Famagusta in North Cyprus and later, find the exact values of each parameter to establish more realistic travel choice models based on socio-economic characteristics of individuals and urban transport attributes for any specific route.

Keywords: Transportation modal choice model, network analysis, logit model

ÖZ

Seyahat talep yönteminde önemli konulardan biri farklı ulaşım araçlarının kullanımının değerlendirilmesi ve şehir içi seyahatin dengelenmesidir. Seyahat için bir kentte otobüs, taksi, özel araç, bisiklet ve yürüme modlarının olduğunu varsayabiliriz. Ulaşım planlayıcılarının en büyük uğraşısı bu varsayılan seyahat modlarından insanların motorsuz modları kısa mesafeler ve otobüsleri de uzun mesafeler için seçmeleri konusunda motive etmektir. Seyahat modu seçimi toplu taşımacılık politikası oluşumunda kilit rol oynadığı için ulaşım planlamasında önemli bir yeri vardır. Toplu taşımacılığın kullanımı, yol ağının daha verimli kullanımına ve özel araçların daha az emisyon salınımına ve kaza olasılığına yol açacaktır.

Bu faydalara ilaveten, özel araç kullananların toplu taşımacılığı tercih etmeleri yol ağındaki hizmet seviyesini de yükseltecektir.

Farklı seyahat modlarını en verimli şekilde seçmek seyahat modunu seçmek kadar önemlidir. Seyahat zamanına bağlı olan optimum seyahat güzergahı tercih edilmelidir.

Geçmişteki araştırmalarda seyahat modu seçim modellerini basitleştirmek için model parametrelerinin değerleri tahmin edilmiş veya sabit değerler kullanılmıştır. Bu çalışmada en iyi güzergah tesbiti beş farklı seyahat modu kullanılarak yapıldı. Bu çalışmada kullanılan model Kuzey Kıbrıs'ta bulunan Gazimagosa kentinin coğrafi bilgi sisteminde oluşturulan yol ağında uygulanmıştır. Çalışmanın sonunda her parametrenin değeri tesbit edilmiştir. Bu sayede, seyahat edenlerin sosyal-ekonomik karakteristikleri ile şehir içi seyahat şartlarına uygun gerçekçi seyahat modu seçimi herhangi bir güzergah için hesaplanabilir.

Anahtar Kelimeler: Ulaşım modal seçim modeli, ağ analizi, logit modeli

I would like to dedicate my thesis to my beloved family

For their endless love, support and encouragement

ACKOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisor Asst. Prof. Dr. Mehmet Metin Kunt for the continuous support of my master study and research, for his patience, motivation, enthusiasm, and immense and generous contribution of knowledge. I would like to convey my warmest gratitude to Dr. Alireza Mansourian, for his guidance, valuable comments and encouragement. I take this opportunity to express my appreciation to my friends in and outside the Civil Engineering Department, Behdad Jafari, Amirkhosro Karimi, Marzieh Parsa and Farzan Morshedian who provided me the possibility to complete this master thesis. Last but not the least important, I owe more than thanks to my family for their support and encouragement throughout my life. Without their support, it was impossible for me to finish my graduate education seamlessly.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 General Introduction

In recent years accurate estimation of the allotment of each transit mode in daily travels is one of the biggest concerns of transportation planners. To accomplish this, various models have been offered, attempting to replicate the system of interest and its behavior by means of mathematical equations based on certain theoretical statements. Most of these transit modal equations are based on utility functions, showing the interest of travellers to a specific alternative based on its costs and benefits. One of the main deficiencies of the models being presented is that, most of them have focused on making predictions to change in the existing available modes or appeal new ones for future matters by generally comparing the probability of choosing transport means with constant or approximate parameters. To better understand the problem, suppose a probability model is presented based on travel time and travel expenditure parameters. Normally, to obtain the probability of choosing one alternative to decide about future modification and enhancement in some modes and attract more travellers one simply needs to put the values for time and cost into the equations, directly getting the results and decides about the available solutions in accordance to them in the next step. But this method is highly susceptible to the following errors:

1. As a result of traffic and other factors the accurate travel time is not available for different times of day and for any desired route;
2. Cumulative travel cost between any origin and destination is unknown
3. Imperfect information about the optimal route between two points may cause bias in obtaining the percentage of probability to choose among transit modes.

The above problems together with the particular taste of each person and, the measurements and observation errors made by the modeler might result in incorrect model generation and consequently wrong decision making.

1.2 Research Objectives

The main aim of this study is to develop an Advanced Traveller Information System (ATIS) for the city of Famagusta in North Cyprus as a sample, and build a network dataset for this study area, using ArcGIS10 by ESRI. This ATIS helps to present an inclusive pack of information for all travellers who intend to travel through this network including travel time, travel expenditure and cumulative travel distance for the best route between any origin and destination base on any desired impedance. Using ‘any desired impedance’ means that if the impedance is time, then the best route is the quickest route and if it is length, the best route is the shortest one. Hence, the best route can be defined as the route that has the lowest impedance, or least cost, where the impedance is chosen by the modeler. Any cost attribute can be used as the impedance when determining the best route and as more accurate data is being used the result will be more realistic. For instance if the live or historical traffic data is used in a network, the best route can be calculated for any date and any time of day.

To extract the required information from any travel route some programming is needed to be done. In this research Python, Mathematica and Visual Basic programming languages together with the Structured Query Language (SQL) are used for the algorithm implementation.

The results of this part of study are used after, to generate the travel modal choice model. With the advantage of working with this comprehensive set of data, the accomplished model is more close to the real world compared to any other previously presented models.

1.3 Tasks

The major and specific tasks of this study are:

1. Investigate the transport modal choice models structure and find the most suitable one for Famagusta
2. Prepare the digitize map of Famagusta in ArcMap 10.0;
3. Do the required data collection and create a multimodal transportation network for the study area;
4. Find the values for all available transit modes attributes;
5. Do a questionnaire for 200 individuals in Famagusta;
6. Analyze the results and finally find the appropriate model for the city.

Chapter 2

LITRETURE REVIEW

2.1 Introduction

Our world, including transport is changing rapidly but we still confront the perpetual problems of traffic congestion, air pollution, collision and travel inconvenience. Countries are going through industrialization and economic development increasingly, more people are emigrating from rural areas to big cities and people become more money rich and time gradually turns into the major problem in different societies. That is the time when, old problems with even bigger magnitude reemerge and the great need for handling these new and complicated problems emanates and that is the beginning of transport modeling.

2.2 Transport Problems

The widespread, rapidly increasing problems of transportation in recent years have been beyond the predictions, incompatibility in resources and demands is among the main difficulties that transport planners face in both developing and industrialized countries and since, the number of travellers is increasingly exceeding the maximum capacity of infrastructures and transit facilities more severe problems have come into view.

The above problems will not disappear without making suitable short, middle and long term policies. These policies should be in accordance to realistic models based on reliable data to maximize the advantages of current and future transportation provision and minimize the negative side effects and money cost.

2.3 Models and Their Roles

A model is a graphical, mathematical (symbolic), physical, or verbal representation or simplified version of a concept or the real world –the system of interest–which focuses on specific aspects of important elements from a certain point of view [1]. Models are generally divided into two main groups:

1. Physical models
2. Abstract models

Physical model refers to a smaller version of an object while, abstract or conceptual model is the one which deals with people's minds. The role of mental models is to help us interpreting and understanding our world and analytical models. One of the most important classes of conceptual models is mathematical models. The aim of this class is to picture the desired part of our world by means of mathematical formulas and parameters based on theoretical statements and available data. Different transportation models like transport modal choice models belong to this group.

2.4 Modeling and Decision Making

It is obvious that if one model is proper for a place it is not necessarily suitable to be applied for another. So first, it is needed to identify which model approach adopts in the study area and decision unit. The optimistic view is that, knowing all about the objectives, every alternative methods of achieving them can be envisaged with their costs and benefits. This way of decision making is known as rational decision making, sometimes referred to as the ‘systems approach’ to planning. The aim is to choose between an inclusive set of options and scenarios by calculating the probability of occurrence for each. To do so, the term ‘Utility’ is defined to quantify each alternative based on its costs and benefits and other factors like convenience, safety and so on but this approach also has its own limitations.

2.5 Choosing Modeling Approaches

It is further explained that there is a great number of transport difficulties and models which should be taken into consideration while defining an analytical approach:

1. Accuracy and precision at the same time. Since, these two concepts are wrongly used identically very often, it is better to define them first in this part. Accuracy is defined as, the ability of a measurement to match the actual value of the quantity being measured, while precision is defined as, the ability of a measurement to be consistently reproduced or the number of significant digits to which a value has been reliably measured [2].
2. The decision-making context: The answer to the question of how many alternatives should be considered to cover different tastes and satisfy the needs of our statistical population. Besides, the decision-making context will also be helpful to express requirements for the model generation and implementing suitable variables.
3. Level of detail required: Can be defined along four main categories: a) geography b) unit of analysis c) behavioral responses and d) the handling of time. The first two are explicit and doesn’t need to be explained so only the third and the last one are described in this part.

The behavioral responses may vary from fairly simple route choice actions in a traffic model to changes in time of travel, mode, destination, tour frequency or even land use and economic activity impacts.

Time can be taken as both continuous or a discrete variable. The main advantage of treating time as a continuous parameter over taking it as a discrete variable is that it enables building more dynamic models [3].

4. Availability of reliable data: In many cases lack of enough suitable data may cause serious modeling problems
5. Resources availability including computer software in addition to technical skills.

2.6 Transport Modeling Issues

As the relationship between transportation problems, different decision making methods and model planning were discussed before, now, it is time to talk about the critical modeling concerns such as preparing the appropriate choice set, model specification, model calibration and finally adjust it to be applied practically in the real world decision making and planning.

Mode choice and carrier selection are part of the decision-making process in transportation that includes identifying relevant transportation performance variables, selecting mode of transport and carrier, negotiating rates and service levels, and evaluating carrier performance [4].

In recent years, various investigations have been made about transport modal choice models in different societies. Inclusive transport planning is one of the most important parts of urban management decision making. The principal goal of these studies is to design short-term (2 to 5 year), middle-term (10 year) and long-term (15 to 20 year) transportation strategies. These kinds of planning have been used or are being developed in almost all civilized societies for decades. The methodologies should reflect the complexity of the travel behavior, the range of factors that impact on the choice process, the interaction between variables during decision making and the variability due to the diversity of travellers making these decisions themselves. By considering these elements conjointly, and not singularly, the choice process can be modeled mathematically [5].

The main problem in generating transport modal choice models in inclusive transportation planning is neglecting the importance of having a comprehensive set of data. This lack of enough information causes wrong prediction of the traveller's behavior and consequently distributes to scheme wrong methods to modify or improve the current transport conditions. Some examples of these kinds of wrong transport planning results can be found in India and Iran. The failure in the monorail project in Tehran after assigning lots of budgets and time due to the wrong decision making happened by implementing insufficient data can be grouped among them.

Probability models research studies in transportation have been started seriously in the late 1960's. In the first years most of the researches were based on binomial choice models. Hence, the principal models were too much simple and without the necessary complexities. The most prominent researches about the principal binomial models were Warner and Shuldiner studies in 1962 [6].

The first advance multinomial transport modal choice studies were held by Spear in 1977. He expressed the socio-economic characteristics of individuals as the most important element in the multinomial models [7]. In 1980, Ort´uzar represented the Logit models implementation in the transport modal choice area for the first time and started a new chapter in transportation modeling [8]. Probability research studies then, reached to its peak by Mac Fadden's efforts in 1981 and Ben Akiva and Lerman in 1985 and different probability models were investigated based on discrete choice models in the years after [9], [10].

After the entry of probability models and multinomial choice methods in transportation, the studies progressed more rapidly and the modeling transport book by Ort´uzar, which its third edition in 2003 is still one of the main resources in modeling studies, was published at that time. Moreover, Green's research in 2005 about the influence of traveller's behavior on urban traffic accomplished all the previously achieved models [11]. All the important academic works in this field are briefly represented in Table 1.

Table1. Important Academic Works in Modeling Transport

Researcher	Subject of research	Research Type	Model
Warner	Binary Choice Models	Book	Multinomial Choice
Wiliams	Choice Models	Book	
Mac fadden	Logit Models Application in Multinomial Choices	Paper	
Spear	Vehicle Choice Models	Book	
Ben Akiva	Demand-Prediction Models	Phd. Thesis	
Greene	Structure of Nonlinear Models	Paper	Discrete Choice Models
Ortuzar	Transport Modeling	Book	
Ben Akiva	Mixed Logit Models Application in Transportation	Paper	
Mac fadden	Multinomial Logit Models Application in Discrete Choice Models	Paper	
Sakamaki	Investigate the Discrete Choice models Application to balance the Transportation Needs In Helsinki	Phd. Thesis	
Hartman	Investigate the Effect of Active Modifications on the Individuals Behaviour	Paper (2011)	Mutinomial Logit Model
Lamendia	Methodology of setting the amount of entertainment trips for intercity travels	Paper (2011)	Logit Model
Senere	Find the Suitable Place for Residentual Construnctions in Accordance to the Individuals Behavior	Paper (2010)	GSCL

As it can be seen, almost in all inclusive transport plans the discrete choice models have been generated and according to the gained results travel allotment and the final results have been investigated.

Chapter 3

RESEARCH METHOD

3.1 Introduction

The goal of this study is to generate a realistic model for the city of Famagusta. Famagusta is a small city on the eastern coast of Cyprus, the bounding geography coordinates of the city is 35°07′30″ North Latitudes and 33°57′00″ East Longitudes and it is the de facto capital of Gazimagusta District of the Turkish Republic of Northern Cyprus. The biggest university of the country, Eastern Mediterranean University, is also located in Famagusta which makes it more a student city by considering the 80% student population of it, according to the last census on 2012.

Like many other small cities, Famagusta suffers from poor public transport infrastructures. Bus is the only public mean of transport available for people in Famagusta and although it is free of charge, it has its own serious problems which low service frequency, unreliable timing, insufficient bus routes and stations and crowdedness during peak hours which all provide discomfort for the travellers are among them. Taxi mode is also too expensive to be used regularly by students and most of the international students face problems with the taxi drivers due to their poor communication skills. Bicycle is another mean of transport which is widely used by students in Famagusta for the short-distance travel purposes, but lack of consideration for bicycles in the physical design of infrastructures and facilities in addition to safety problems and inadequate exclusive bicycle parking spaces create some difficulties for people, using this non-motorized vehicle.

All the above problems with different travel modes cause people to have a great preference to use their private cars, if available, for their daily travels.

3.2 Discrete Choice Models

In recent years, most of the transport modal choice systems have been generated based on the discrete choice theory which postulates that the probability of individuals choosing a given option is a function of their socioeconomic characteristics and the relative attractiveness of the options.

To build the model, the first step is to prepare the appropriate choice set containing all the available options for the travellers. Three requirements should be met in the choice set for a discrete choice model [12]:

1. The set of alternatives must be exhaustive, meaning that all possible alternatives must be in our choice set and it is obligatory for any person to select from that specific set.
2. The alternatives must be mutually exclusive. This condition indicates that multi selection is prohibited and allows each person to choose only one alternative from the choice set.
3. Unlike the regression analysis parameters, variables in discrete choice analysis can only take a finite number of values which means that the choice set for a discrete choice model cannot include an infinite number of options.

To represent the attractiveness of the alternatives, the concept of utility (which is a convenient theoretical construct defined as what the individual seeks to maximize) is used. The main premise is that, the behavior of the travellers is utility maximizing.

3.3 Utility

'Utility' is a concept representing the degree of attractiveness of an option over a set of alternatives [13]. In other words, U_{ij} it is the well-being or benefits that person i obtained from selecting alternative j based on the hypothesis that person i chooses the alternative with the highest utility number. In our case, utility is formulated base on different characteristics of the alternatives and individuals and the relative effect of each attribute, contributing to the overall satisfaction by choosing an alternative is represented by its coefficients.

If we have a predefined set $B = \{B_1, . . . , B_j, . . . ,B_n\}$ of available options and a set V of vectors of measurable characteristics of alternatives and individuals, each individual i faces a specific set of attributes $v \in V$ and each option $B_j \in B$ associates a net utility U_{ij} for individual i. Obviously, it is impossible to say a modeler has full knowledge about all the elements considered by the individuals while selecting an option therefore, it is presumed that each utility function has two components: the first one can be observed by the researcher, it means that it can be formulated based on the observable parameters of the option, and the second one cannot be seen by the modeler and includes both the particular taste of each person and the errors made by the researcher.

By decomposing the utility function into these two parts to mathematically use it for modeling purposes the formula below is obtained:

$$U_{ij} = V_{ij} + \mathcal{E}_{ij} \tag{3.1}$$

where:

U_{ij}: Utility of choosing alternative *j* by individual *i*

V_{ij}: Measurable aspect of the alternative which is a function of its measured attributes

and

ε_{ij}: Unobservable part or random error

Which allows two obvious 'illogicalness' to be expressed: that two individuals facing an identical choice set, with similar attributes do not necessarily select the same option and that some individuals may opt for some alternatives which are not the best in the view of the researcher. These two facts clearly show the important role of the random unobservable part in setting up a realistic model.

3.4 The Unobservable Component of Utility ε

Generally, residual ε is the most important part of the utility function and as long as there is no esteemed value for it, no model can be defined for choosing between transit modes. The only thing known from this part of the function is that, they are random variables with mean zero and a probability distribution, $f(\varepsilon) = f(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_3)$ to be found, and the chance of selecting each product defined accordingly.

Mathematically, the utility function $f(u)$ exactly has the same distribution as well, because *V* is the measurable part of the utility function with a true value, which does not affect the distribution.

As said before, travellers are mostly interested in alternatives with higher utility number. Means that, one chooses alternative *j* only when $u_j > \{u_{j'}\}_{\forall j' \neq j}$.

The main goal is to find the probability of choosing alternative *j* by individual *i* therefore, the value of the utilities should be transformed into a probability number between 0 and 1. To do so the answer of the following equation has to be found:

$$P_{ij} = Pr\left[u_{ij} > \{u_{j'}\}_{\forall j' \neq j}\right] \tag{3.2}$$

$$= Pr\left[v_{ij} + \varepsilon_{ij} > \left\{v_{ij'} + \varepsilon_{ij'}\right\}_{\forall j' \neq j}\right] \tag{3.3}$$

But first, two concepts of CDF and PDF are briefly defined in this part.

3.4.1 PDF and CDF

Generally, random variables can be grouped into these two categories:

1. Discrete: involves discrete or countable range of real numbers
2. Continuous: involves continuous or uncountable range of real numbers

If $P(x)$ represents the probability function of a discrete random variable, the cumulative distribution function of that variable is given by:

$$F_X(x) = P(X \leq x) \tag{3.4}$$

and while, the CDF of a continuous random variable X cannot be expressed algebraically, it can be expressed analytically as the integral of its probability density function f_x as follows:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \tag{3.5}$$

To solve the Eqn. (3.2) the cumulative distribution function (CDF) of the vector of residual errors, $\varepsilon_i = [\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{in}]^T$ that is $F_\varepsilon(\varepsilon)$, together with its probability density function (PDF) $f_\varepsilon(\varepsilon)$ are used.

If the equation is needed to be solved for two alternatives:

$$P_1 = Pr[v_1 + \varepsilon_1 > v_2 + \varepsilon_2] \tag{3.6.1}$$

$$= Pr[\,\varepsilon_2 < v_1 - v_2 + \varepsilon_1] \tag{3.6.2}$$

By comparing Eqn. (3.6.2) with Eqn. (3.4) it can easily be understood that the above equation shows the cumulative distribution function of ε_2, calculated at the point$(v_1 - v_2 + \varepsilon_1)$. Therefore according to the definition of CDF:

$$P_{1=} \int_{\varepsilon=-\infty}^{\infty} \left(\int_{\varepsilon_2=-\infty}^{v_1 - v_2 + \varepsilon_1} f_\varepsilon(\varepsilon_1, \varepsilon_2) d_{\varepsilon 2} \right) d\varepsilon_1 \tag{3.7}$$

So if the Eqn. (3.7) is generalized for a set of alternatives:

$$P_j = Pr[v_j + \varepsilon_j > \{v_k + \varepsilon_k\}_{\forall k \neq j} \quad (3.8.1)$$

$$= Pr[\{\varepsilon_k < v_j - v_k + \varepsilon_j\}_{\forall k \neq j} \quad (3.8.2)$$

$$= \int_{\varepsilon_1}^{\infty} (\int_{\varepsilon_1=-\infty}^{v_j+v_1+\varepsilon_j} \int_{\varepsilon_2=-\infty}^{v_j+v_2+\varepsilon_j} \dots \int_{\varepsilon_j=-\infty}^{v_j+v_j+\varepsilon_j} f_{\varepsilon}(\varepsilon) d\varepsilon) d\varepsilon_j \quad (3.9)$$

$$= \int_{\varepsilon_j=-\infty}^{\infty} f_{\varepsilon_j}(\varepsilon_j) (\prod_{k \neq j} (\int_{\varepsilon_k=-\infty}^{v_j-v_k+\varepsilon_j} f_{\varepsilon_k}(\varepsilon_k) d\varepsilon_k)) \, d\varepsilon_j$$

$$= \int_{\varepsilon_j=-\infty}^{\infty} f_{\varepsilon_j}(\varepsilon_j) \ (\prod_{k \neq j} F_{\varepsilon_j}(v_j - v_k + \varepsilon_j)) d\varepsilon_j \quad (3.10)$$

where:

$$d\varepsilon = d\varepsilon_j \dots d\varepsilon_{j+1} \, d\varepsilon_{j-1} \dots d\varepsilon_2 d\varepsilon_1$$

3.4 Discrete Choice Models Classification

Discrete choice models could be first clustered into these two main groups based on the number of available options:

1. Binomial models : in this case there are two options to choose from
2. Multinomial models : in this case there are 3 or more options to choose from

Multinomial models can also be classified based on their feature as below:

- Models, that allow correlation between unobserved part of the utility functions between alternatives

- Models that do not allow any correlation between the unobserved sections among alternatives [14].

The main goal in this study is to model the choice probability between more than two options with no correlation in the error terms. Therefore, the model has to be set up by considering the above conditions. Further in this study, the most commonly used methods are inspected to find the finest model for Famagusta.

After finding the relation between CDF and comparing the utility functions, to forecast which option will be selected by the travellers, the most suitable probability function will be selected for this research in the next step.

For this, two well-known probability functions are more common than the others:

1. Logit

$$P_1 = \frac{e^{V_1}}{\sum_k e^{V_k}}$$

and

2. Probit

$$P_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{V_1+V_2+x_1} \frac{exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1}{\sigma_1}\right) - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \left(\frac{x^2}{\sigma^2}\right)^2\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} dx_2 dx_1$$

The important common feature between these two models is that they both have an S-shaped design.

3.5.1 Similarities and Differences

Both logit and probit are sigmoid functions ranging between 0 and 1, that makes both of them quantile functions. In fact, the logit is the quantile function of the logistic or Gumbel distribution (Figure 1), while the probit is the quantile function of the normal distribution [15].

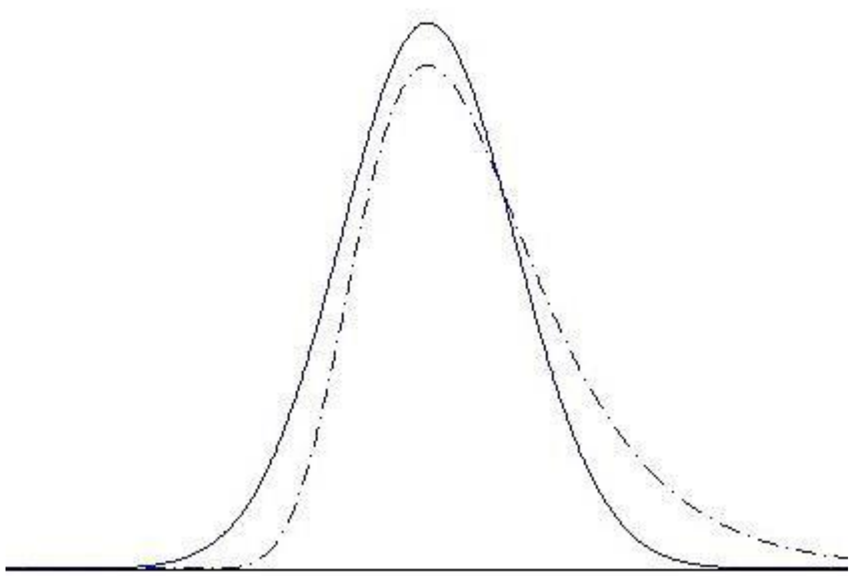


Figure 1. Gumbel Distribution

If $\Phi(x)$ is the cumulative distribution function of a normal distribution as mentioned below:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx$$

Then the corresponding probit function can be shown by $\Phi^{-1}(x)$.

Figure 2 shows the similarity of logit and probit functions particularly when their slope matches at the point y=0.

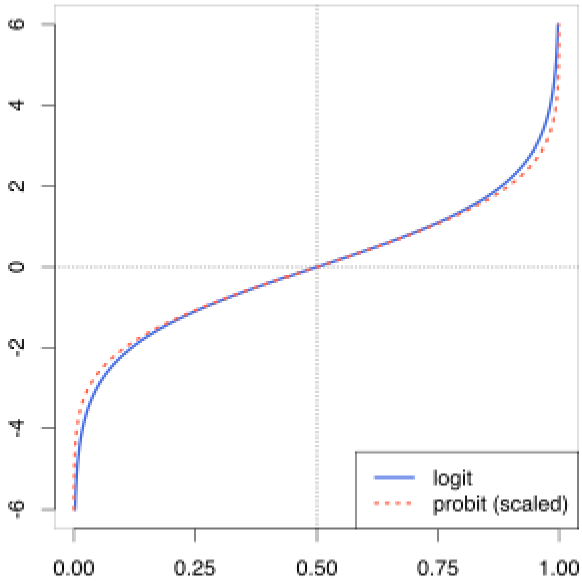


Figure 2. Logit and Probit Models

3.5 Logit Model

Generally, if ε_{in} has normal distribution, its CDF results the probit model and alternatively, if it has Gumbel (Weibull) distribution logit model derives from it. The important assumption in logit models is that the error terms are all distributed identically, independently and following double exponential distribution:

$$F_{\varepsilon}(\varepsilon_j) = exp(-e^{-\varepsilon_j})$$

$$f_{\varepsilon}(\varepsilon_j) = e^{-\varepsilon_j}.exp(-e^{-\varepsilon_j}) \quad (3.11)$$

Using Eq. (3.10) in Eq. (3.11):

$$P_j = \int_{\varepsilon_j=-\infty}^{\infty} e^{-\varepsilon_j}.exp(-e^{-\varepsilon_j}) \prod_{k\neq j} exp(-e^{v_j-v_k+\varepsilon_j})\, d\varepsilon_j \tag{3.12}$$

Since v_{ij}–v_{ij}=0

$$P_j = \int_{\varepsilon_j=-\infty}^{\infty} e^{-\varepsilon_j} \prod_{k\neq j} exp(-e^{v_j-v_k+\varepsilon_j})\, d\varepsilon_j$$

$$P_j = \int_{\varepsilon_j=-\infty}^{\infty} e^{-\varepsilon_j}.exp(-\sum_k e^{v_j-v_k+\varepsilon_j}) d\varepsilon_j \tag{3.13}$$

$$P_j = \int_{\varepsilon_j=-\infty}^{\infty} e^{-\varepsilon_j}.exp(-\sum_k e^{v_j-v_k}) d\varepsilon_j$$

By doing a change in the variables and take t as $t = e^{-\varepsilon_j}$ then $dt = -e^{-\varepsilon_j} d\varepsilon_j$ and $d\varepsilon_j = -\frac{dt}{t}$. It is obvious that as ε_j approaches infinity t approaches 0 and reverse. Replacing $e^{-\varepsilon_j}$ with t in equation (3.13) we have:

$$P_j = -\int_{t=-\infty}^{0} exp\left(-t \sum_k e^{v_k-v_j}\right) dt$$

$$= -\left(\frac{1}{-\sum_k e^{v_k-v_j}}\right) exp(-t \sum_k e^{v_k-v_j}) |_{-\infty}^{0}$$

(3.14)

$$= \left(\frac{1}{-\sum_k e^{v_k-v_j}}\right) = \left(\frac{1}{-e^{-v_j}\sum_k e^{v_k}}\right)$$

and finally:

$$P_j = \frac{e^{v_j}}{\sum_k e^{v_k}} \tag{3.15}$$

3.7 Multinomial Logit Model (MNL)

This is the simplest and most popular practical discrete choice model. The important characteristic of this model is that in it, usually the utility functions are assumed to be linear in parameters.

There are other kinds of logit models like the Nested Logit (NL) and the Mixed Logit (ML) as well which are both more complicated than MNL but, as the area of concentration in this thesis is increasing the accuracy of the observed part of the function and by considering the homogeneity of the individuals in Famagusta, this model, even with is simple form, is extremely reliable and will effectively fulfill the needs if this research.

Now that the model is specified it is the time for working on the observed part and generating the model afterwards.

3.8 The Observable Component of Utility *V*

The measurable part can be defined as a function of the observed characteristics of individuals and vehicle as follows:

$$V_{ij} = \sum_k \theta_{kj}\, x_{ijk} \tag{3.16}$$

where θ is the related coefficient for attribute k of the alternative j or socio-economic characteristic of individual *i*.

As the goal of this study is to find the probability of choosing different means of transport in Famagusta, first it is needed to establish the utility function for all the available alternatives. The main factor that makes this research special is the idea of using ArcGis to generate the observed part of the utility function.

To better understanding the subject suppose that a person decides to go to work at 8:00, come back home at 16:00, go shopping with his family at 18:00 and return home at the end of day at 23:00. Our goal as a transport planner is to find the probability of choosing each mean of transport for every single trip that this person wants to make. For this purpose as mentioned before, it is needed to establish the utility function for all the available means of transport and this work requires modeling both the observed and unobserved parts of the function.

In almost all utility functions, travel time, travel cost, waiting time and availability are the four major parameters, describing the attractiveness of conveyance by each mean, but there are also some difficulties in calculating the true values for these parameters to reach a fine model. Some of these difficulties relating to each one of the above parameters are listed as:

1. Travel time: Since it strictly depends on the travel distance, travel route and time of day, it cannot be treated as a constant value as it has been treated in most of the previous researches.
2. Travel cost: Is also a complex concept and not easy to be determined and that is because, it is formed by the summation of a wide range of parameters including maintenance, fuel and annual insurance cost for private cars and ticket price and fares for public transportation vehicles.

3. Waiting time and availability are also two parameters that differ according to time and from one location to another.

To tackle these problems a GIS-based decision support system (DSS) is established for the sample study area and with the advantage of implementing and visualizing spatial data an intelligent system is built to calculate the probability of choosing each mean of transport for any desired time of day, based on the traffic data and cumulative travel cost by considering the traffic condition and cumulative distance which are two elements, being used in taxi meter algorithms. In the next part the procedure of making this DSS is explained in detail.

3.9 Digitized Map of Famagusta

Since there is no available digitized map of Famagusta to work on, the starting point is to prepare a digital map of the study area in the ArcGis software. This time-consuming work is done by the editor toolbar of the ArcMap 10.0 and the imagery base map provided by Bing (Figure 3). For this purpose, first a new File Geodatabase is created and named Famagusta.gdb by ArcCatalog 10.0 and then a feature dataset (Transportation) including three feature classes is built in it. According to this study these feature classes are named Streets, Bus_Routes and Bus_stops which the first one is polyline and the other two are point feature classes. The dependent attribute table of this spatial data is also built simultaneously as it will be discussed later in this topic.

Figure 3. Digitized Map of Famagusta

3.10 Creating Network Dataset

After finishing the digitizing part creating the network dataset is the next step with the following main targets:

1. To prepare an environment where in, different means of transport can be compared equally and in an efficient way.
2. To have the accurate values for time, cost and distance between each origin and destination and for any desired date and time.
3. To have a system which is capable of assessing any number of transportation means simultaneously.

4. To have the ability to update the model by simply updating the database.

Network datasets are well suited to model transportation networks. They are created from source features, which can include simple features (lines and points) and turns, and store the connectivity of the source features. When performing an analysis using the ArcGIS Network Analyst extension, the analysis always happens on a network dataset. There are three types of network elements, constituting network datasets: edges, junctions and turns which these network elements itself, are created from the same source features as those were implement to create our network dataset from. In addition, network elements have attributes that control navigation over the network. These three elements can be defined as below:

1. Edges (or street sections in most of the studies) are line features which the agents are able to travel through.
2. Junctions are point features, connecting the edges. The application of junctions is to simplify navigation between edges.
3. Turns are line features, storing information that can affect movement from one edge to another.

All these three kinds of network elements are created to build the Famagusta network dataset.

3.11 Collecting Traffic Data

By considering this fact that if route planning is done from one location to another without account for traffic, the expected travel and arrival times could be far from accurate and moreover, some routing options that save time by avoiding the slower, more congested roads may be missed, and besides the travel mode choice can strongly be affected, the traffic data is collected precisely for the whole 24 hours of a day from Monday to Sunday with GPS for all the edges of our network. The following figures show how the traffic condition can affect the best route choice in different time of days (Figure 4) and how one mode of transport can become more preferable in traffic (Figure 5).

Figure 4. How Traffic Affects the Best Route Choice

Figure 5. Which Travel Mode Is Better in Traffic

It has to be mentioned that today, in most of the civilized countries live traffic data is used for these kind of studies, providing current speed using a number of sources, such as GPS receivers in vehicles and speed sensors on roads. But as live data is not available for Famagusta, historical traffic data is used in this study and all the required data is collected manually.

One of the several available options for generating historical traffic data is to store the time related cost for each edge. These costs should contain the speeds for each edge or streets for a whole week. To do so, a week is partitioned into 336 discrete, thirty-minute intervals which mean that we should have 336 cost attributes for each edge of this network, representing how the speed changes and in other words shows how traffic influences the vehicle movements during a week.

To avoid duplication, instead of storing all these values for every single street that needs a large space especially for bigger networks another method is used in ArcGis Network Analysis by noticing that many streets following the same trend of speed changing. Therefore rather than store all these 336 values per feature a related table is created to hold all these information. The rows of this table are speeds or travel times which have a linear relation with speeds for every thirty-minute in a day. A traffic profile can be drawn by connecting the discrete values of each row that means for any number of edges that share the same traffic profile just a single, unique row needs to be stored.

It is obvious that those streets referring to the same profile should not necessarily have the same speed limits for all the time intervals. The only thing that must be in common between them is the pattern they fallow throughout a day.

To create these profiles first the free-flow speed should be determined for each road section. To do so, for the roads with certain speed limits the maximum legal speed numbers are used as the free-flow speed and for the others the average speed of a vehicle in the absence of other traffics are used. Collecting the speed values was started for the whole day at thirty-minute intervals then. This work was done by a RoyalTek GPS logger RGM-3800 (Figure 6) and the results were uploaded to the http://www.gpsvisualizer.com/ website and converted to Google Earth KML files.

Figure 6. RoyalTek GPS Logger RGM-3800

One of the great advantages of this website is that the obtained data can be labeled by time and colorize by speed. A sample output of this work can be seen in Figure 7.

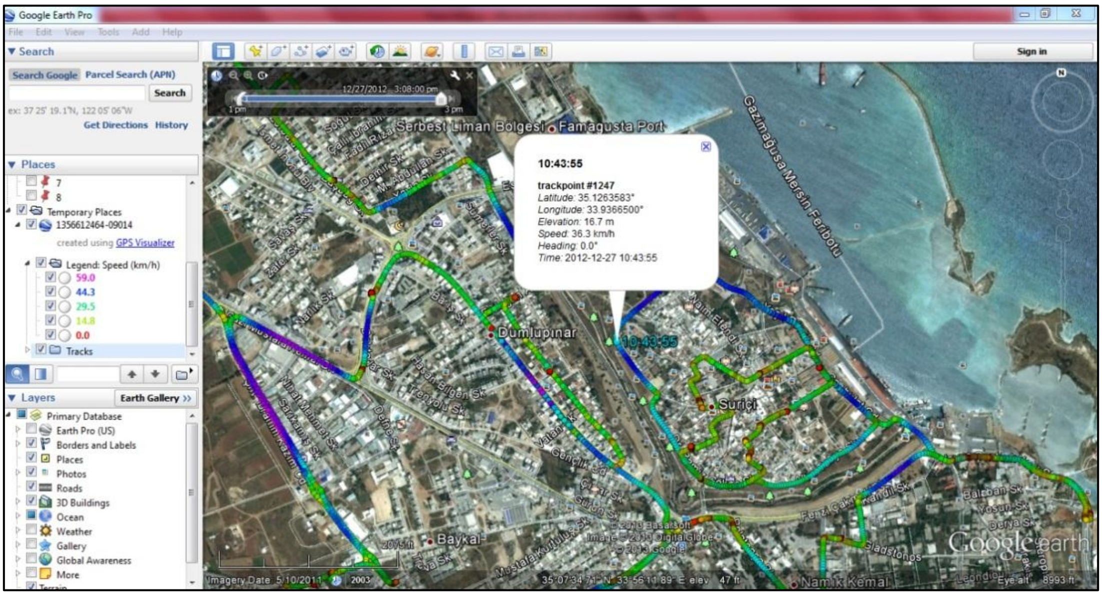


Figure 7. Famagusta KML files in Google Earth

It has to be mentioned that after doing some data collection it was found out, due to the similarity in traffic patterns for the weekdays and weekends, for 90% of the streets in Famagusta only two traffic profiles need to be generated, one for the weekdays and one another for the weekends.

By having this data, they are normalized for each edge by taking it’s free-flow speed as 1 and others as scale factors between 0 and 1 of that number using this formula:

$s = V/V_f$

(3.17)

where:

V=free-flow speed

It is obvious that in this case, other speeds are always less than the free-flow. For example if the free-flow speed is 70 mph and the travel speed is 28 mph at 8:00 AM and 60 mph at 18:00 the relative scale factors would be 1, 0.4 and 0.85 respectively (Figure 8).

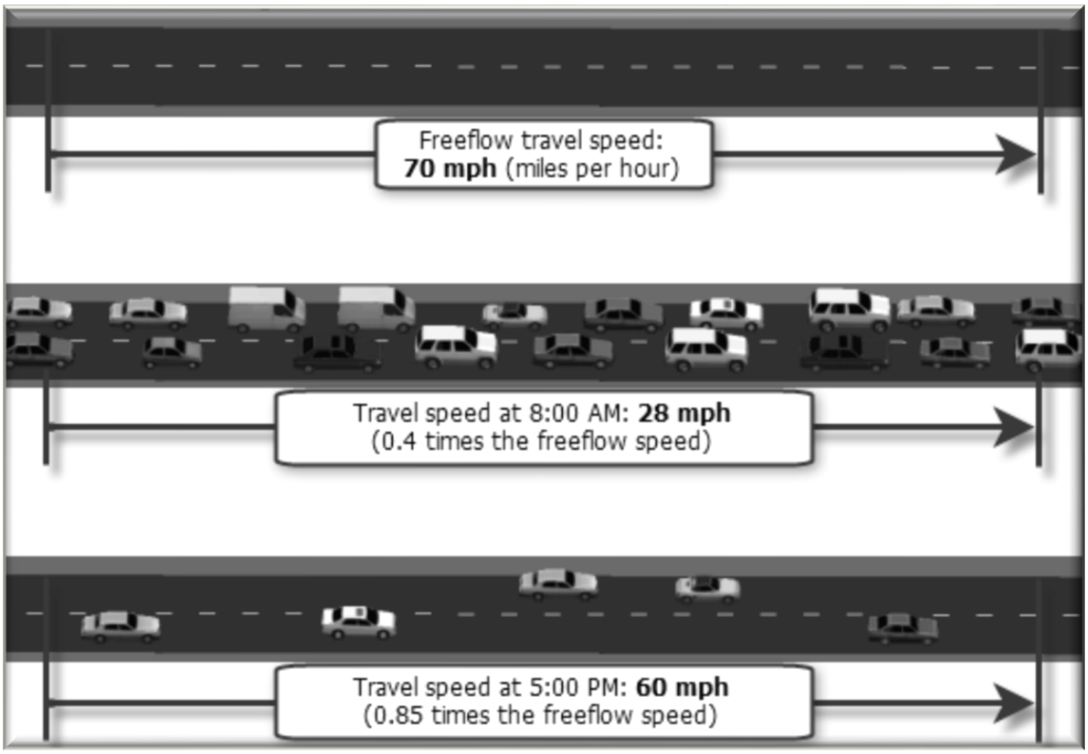


Figure 8. Travel Speed and Scale Factors

Once the observation is completed and the profiles are plotted, all unique profiles are stored in a reference table and linked to the associated streets (Figure9)

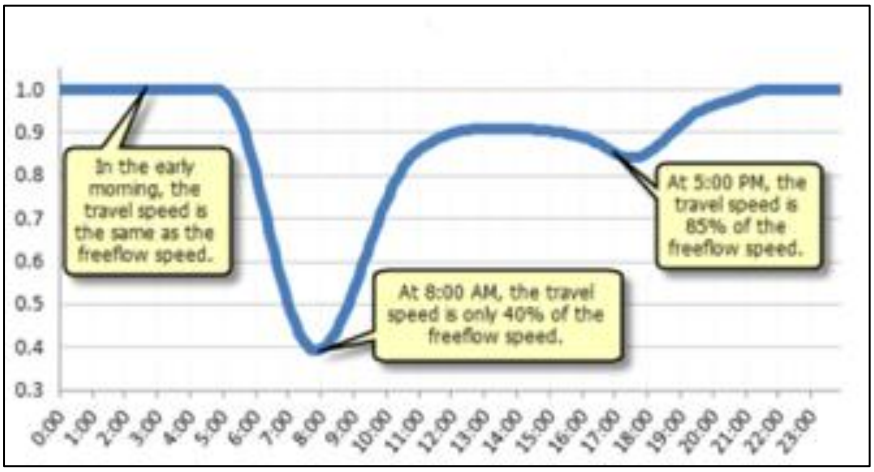


Figure 9. A Sample Profile

As it is needed to use the scale factors of travel time for Famagusta, the inverse values of the obtained output of Eqn. (3.17) are calculated and the profiles are simply drawn through Microsoft Excel in different times of day (X-axis: Time starting from 00:00, Y-axis: Time scale factor).

After comparing the travel time for the all single streets of the city, 36 unique profiles are plotted and assigned to our edges (Figure 10). As all the streets follow the same strategy, the procedure is only described for the primary roads of Famagusta here (As mentioned before it is presumed that, free-flow speed is the same for all seven days of week). These three main roads of the city are listed as below:

1. Ismet Inonu Blv;
2. Gazi Mustafa Kemal Blv and
3. Fevzi Cakmak Blv.

Due to different traffic flows the first road is splinted into seven sections for each direction, with the total number of 14 sections, to make the analysis more accurate. These road sections are listed as below:

1. From DAU Sqr. to Bahriyeli
2. From Bahriyeli to Faiz Kaymak
3. From Faiz Kaymak to Cahit Sitki Taranci
4. From Cahit Sitki Taranci to Savas
5. From Savas to Cami roundabout
6. From Cami roundabout to Erdogan Acar
7. From Erdogan Acar to Zafer roundabout

The same division is also done for the second road with two sections:

1. From Sabanci Entrance to Famagusta-Nicosia roundabout
2. From Famagusta-Nicosia roundabout to Zafer roundabout

Besides, according to the data, collected by GPS, 17 more profiles are prepared for the other streets in Famagusta, which two of them (Fevzi Cakmak and Havva Senturk) have only hourly traffics at schools closing time. Four profiles are presented here as a sample in Figures 11 to 14 for Bahriyeli to Faiz Kaymak, Cahit Sitki Taranci to Savas, Cami roundabout to Erdogan Acar and Famagusta-Nicosia roundabout to Zafer roundabout respectively.

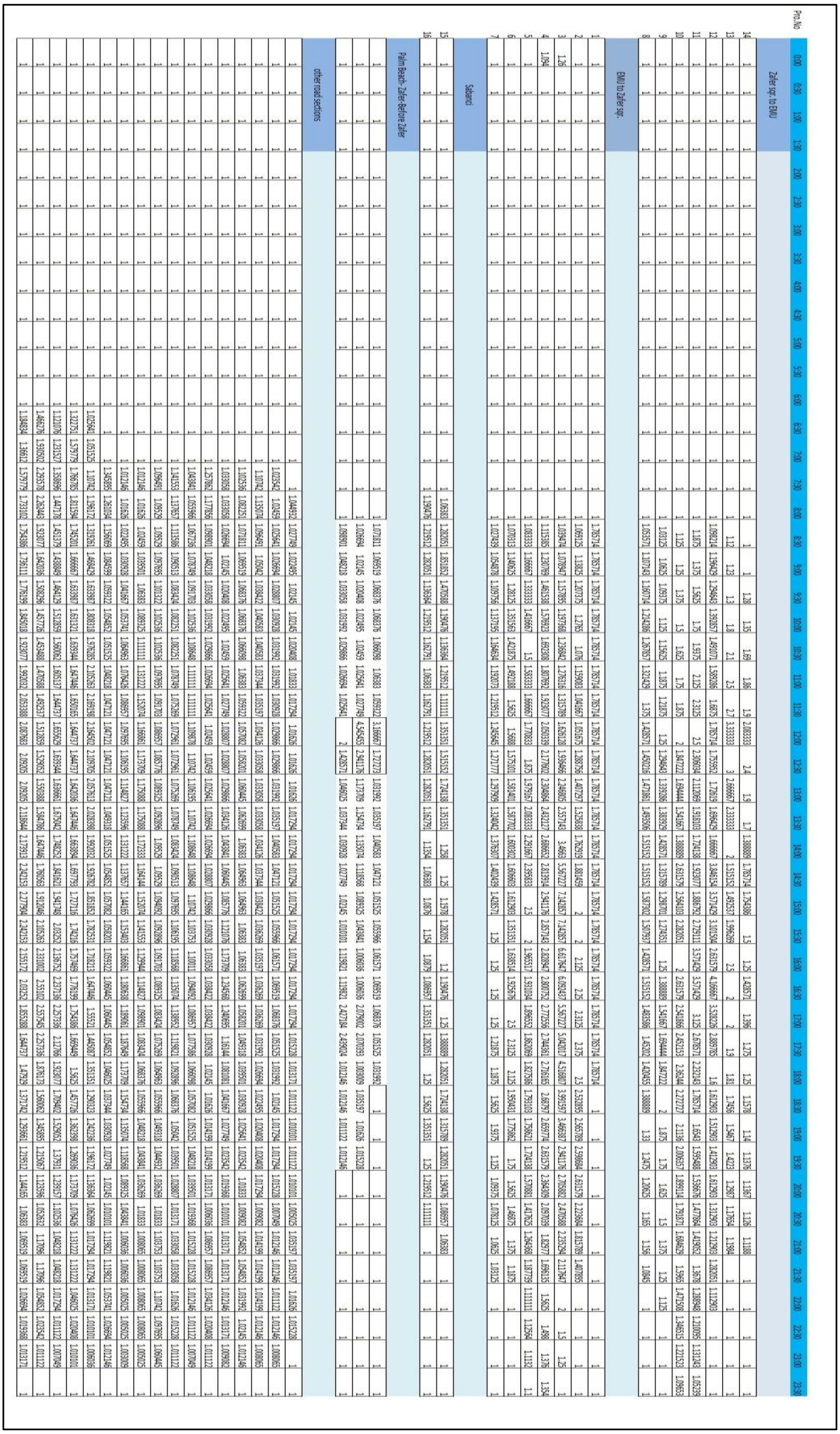


Figure 10. Famagusta Streets Travel Time

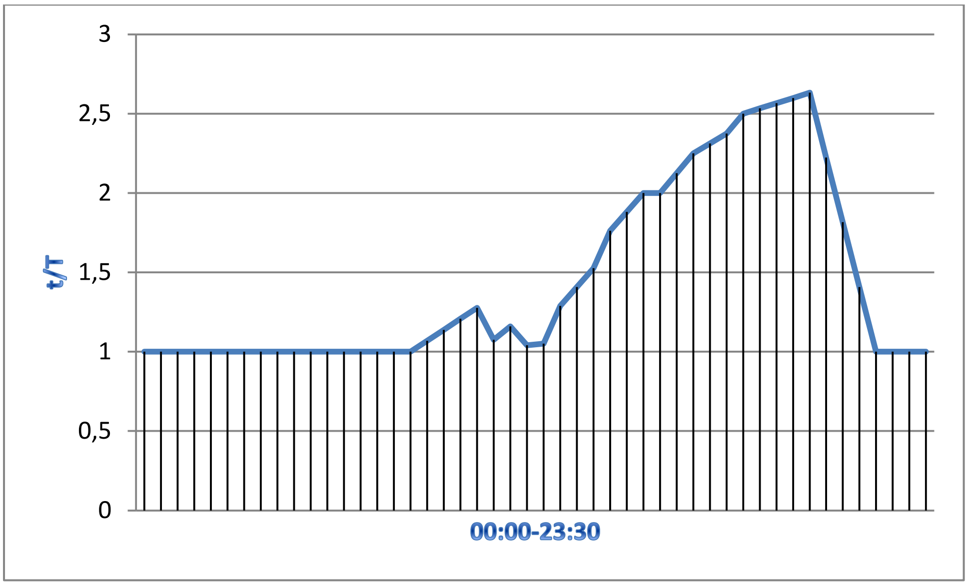


Figure 11. Bahriyeli to Faiz Kaymak

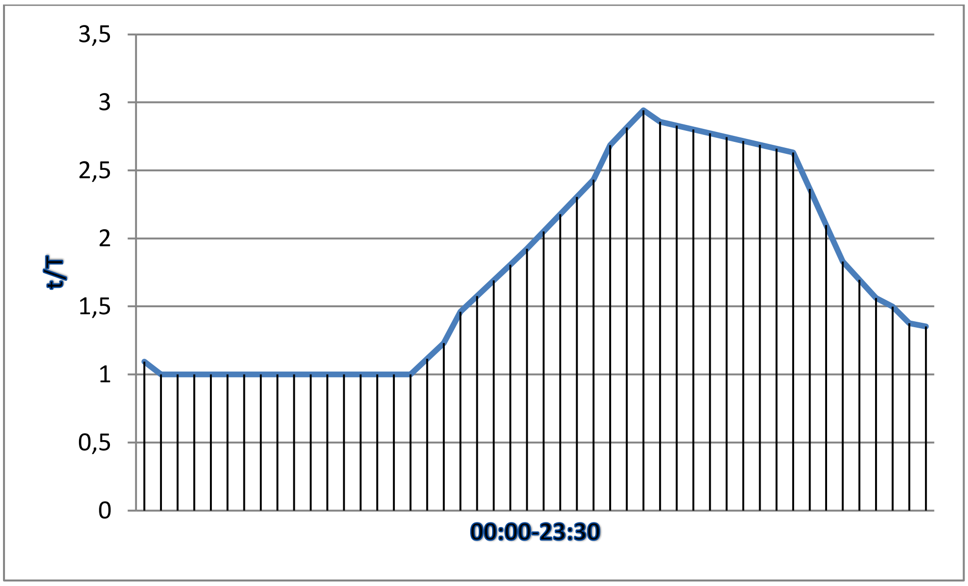


Figure 12. Cahit Sitki Taranci to Savas

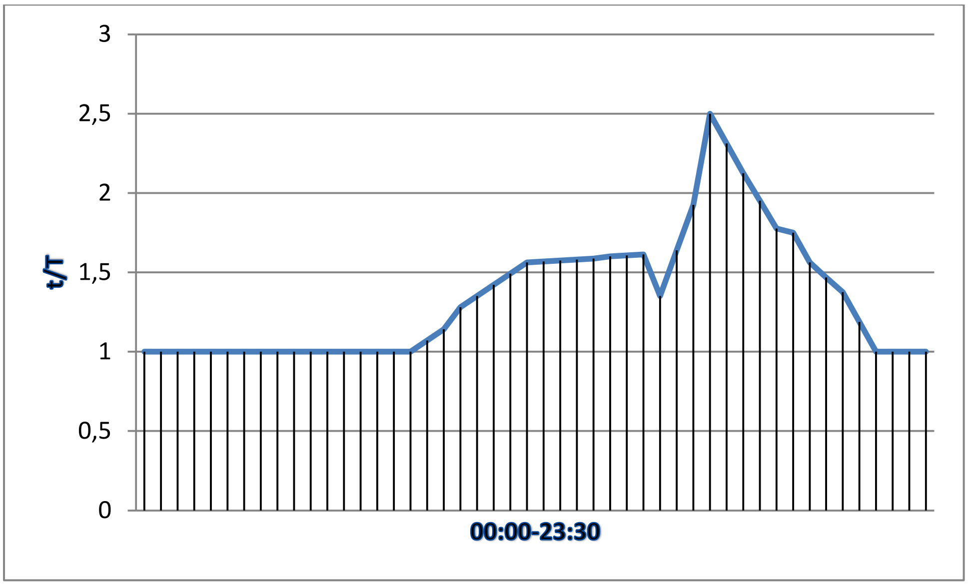


Figure 13. Cami Roundabout to Erdogan Acar

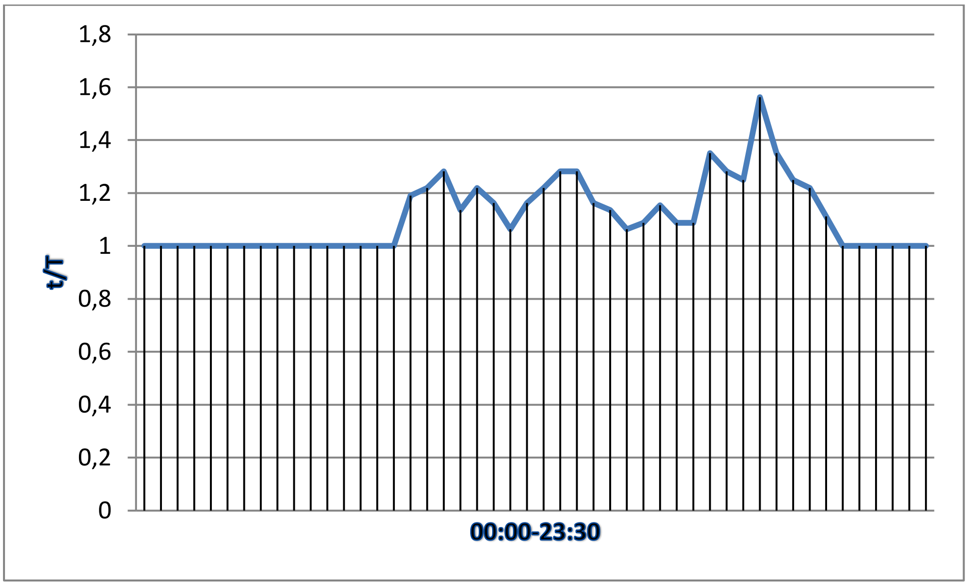


Figure 14. Famagusta-Nicosia Roundabout to Zafer Roundabout

Next, a field named 'Profile' is added to the street attribute table and the corresponding profile number is imported for each road section (according to the profile No. in the excel file). It should be mentioned that although, the profile numbers are imported for three primary roads, the profiles should be imported while building the network data set.

It is obvious that the final results can be used in finding the best route with the additional parameter of Traffic Flow.

Finally, a table is created, storing the streets ID, free-flow speed and the profile number for each road segment from Monday to Sunday called the Streets-Profiles join table in the Famagusta geodatabase. Technically, for a network dataset, which in this case is a multimodal network; with historical traffic data two more tables should be created as well. One, storing the profiles characteristics which include the scale factors for every thirty minutes and the other the street-profile join table. It is explained later in detail that how these tables are used in the network generation.

3.12 Multimodal Networks

As the aim is to investigate the operation of four means of transport, a single network cannot fulfill the expectations and a new type of network is required to be built and worked on for this research. Multimodal networks enable the researcher to model more than one mode of transport over a single feature dataset.

For this research, a multimodal transportation network of four modes is built; private car, taxi, bus and bicycle. This model is also capable to show the travel time for those individuals who prefer to travel on their feet.

To set up this network, first it is needed to create all the necessary feature classes. Private cars and taxies both use the streets and only have direction constrains like what we have for one way streets. Buses on the other hand can only travel through their defined routes according to their time table and although in the real world there are no exact limitations for pedestrians and bike riders, as it will be discussed later their scope has to be restricted to the street feature class.

To create the network dataset first the feature classes participating in the network which are as mentioned before streets, bus routes and bus stops should be selected. Then, the connectivity groups should be defined.

Connectivity groups show the way that network elements are connected to each other; it means that two edges can only connect to each other if they have one of these two conditions:

1. They both belong to the same connectivity group or
2. If they belong to two different connectivity groups they should join by a point or junction that is in that two groups at the same time.

As it is required to restrict the bus network to the 'Bus_Routes' feature class, we two connectivity groups are created for the network dataset. Connectivity group 1 represents the street network and connectivity group 2 represents the bus system and these two are joint by the 'Bus_stops' feature class which belongs to both group 1 and group 2.

Three different time costs are created for this network, the pedestrian time, bus time and drive time and while performing network analyses this gives the ability to have the travel time for each mode of transport or even for multiple transport modes like what we have when one wants to travel from point A to point B by bus. In this case normally, the person cannot directly board the bus and goes to his destination. He first needs to walk along the street to reach the closest bus stop and then rides the bus for the rest of his travel. Therefore it is required to have the walking time together with the bus travel time to calculate the total time for travel.

For the pedestrians and bicycles the average number of 3km/h and 10km/h are taken respectively and to apply them in the network the related queries are written in Visual Basic which enables the user to calculate the cumulative travel time for each of them.

3.13 Taxi Fares

Although there is no standard rate for fares and no control on the amount of money that taxi drivers take from their passengers, to have an approximate number for the model calculations a simple code which is implemented in taxi meters is written. This algorithm describes that the taxi meter charges every passenger 2 Turkish Lire (TRY) immediately after taking into the taxi for the first 250 meters of travel and after that, a rate of 0.75 TRY for every further 265 meters is added to the total fare. This algorithm is formulated in this research and the following equation is presented for it:

$$C[TRY] = 2.0 + \left(\frac{d[m]-250}{250}\right) * 0.75 \qquad (3.17)$$

where:

C= Taxi fare in Turkish Lira and

D= Travel distance in meters

It can also be mentioned that although most of the taximeters calculate the total fare based on both distance and time but as Famagusta is a small city with fluent traffic by ordinary, the role of traffic is not taken into consideration in our equation (Figure 15).

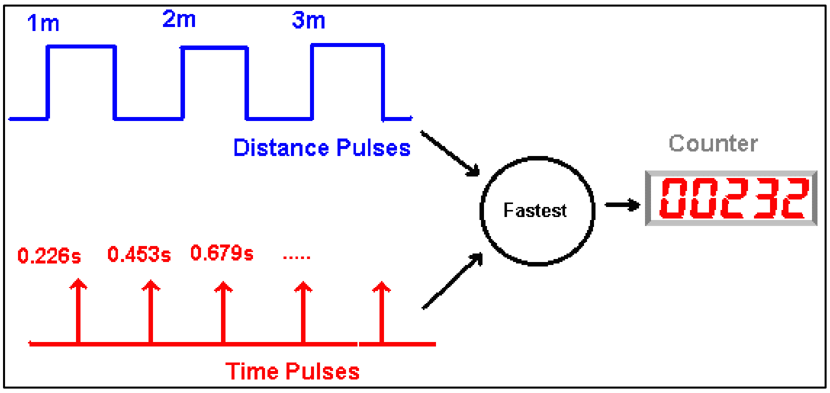


Figure 15. Taxi Meters Algorithm

3.14 How to Work With Developed Network Dataset

Defining the network attributes and assigning traffic data to that is the last step and after, the output is ready be used in analyzing the transportation network in our study area.

As expressed before, one of the major problems with previous transportation modal choice models is to prepare an environment where in, different means of transport can be compared equally and in an efficient way. This can be reached by ArcMap because it uses the well-known Dijkstra's algorithm to calculate the shortest or quickest path for all the modes in different times of day, therefore it is no longer needed to concern about this question that what if taxi is better for going from point A to point B in all aspects compared to bus, but a taxi driver by choosing a travel route other than the best or the fastest one, makes it less preferable in real world.

The classic Dijkstra's algorithm solves the single-source, shortest-path problem on a weighted graph in ArcGIS. To find the shortest path between points, the weight or length of a path is calculated as the sum of the weights of the edges in the path. Dijkstra's algorithm finds the shortest path from point A to point B in order of increasing distance from point A. That is, it chooses the first minimum edge, stores this value and adds the next minimum value from the next edge it selects. The solving process starts out at one vertex and branches out by selecting certain edges that lead to new vertices which is similar to the minimum spanning tree algorithm, in that it is "greedy", always choosing the closest edge in hopes of an optimal solution.

As it can be see in Figures 16 and 17 a sample origin and destination is defined and the best route is found for 6/14/2013 at 10 AM. Since a multimodal transportation network is implemented, the time cost for each mode of transport plus the taxi fare is calculate for this route as well (Figure 18).

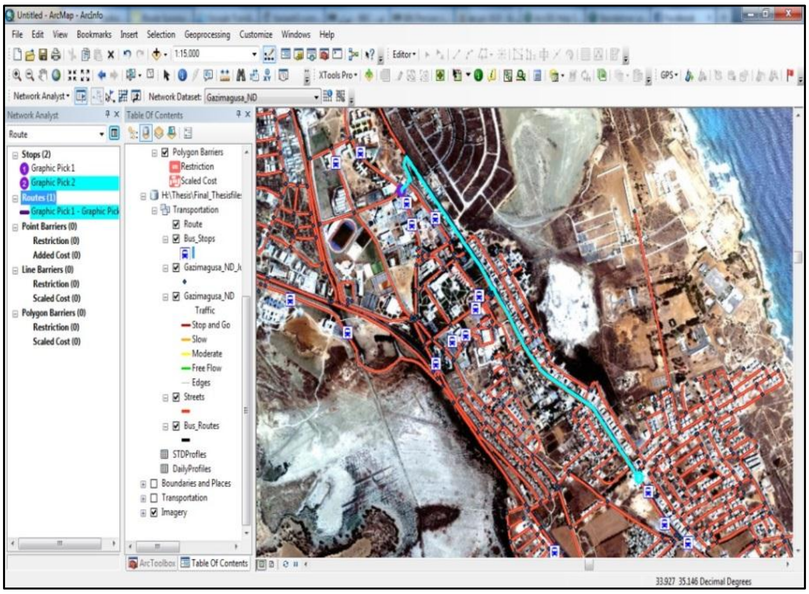


Figure 16. Best Route for the Sample Origin and Destination

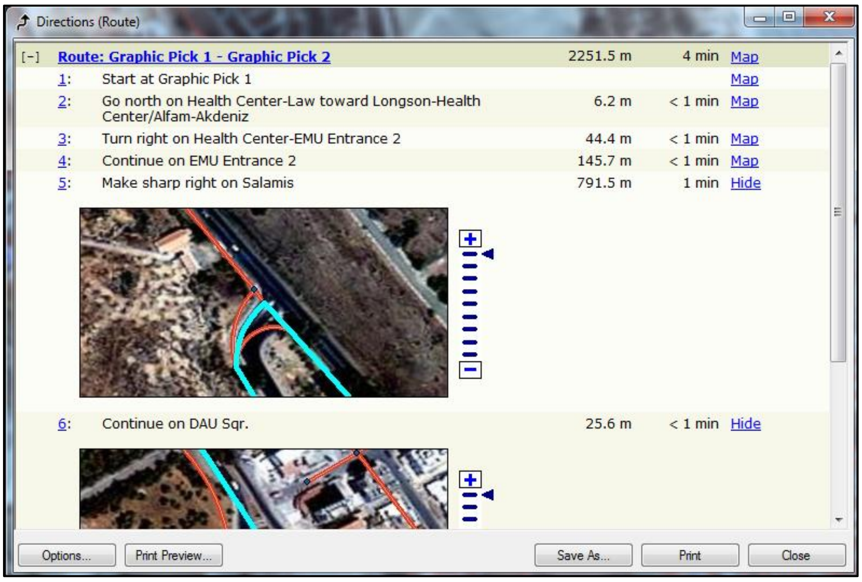
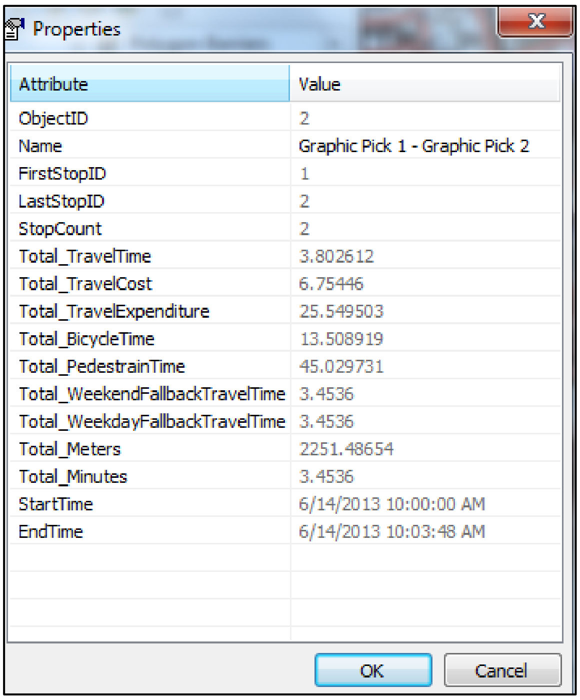


Figure 17. Detailed Best Route for the Sample Route


Properties

Attribute	Value
ObjectID	2
Name	Graphic Pick 1 - Graphic Pick 2
FirstStopID	1
LastStopID	2
StopCount	2
Total_TravelTime	3.802612
Total_TravelCost	6.75446
Total_TravelExpenditure	25.549503
Total_BicycleTime	13.508919
Total_PedestrainTime	45.029731
Total_WeekendFallbackTravelTime	3.4536
Total_WeekdayFallbackTravelTime	3.4536
Total_Meters	2251.48654
Total_Minutes	3.4536
StartTime	6/14/2013 10:00:00 AM
EndTime	6/14/2013 10:03:48 AM

OK Cancel

Figure 18. Transit Modes Attributes

Now, almost all the required parameters of the deterministic part of the utility function for the four modes of transportation in Famagusta are available. To generate the utility functions for each mode the parameters participating in our models structure have to be signified. It is crucial then to construct our model sensitive to those parameters of travel that affect individual modal selection. It will be discussed how this necessity can be satisfied by choosing the appropriate parameters, well describing the attributes of travel. To do so, the important factors affecting the choice of model are probed in the next part and then, those which are important in the traveller's point of view in Famagusta are selected.

3.15 Factors Affecting the Modal Choice

The items influencing mode choice can fall into these three categories:

1. Characteristics of the traveller: The most important one among all the characteristics can be listed as:

- Car availability;
- Family type (young couple, couple with children, retired, singles, etc.);
- Average income;
- Residential density.

2. Characteristics of the journey:

- The trip purpose; for instance, daily travels to work or the university are easier and consequently more probable to be done by public transport vehicles. Alternatively, journeys which are made with entertainment purposes are more likely to be done by taxi or private cars.
- Time of the day, for example in many cities public transportation vehicles are not available for late time travels, besides, traffic conditions and travel time are two important factors which directly depend on the time of day.
- Either the journey is undertaken alone or in a group.

3. Characteristics of the transport facility. Two classes can be defined for this group: Firstly, quantitative attributes such as:

- Travel time: combination of access, waiting and in-vehicle times for each mode;
- Travel expenditure: components of fuel, annual insurance, maintenance, ticket and other costs;
- Availability of parking spaces;
- Reliability of travel time and
- The regularity of the service.

Secondly, hardly determined qualitative factors like:

- Comfort and convenience;
- Safety and security;
- Possibility to do other activities during travel (use the phone, read, etc.).

Note that the work will be more valuable and closer to the real world if tours are considered for our model generation with trips as their components. The simplest form of a tour can be a from-home and back trip. This will also allow us to reach to more generalized models and better investigate the use of multiple modes of transportation between any origin and destination.

Additionally, it should be mentioned that different people may perceive costs in different ways, means that although for some people time is more important than expenditure for some others money might be the most important factor. To draw a good deal of interest a questionnaire was prepared and nine questions were asked from 200 individuals about the mean of transport they use for their daily travel.

Besides, their idea about some possible solutions for the transportation problems in Famagusta was come into question (see Appendix 1). The questions are discussed and the approached results are analyzed in the next chapter.

CHAPTER 4

RESULT AND DISCUSSION

4.1 Introduction

In this chapter the results of the questionnaire will be evaluated and our model will be built accordingly. Moreover, as people are asked about the possible solutions for some public transportation deficiencies an attempt will be made to predict the changes in the interest of travellers for public transportation vehicles after doing these modifications as well.

4.2 Evaluation

For better evaluation purposes all the obtained results are imported in Microsoft Access10.0 where in, by simply writing the proper SQL queries more accurate results are acquired in a short time.

The result shows 89% of car owners, which are 21% of the whole population, use their own car for their daily travels. This can be interpreted as poor public transport service in Famagusta that most car owners have a great preference to use their own cars by considering this fact that all the rest 11% of them that choose other options only need to spend less than 10 minutes to go from their home to work or university.

In spite of all the problems with the bus service in Famagusta, according to statistics 24% of travellers use this mean for their daily trips which, more than 50% of them live far from university or their work place.

Majority of individuals that are 41% of all questioned persons say that they go to work or university on foot but this is not surprising if we note that only 12% of them live far from their destination.

And finally, 10% of people choose bicycle as their transport mean that is still 2% more than the travellers, using taxi frequently for their everyday travels.

Figure 19 briefly shows the percentage of people choosing each mode of transport.

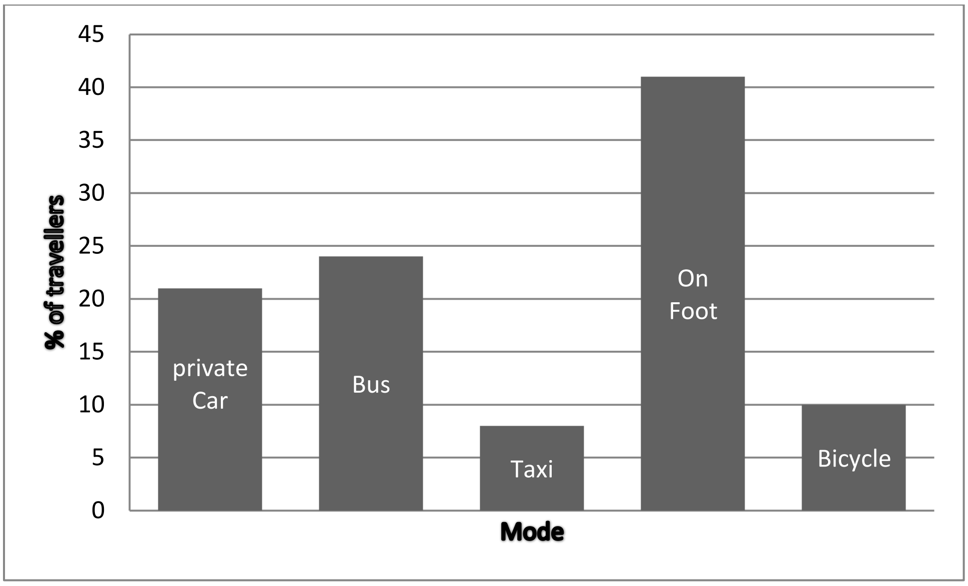


Figure 19. Percentage of People Choosing Each Mode of Transport

4.3 Form the Utility Function

To decide which variables enter the utility function a search process is normally employed starting with a theoretically appealing specification [16]. The important thing is that, none of the alternatives can be looked individually and even for each alternative the important factor is the relevant influence of each attribute representing by the coefficients rather than the absolute value of them. So, the other important work that should be done for the model generation is to set one of the values of attribute x of alternative B_j equal to 1 and take the corresponding coefficient θ_k as the alternative specific constant or ASC and then find the relative values for the others.

A well-known commonly used method for finding the coefficients is the maximum likelihood estimation. In the next part this method is explained and the applicability of it to our study is investigated.

4.4 Maximum Likelihood Estimation

To find the proper value for the coefficients a method called maximum likelihood estimation is used. Applying this, leads the model to make prediction in a form that better matches the observed data [17]. In this case as it is required to use MNL to find the possibility of each mode to be chosen by the travellers the main assumption is to have double exponential distribution for the error term.

In this method there is a set of values for the observed part (V_{ij}) and the goal is to find the θ_{jk}s which maximize the possibility that the model generates the observed data. If we have a specific choice set and an overall knowledge about the people's interest, we can find the probability of the model, choosing the same option as what we know through our questionnaire for any individual *i* like this:

Assume:

$$\prod_j P_j^{\varphi_{ij}} \tag{4.1}$$

where $\varphi_{ij} = 1$ if individual i chooses product j, and $\varphi_{ij} = 0$ otherwise. If the process is repeated for n individuals, the total number of individuals selecting alternative j is $n_j = \sum_i \varphi_{ij}$ and the probability of the model generating the observed choices is

$$\prod_j P_j^{n_j} \tag{4.2}$$

As the coefficients that maximize the quantity are sought, a common way to reduce the complexity and numerical difficulties is to maximize the logarithm of the Eqn. (4.2) rather than working on it directly. Mathematically, the value that maximizes Eqn. (4.2) is exactly the same as the one which maximizes its logarithm with this difference that the logarithmic form becomes into the simple linear shape when it is differentiated. This is called the log-likelihood, usually written LL. The maximum log likelihood θ terms are therefore:

$$\theta = argmax(\sum_j n_j log P_j) \tag{4.3}$$

where, as said before, for logit case:

$$P_j = \frac{e^{v_j}}{\sum_k e^{v_k}}$$

For the model approaching in this study one of these two groups of attributes can be used:

1. Generic, if attribute participates in all the utility functions of the available alternatives. In this case it can be postulated that their coefficients are identical and can be shown by a unique symbol (θ_k rather than θ_{jk}).
2. Specific, if the attribute does not appear in all the utility functions.

To check if the maximum likelihood estimation is appropriate for this study, generic attributes are used to set the model and the parameters decided to apply are access time, waiting time, in-vehicle time, travel cost and relative travel comfort. To better understand, this method is expressed here as an example.

Assume the probability of choosing each mode of transport for a specific route (i.e. from Longson dormitory to Cami square) is needed to be found. First the origin and destination and the desired departure time have to be specified in the ArcMap10.0 and the network analysis extension should be used after to find the best route. By doing this, an inclusive pack of required information is prepared, including travel time for each mode together with the taxi fare by considering that using bus is free of charge in the city(Figures 20 and 21).

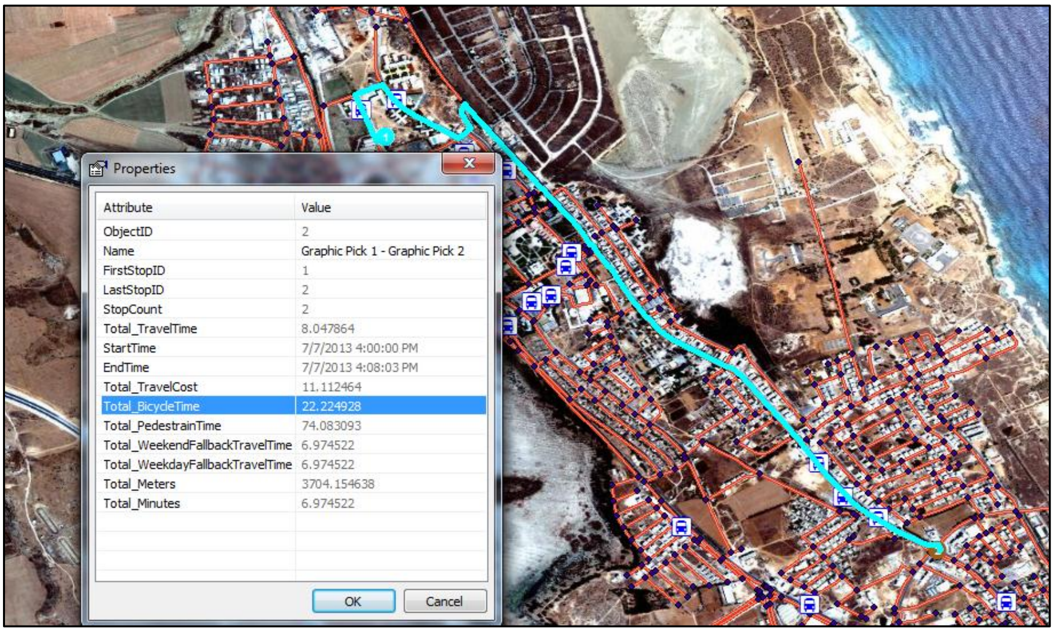


Figure 20. Transit Modes Attributes for the Sample Route

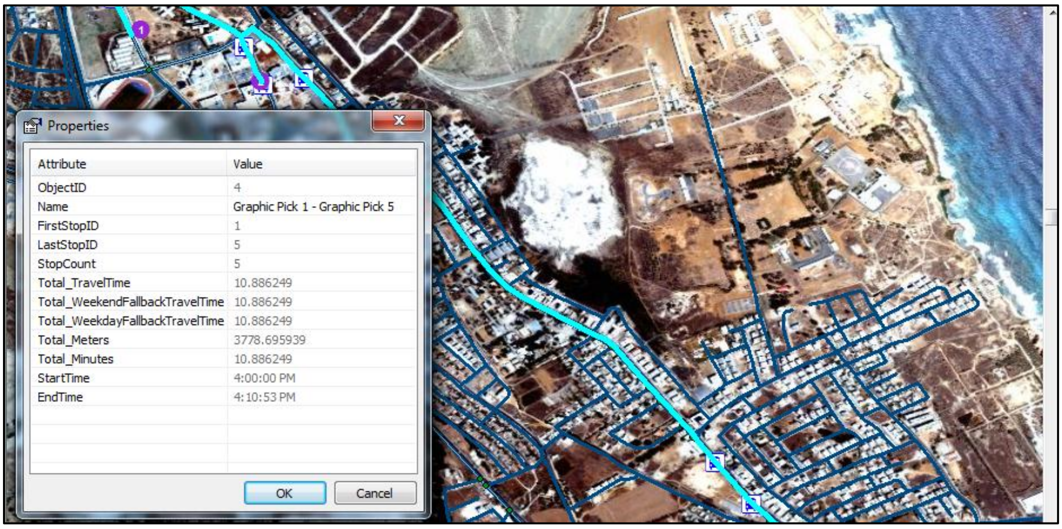


Figure 21. Bus Attributes for the Sample Route

For simplicity all the obtained values are shown in the table below. Note that, travel expenditure for private vehicle is an approximate result of maintenance, annual insurance and fuel cost.

Table 2. Travel Attributes for Different Modes

Mode	Access Time(min)	Waiting Time(min)	In-Vehicle Time(min)	Total Time(min)	Travel Expenditure (TRY)	Relative Comfort
Private car(pc)	0	0	6	6	5	10
Taxi(t)	0	2	6	8	11	10
Bus (bu)	7	15	10	32	0	7
Bicycle(bi)	19	0	0	19	0	4
On Foot(f)	62	0	0	62	0	0.5

Using Eqn. (4.3) we have:

$$maximize(21 * \log P_{pc} + 8 * \log P_t + 24 * \log P_{bu} + 10 * \log P_{bi} + 41 * \log P_f)$$

where

$$P_j = \frac{\exp(\theta_0 v_{1j} + \theta_1 v_{2j} + \theta_2 v_{3j} + \theta_3 v_{4j} + \theta_4 v_{5j})}{\sum_k \exp(\theta_0 v_{1k} + \theta_1 v_{2k} + \theta_2 v_{3k} + \theta_3 v_{4k} + \theta_4 v_{5k})}$$

After writing the queries in Mathematica 8.0 the following values are obtained for the coefficients: $\theta_{0=}$ 0.0516819, $\theta_{1=}$ -0.0683901 and θ_2= 0.238084 as shown in Figure 22.

```
In[1]:= a = Exp[8 t0 + 5 t1 + 10 t2] + Exp[10 t0 + 11 t1 + 10 t2] +
          Exp[28 t0 + 7 t2] + Exp[23 t0 + 4 t2] + Exp[64 t0 + 0.5 t2];
        p1 = Exp[8 t0 + 5 t1 + 10 t2] / a;
        p2 = Exp[10 t0 + 11 t1 + 10 t2] / a;
        p3 = Exp[28 t0 + 7 t2] / a;
        p4 = Exp[23 t0 + 4 t2] / a;
        p5 = Exp[64 t0 + 0.5 t2] / a;
        b = 21 Log[p1] + 8 Log[p2] +
          24 Log[p3] + 10 Log[p4] + 41 Log[p5];
        NMaximize[b, {t0, t1, t2}]
Out[8]= {-152.937, {t0 → 0.0516819, t1 → -0.0683901, t2 → 0.238084}}
```

Figure 22. Solving MLL in Mathematica

As it can easily be seen, although θ_0 must enter the utility function with negative sign (as increase in total time should lead to decrease in utility) the output of the programming is a positive signed value which results in wrong answer. After investigation, it is found that applying this method has the following problems which may cause bias and contributes to wrong modeling:

As Famagusta is a very small city, travel time by car, taxi or even bus does not differ much through long-distance and short-distance routes. Hence, to do a correct modeling the study area should be partitioned into very small zones and the required manual data collection and corresponding statistical analyzes must be done for each, which will be very time-consuming and fallible. Besides, loosing generality will be the other result of the above problem in Famagusta

Additionally, as previously mentioned in this chapter, lack of an integrated transportation system and great favor to use private vehicles are two of the most important environmental factors in Famagusta. These two factors contribute to increase in the number of people, using private cars for their inter-urban journeys.

Therefore, if the model construction is based only on the results of the origin-destination surveys the outcome will be far from real and will not reflect the true dominant situation of the society.

The most suitable way to accord the outcomes of the transportation modal choice model with the true situations in Famagusta is to implement a model which in, the great portion of selecting private vehicles for the car owners is considered and this premise is taken as the base for the model.

By considering the above conditions a possible solution is to build a relative model with private car as the reference and define the other alternatives accordingly. To do so, the utility function for private vehicle is set equals to zero ($U_{ij}^{C}=0$) and the influence of this item in people's daily travels is modeled accordingly. This is another theoretical identification issue associated with the MNL.

To build the new model, first the private car is excluded from our options and then the frequency of selecting other transit modes are found in a new questionnaire. The obtained results are shown in the table below for the same route as before:

Table 3. Frequency of Choosing Each Transit Mode After Excluding Private Car

Mode	Frequency
Taxi	30
Bus	180
Bicycle	80
On Foot	110

After considering the questionnaire outcome and do the required programming (Figure 23) the following model is found proper to be used in Famagusta:

```
In[667]:= a = Exp[80 t0 + 100 t3 + 10 t4] + Exp[320 t0 + 0 t3 + 7 t4] +
             Exp[190 t0 + 0 t3 + 4 t4] + Exp[620 t0 + 0 t3 + 0.5 t4];
          p1 = Exp[80 t0 + 100 t3 + 10 t4] / a;
          p3 = Exp[320 t0 + 0 t3 + 7 t4] / a;
          p4 = Exp[190 t0 + 0 t3 + 4 t4] / a;
          p5 = Exp[420 t0 + 0 t3 + 0.5 t4] / a;
          b = 30 Log[p1] + 180 Log[p3] + 80 Log[p4] + 110 Log[p5];
          NMaximize[b, {t0, t3, t4}]
Out[673]= {-494.44, {t0 → -0.00174147, t3 → -0.0197932, t4 → 0.016947}}
```

Figure 23. Find the Coefficients in Mathematica

$$U_{pr} = 0$$

$$U_t = -1.76258 * \mathrm{T} - 19.8813 * \mathrm{E} + 15.0559 * \mathrm{C} + \ln Car$$

$$U_{bu} = -1.76258 * \mathrm{T} - 19.8813 * \mathrm{E} + 15.0559 * \mathrm{C} + \ln Car + \ln Bus \qquad (4.4)$$

$$U_{bi} = -1.76258 * \mathrm{T} - 19.8813 * \mathrm{E} + 15.0559 * \mathrm{C} + \ln Car$$

$$U_f = -1.76258 * \mathrm{T} - 19.8813 * \mathrm{E} + 15.0559 * \mathrm{C} + \ln Car$$

where

$$Car = \begin{cases} 0 & if\ car\ is\ available \\ 1 & if\ car\ is\ not\ available \end{cases}$$

and

$$Bus = \begin{cases} 0, & dq < 3000\ meters \\ 1, & dq \geq 3000\ meters \end{cases}$$

The first condition expresses that for any individual who has private car, the probability of choosing other means of transport is close to zero and the percentage of car owners, using their private car clearly shows that it can be taken as a good esteem. The second condition has also been set to exclude choosing bus from the choice set of the individuals whose the distances of their departure locations are more than 3 kilometers from the closets bus stop.

This time, θ_0 is negative, indicating that increasing total travel time will decrease utility. Similarly, as θ_1 is negative and θ_2 is positive it can be deduced that increase in travel cost and decrease in comfort will diminish the value of utility.

After this model generation, as the travel attributes for each mean of transport are available by GIS, the probability of choosing each mode can be found by simply find their utility number and place them in Eqn. (3.15). Moreover, using the proper coefficients in this model, short-term and long-term decisions can be made about adding new modes or make positive changes to existing ones.

As an example, suppose one wants to decrease taxi fares with the goal of attract 20% more travellers. To find how much the new price should be to reach this goal the following equation can be solved:

$$P_t = \frac{exp(\theta 0v1t + \theta 1v2t + \theta 2)}{\sum_k exp(\ \theta 0v1k + \theta 1v2k + \theta 2v3k)} = 0.09$$

Which in this case the result will be 8.9 TRY.

Chapter 5

CONCLUSION

A key decision in transport planning is choosing the transportation mode. Transport planners typically consider multiple attributes when making this decision, often focusing on cost, comfort and transit time as the primary criteria. This is not a trivial decision, however, as the process often involves multiple criteria, some of which are not readily quantified. Additionally, difference in the environmental and social parameters in different places causes dissimilarity in the importance of individual factors from one location to another.

Mode choice is part of the decision-making process in transportation that includes identifying relevant transportation performance variables. Today, as almost all the countries are going through industrialization, more people are moving from rural areas to big cities and future transportation planning to fulfill the needs of these large group of people even becomes more important than before. Providing new transit modes build sufficient parking spaces, roads and infrastructures and besides improve the level of service and attraction of the current public transportation means are among the necessary works that should be done in this issue.

A great number of attempts have been made by researchers and urban planners to predict the current and future behavior of travellers since the Mid-19th Century. But, as providing new means of transportation and motivating people to use them rather than their private vehicles require lots of efforts and investments and moreover as mentioned before, the rapidly increasing number of people immigrating to big cities is changing the traffic patterns and transportation needs perpetually, modal selection prediction has become increasingly complex in recent years and the previously presented models cannot meet our planning needs anymore. That is the time when the great need for generating updatable models shows up.

In this research a precise dynamic transportation modal choice model was set up by using GIS for the city of Famagusta. To do so, first a transportation network including all the available inter-urban transit modes was built in ArcGIS and then the multinomial logit model was selected to generate the prediction model for the travellers.

The result of this study showed that using GIS for modeling contributes to better investigation and consequently leads to generate more realistic models since it can be used to achieve precise values for the attributes of all available transit modes in the study area and besides, the model can be easily updated according to new situations.

In conclusion, the finally achieved model was found really efficient to be used for the transportation planning in Famagusta. It should be mentioned that although, this research was a case study for a very small town with relatively fewer and simpler transportation problems compared to big cities and metropolitan areas the same method can be used in any other cases with any transportation conditions and this approach can be a starting point for all future studies in this field.

REFERENCES

[1] F., Hutson Gregory, "Cause, Effect, Efficiency & Soft Systems Models," Journal of the Operational Research Society, pp. 149-155, 1993.

[2] ISO. (2008) Bureau International des Poids et Mesures. [Online]. http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2008.pdf

[3] Neil A. Gershenfeld, The Nature of mathematical Modeling.: Cambridge University Press, 1999.

[4] R., Trent, R. and Handfield, R. Monczka, Purchasing and Supply Chain Management.: Thomson South-Western, Mason, OH, 2005.

[5] D.A., Rose, J.M. and Greene,W.H. Hensher, Applied Choice Analysis: A Primer.: Cambridge University Press, 2005.

[6] S.L. Warner, Strategic Choice of Mode in Urban Travel: A Study of Binary Choice.: Northwestern University, 1962.

[7] B.D. Spear, cations of New Travel Demand Forecasting Techniques to Transportation: A Study of Individual Choice Models.: Final Report to the Office of Highway Planning, Federal Highway Administration, US, 1977.

[8] J. de D. Ort´uzar, Mixed-mode demand forecasting techniques.: Transportation Planning and Technology.

[9] D. McFadden, Econometric models of probabilistic choice. In C. Manski and D. McFadden (eds.), Structural.: The MIT Press, Cambridge, Mass, 1981.

[10] M.E. and Lerman, S.R. Ben-Akiva, Discrete Choice Analysis: Theory and Application to Travel Demand.: MIT, 1985.

[11] W.H., Hensher, D.A. and Rose, J.M. Greene, Accounting for heterogeneity in the variance of the unobserved., 2005.

[12] http://en.wikipedia.org/wiki/Discrete_choice.

[13] A., Marshall, Principles of Economics. An introductory Volume, 8th ed., 1920.

[14] K. Train, Discrete Choice Methods with Simulation.: Cambridge University Press , 2003.

[15] DW., Hosmer and S., Lemeshow, Applied Logistic Regression , 2nd ed., 2000.

[16] J. de D. Ort´uzar, Fundamentals of discrete multimodal choice modelling.Transport Reviews 2., 1982.

[17] Michiel, ed. Hazewinkel, "Maximum-likelihood method", Encyclopedia of Mathematics," Springer, no. 978-1-55608-010-4.

APPENDIX

Appendix: Questionnaire

Transportation in Famagusta

Instructor:	Sina Darban Khales	Name:	______
Class:	Master student of transportation	Date:	______
	Engineering	Results:	______

Instructions

Thank you for your time

1) _____ **Do you have private car?**

a. Yes

b. No

2) _____ **Which mode of transport do you use to go from your home to the university?**

a. My own car

b. Bus

c. Taxi

d. Bicycle

e. On foot

3) _____ **Which parameter is the most important in your view for choosing a transit mode?**

a. Travel Time

b. Travel Cost

c. Travel Comfort

4) _____ **Which parameter is the second most important in your view for choosing a transit mode?**

a. Travel Time

b. Travel Cost
c. Travel Comfort

5) _____ **Where do you live? (Please mention how much time it needs to go from your home to the university)**

a. Far from university
b. Near the university
c. In campus
d. None of the above

6) _____ **Please select the reasonable taxi fare in your opinion from the university to the mosque square.**

a. 7.0 TL
b. 5.0 TL
c. 4.0 TL
d. 3.0 TL

7) _____ **Will you use taxi more frequently if taxi drivers or the government decrease the prices?**

a. Yes
b. No
c. I don’t Know

8) _____ **Which one is the biggest problem of bus service in Famagusta?**

a. Bad Scheduling
b. Bus-Stop locations
c. Crowdedness
d. Insufficient bus routes and stations

9) _____ **Is that a good idea to sell cheap bus tickets to provide the required budget with the goal of enhancing the level of service?**

a. Yes
b. No
c. It depends on the price